Particle Physics and Cosmology: The Interface

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Springer (formerly Kluwer Academic Publishers) in conjunction with the NATO Public Diplomacy Division.

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Springer (formerly Kluwer Academic Publishers)
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Springer (formerly Kluwer Academic Publishers)

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are "Advanced Study Institutes" and "Advanced Research Workshops", and the NATO Science Series collects together the results of these meetings. The meetings are co-organized by scientists from NATO countries and scientists from NATO's Partner countries — countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series was re-organized to the four sub-series noted above. Please consult the following web sites for information on previous volumes published in the Series.

http://www.nato.int/science
http://www.springeronline.com
http://www.iospress.nl

Series II: Mathematics, Physics and Chemistry – Vol. 188

Particle Physics and Cosmology: The Interface

edited by

D. Kazakov
JINR, Dubna and ITEP,
Moscow, Russia

and

G. Smadja
Université Claude Bernard,
Lyon 1, France

Springer

Published in cooperation with NATO Public Diplomacy Division

Proceedings of the NATO Advanced Study Institute on
Particle Physics and Cosmology: The Interface
Cargèse, France
4–16 August 2003

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 1-4020-3160-2 (PB)
ISBN 1-4020-3159-9 (HB)
ISBN 1-4020-3161-0 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Springer,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Springer,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Contents

Contributing Authors

Riccardo Barbieri Professor at Scuole Normale Superiore, Pisa

Pierre Binetruy Professor at University Paris 11 and Paris 7

Jacques Bouchez Research staff at DAPNIA, CEN Saclay

Wilfried Buchmuller Professor at the University of Hamburg and DESY staff.

Wim De Boer Professor at the University of Karlsruhe

Dimitri Kazakov Research staff at JINR (Dubna) and ITEP (Moscow, Russia)

David Langlois Research staff at IAP (Paris)

Hans-Peter Nilles Professor at the university of Bonn

Stefan Pokorski Professor at the University of Warsaw

Riccardo Rattazzi Research staff at CERN (Geneva)

Joseph Silk Professor at the University of Oxford

Gerard Smadja Professor at the University Claude Bernard (Lyon 1)

George Smoot Professor at the University of California at Berkeley

Achille Stocchi Research staff at LAL-Orsay (CNRS)

Foreword

This Cargèse school of Particle physics is meant to bridge the narrowing gap between astrophysical observations and particle physics. The lectures supply the students with a theoretical background which covers several aspects of the cosmological scenario: matter-antimatter asymmetry, the nature of dark matter, the acceleration of the expansion and the cosmological constant and the geometry of the universe as well as modern views on particle physics including supersymmetry, extra dimensions scenarii and neutrino oscillations.

Preface

The investigation of nuclear abundances by Alpher, Bethe, and Gamow (1948) was the first intrusion of subatomic physics into cosmology. In contrast with their assumption, most nuclear species are now known to be produced in stars, but their bold step led to predictions which have largely been proven to be right:
-a crude estimate of the densities during primordial nucleosynthesis
-the presence of a residual 3K radiation today.
the issues they addressed are still relevant. The origin of matter is not fully understood,and the CMB has grown into a powerful tool to investigate the early eras of the universe.

The progress of cosmological observations has now led to a 'standard' slow-roll inflation model, which accounts quantitatively for many observed features of the universe. As the lectures will show, it still leaves large unchartered areas, and the underlying particle physics aspects are yet to be elucidated.

At the same time despite the unprecedented success of the Standard Model of fundamental interactions there are still remaining "white spots". Possible physics beyond the SM may have its manifestation in astrophysics and cosmology which may serve as a sky laboratory to reveal the new features of the microworld.

The present volume contains the collection of lectures on the hottest topics in particle physics and cosmology given by the experts which describe the modern status and the perspectives of development of experimental and theoretical activities in these fields.

Acknowledgments

The school has been supported by the following institutions:
NATO, CNRS (Formation permanente, IN2P3, INSU/PNC, Collectivité Territoriale Corse, DAPNIA/CEN-Saclay

We also want to acknowledge the major contribution of the IPNL staff who helped organize the school: M. Chartoire, M. Croizé, D. Jarroux-Déclais, S. Florès, and Anne-Marie Ferrer, who took care of all financial problems. At the Cargèse Institute, Brigitte Cassegrain, Nathalie Bedjai, and Pierre Grossi were very helpful and ensured the smooth operation of the lectures and the distribution of the teaching material.

ELECTROWEAK SYMMETRY BREAKING AS OF 2003, ON THE WAY TO THE LARGE HADRON COLLIDER*

Riccardo Barbieri
Scuola Normale Superiore and INFN, Pisa, Italy
riccardo.barbieri@sns.it

Abstract I review the status of the ElectroWeak Symmetry Breaking problem. The lectures are naturally divided into two parts. The first is mostly devoted to overview the impact of current data on the issue of EWSB. The tools are known, the latest data are included. Always in the first part, I say why I care about the "little hierarchy" problem and I summarize how some proposals for EWSB, recent and less recent, are confronted with this problem. Motivated by these considerations, in the second part I describe the essential features of a proposal for breaking supersymmetry, and consequently the electroweak symmetry, by boundary conditions on an extra dimension.

Keywords: Electroweak, Symmetry, Breaking, Higgs mass, Supersymmetry, Kaluza-Klein, MSSM

The data (their interpretation) summarized

There are several good reasons for being interested in the problem of how the electroweak symmetry gets broken. Above all, the physical origin of the Fermi scale has not been identified, yet. Consequently, and not less importantly, this ignorance acts as a cloud on every attempt to design a theory of the fundamental interactions beyond the Standard Model (SM). Last, but not least, the exploration of the TeV scale of energy expected at the Large Hadron Collider (LHC) should finally allow

*Lectures given at the Cargese School of Physics and Cosmology - August 2003 - Cargese - France

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 1–29.

a direct comparison with experiment of every theoretical idea on this matter.

These are not the first lectures on the subject of ElectroWeak Symmetry Breaking (EWSB). Nevertheless, I find that it may be useful to overview the present status of the subject when we still have a few years before the start of the LHC and when the program of the ElectroWeak Precision Tests (EWPT), in particular with the completion of most of the data analyses by the LEP experiments, is in a mature stage.

The EWPT are still the most important source of experimental information, although indirect, on EWSB. It so happens that I already lectured in Cargese on the EWPT in 1992, when the accumulation of significant experimental results on the EWPT was about to start and the top quark had not yet been discovered. At that time the bulk of the radiative effects seen in the data was still of electromagnetic origin. Now we know that several per-mil effects of pure electroweak nature are crucial in allowing an effective description of the data and that these effects are contained in the SM. Things might have gone differently.

What is it then that we learn on the EWSB problem? Among the conclusions of my 92 lectures, I argued that the program of the EWPT should have made possible to discriminate between a perturbative and a strongly interacting picture of EWSB, the prototype examples for the two cases being respectively supersymmetry and technicolor. It is now pretty clear that the data support a perturbative more than a non perturbative description of EWSB, as illustrated more precisely later on. Inside this framework, a relevant piece of information, also coming from the EWPT is the indication for a light Higgs, most likely lighter than about 200 GeV. All this seems in fact to make a rather coherent picture of EWSB, and maybe it does. I will argue, however, that the direct lower limit on the Higgs mass, $m_H > 115\,GeV$ [1], and the absence so far on any deviation from the expectations of the SM may also require some interpretation with a possible impact on the picture of EWSB.

The lectures are naturally divided into two parts, to be found in Sections 2 and 3 respectively. The first is mostly devoted to overview the impact of current data on the issue of EWSB. The tools are known, the latest data are included. Always in the first part, I will illustrate why I care about the "little hierarchy" problem and I will summarize how some proposals for EWSB, recent and less recent, are confronted with this problem. Motivated by these considerations, in the second part I will describe the essential features of a proposal for breaking supersymmetry, and consequently the electroweak symmetry, by boundary conditions on an extra dimension. In Section 4 I summarize the line of reasoning that motivates mostly these lectures and I conclude.

1. Un updated overview

1.1 The data (their interpretation) summarized

Experiment versus theory with generic "oblique" corrections.

I begin by referring to the data on the EWPT from the LEP, TEVATRON and SLC experiments, as summarized by the LEP ElectroWeak Working Group in the summer of 2003 [2]. These data allow a stringent test of the SM, sensitive to the radiative corrections of electroweak nature. The test is successful, with no serious reason of concern, in my view, for those measurements that appear in some tension with the SM prediction.

The EWSB sector of the SM has some impact on this test. For this very reason, taking into account that this is the physics whose nature we are wondering about, one has advocated since the beginning of the experimental program an analysis valid in a broader class of theories. Such are those theories that differ from the SM only in the so called "oblique" corrections [3], i.e. those corrections that come from vacuum polarization amplitudes of the vector bosons. To this purpose one has defined three dimensionless experimental quantities, $\epsilon_i, i = 1, 2, 3$, [4] with the property that they encapsulate, among other effects, the "oblique" corrections. Furthermore, since one of them, ϵ_2, is unlikely to contain new physics, one often freezes it to its SM value [1]. Fig. 1 shows the determination of the two remaining parameters, ϵ_1 and ϵ_3, with their correlation, as obtained from the set of data mentioned above. In the same Figure, the SM prediction is indicated for different values of the Higgs mass.

The "oblique" contributions to ϵ_1 and ϵ_3 can be expressed in terms of the transverse components of the usual vacuum polarization amplitudes $\Pi_{a,b}(q^2)$, with $a, b = 1, 2, 3$ for the SU(2) or $a, b = B$ for the U(1) gauge bosons, as

$$\Delta\epsilon_1 = \frac{\Pi_{33}(0) - \Pi_{11}(0)}{M_W^2}, \tag{1}$$

$$\Delta\epsilon_3 = \frac{g}{g'}\Pi'_{3B}(0), \tag{2}$$

where g and g' are the SU(2) and U(1) gauge couplings.

The data speak by themselves. The agreement with the SM is remarkable and constitutes indirect evidence for the existence of the Higgs. It has even become difficult, if not impossible at all, to try to reconcile

[1] The slight dependence of ϵ_2 on m_H is practically irrelevant. Here ϵ_2 is taken at $m_H = 115\,GeV$.

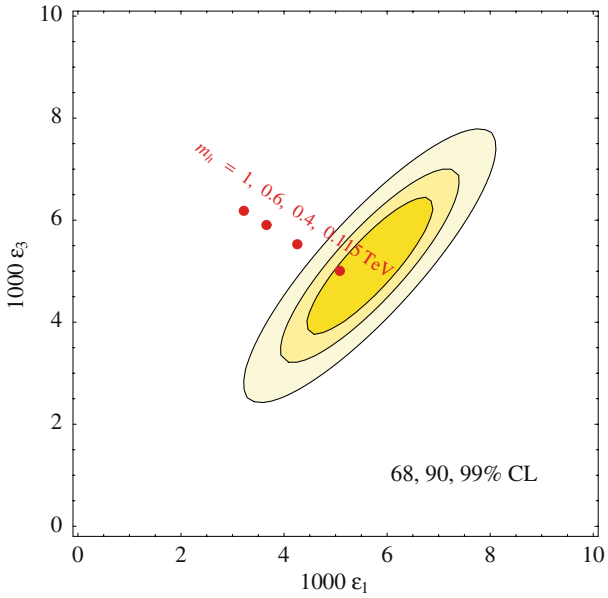

Figure 1. ϵ_1 and ϵ_3, with their correlation, as determined from the EWPT. The SM prediction is indicated for different values of the Higgs mass.

with the data a heavy Higgs, in the TeV range of mass where the very concept of the Higgs becomes elusive, by adding an extra negative contribution to ϵ_3. As commented below, a positive extra effect on ϵ_1, or a combination of two effects, would appear more useful in this respect. On the other hand, and always with a heavy Higgs, a positive extra contribution to ϵ_3 above 2 units in 10^{-3} is excluded, no matter what ϵ_1 does. If there is a Higgs, which seems more likely, Fig. 1 makes also clear the preference of the EWPT for a low mass value, close to the direct lower bound of 115 GeV.

Constraining the cut-off of the SM. An alternative way to try to appreciate the impact of the EWPT is to view the Lagrangian of the SM as an effective low energy theory with possible modifications introduced by operators $\mathcal{O}_i^{(4+p)}$ of dimension $(4+p)$,

$$\mathcal{L}_{eff}(E<\Lambda) = \mathcal{L}_{SM} + \sum_{i,p} \frac{c_i}{\Lambda^p} \mathcal{O}_i^{(4+p)}. \qquad (3)$$

How does this modified Lagrangian compare with the most recent EWPT [5]? Table 1 gives a list of some gauge invariant operators, together with the lower limit that the same EWPT set on the scale Λ that multiplies

Dimension six operator	$c_i = -1$	$c_i = +1$
$\mathcal{O}_{WB} = (H^+\sigma^a H)W^a_{\mu\nu}B_{\mu\nu}$	9.0	13
$\mathcal{O}_H = \|H^+ D_\mu H)\|^2$	4.2	7.0
$\mathcal{O}_{LL} = \frac{1}{2}(\bar{L}\gamma_\mu\sigma^a L)^2$	8.2	8.8
$\mathcal{O}_{HL} = i(H^+ D_\mu H)(\bar{L}\gamma_\mu L)$	14	8.0

Table 1. 95% lower bounds on Λ/TeV for the individual operators (and the coefficients c_i as indicated) with $m_H = 115\,GeV$. σ^a are the Pauli matrices acting on the $SU(2)_L$ doublets.

each of them. One operator at a time is taken, with the dimensionless coefficients $c_i = +1$ or $c_i = -1$ (the interference with the SM amplitude matters) and the Higgs mass (which affects the SM amplitude) at 115 GeV.

Table 1 gives a feeling on the possible lowest energy at which a change of regime can intervene relative to the SM physics, called the "cut-off " of the SM. To be able to make a more precise statement, the coefficients c_i would have to be known. On general grounds, the most important conclusion from Table 1 is that, if a new strong interaction intervenes at Λ, it is unlikely that this scale be lower than about 10 TeV. A different statement might be defendable, but only on a case by case basis.

In the case of the operators $\mathcal{O}_{WB}$ and $\mathcal{O}_H$, but only in these cases, there is a direct connection of Table 1 with Fig. 1, since these operators affect the EWPT only by contributing to the ϵ-coefficients, via

$$\delta\epsilon_3(\mathcal{O}_{WB}) = \frac{2g}{g'}c_{WB}\frac{v^2}{\Lambda^2}, \tag{4}$$

$$\delta\epsilon_1(\mathcal{O}_H) = -c_H\frac{v^2}{\Lambda^2}. \tag{5}$$

It is also clear therefore that a theory capable of modifying the SM by the inclusion of a single dominant operator $\mathcal{O}_H$ would allow a fit to the EWPT with a heavier Higgs (with $c_H = -1$ and a corresponding scale lower than in Table 1) [6, 5].

1.2 The "little hierarchy" problem

It is generally believed by many, including myself, that the discovery of the Higgs would not identify, *per se*, the physical origin of the Fermi scale. It is also a widespread opinion that the LHC experiments will reveal new phenomena related to EWSB and not included in the SM. Let us review the argument that supports this view. The general argument is old. Its numerical terms are new inasmuch as one uses in an essential

way the precise knowledge of the top mass and the indirect information on the Higgs mass, as just described.

For a fixed value of the Higgs vacuum expectation value (vev) or of the Fermi scale, the physical Higgs mass can either be computed from the quartic Higgs coupling or from the curvature of the Higgs potential at zero field. Using this last quantity, the one loop radiative correction to the squared Higgs mass in the SM is

$$\delta m_H^2 = \frac{6G_F}{\sqrt{2}\pi^2}(m_t^2 - 1/2m_W^2 - 1/4m_Z^2 - 1/4m_H^2)\int^\Lambda EdE, \qquad (6)$$

where $E = \sqrt{k^2}$, k is the momentum running in the loop, and terms of relative order $(m/\Lambda)^2$ for any particle mass are neglected. For the cut-off of the integral, taken equal for all the individual contributions, the use of the same symbol as the one employed for the cut-off of the SM is not without reason. As apparent from eq. (6), if one accepts that the Higgs mass is below 200 GeV, as I do in the following unless otherwise stated, the top loop contribution, proportional to the top mass squared, is clearly dominant. Furthermore it becomes numerically significant for a pretty low value of Λ, since, using $m_t = 174GeV$,

$$\delta m_{H,top}^2(SM) \sim (115GeV)^2(\frac{\Lambda}{400GeV})^2. \qquad (7)$$

It will be useful for later use to define a kind of "Higgs mass spectral function" $\Delta m_H^2(E)$ which enters the expression of the radiative Higgs squared mass as

$$\delta m_H^2 \equiv \int^\Lambda \frac{dE}{E}\Delta m_H^2(E). \qquad (8)$$

In the SM, $\Delta m_H^2(E)$ grows quadratically with E, it becomes rapidly large and it gives the main contribution to the Higgs mass for E close to the cut-off where it cannot be trusted.

Eq. (7) is the main argument for expecting new physics to show up at the LHC. How can it possibly be that $\Delta m_H^2(E)$ keeps growing unmodified up to energies unaccessible to the LHC, when already at 400 GeV the radiative correction to the Higgs mass is around the best value implied by the EWPT as discussed above? Note that this is not an inconsistency of the SM. Since the curvature of the Higgs potential is a free parameter of the SM Lagrangian, it only takes a little counterterm to subtract away the loop effect that we are discussing. I prefer to think - or to assume, or to hope, as one prefers - that the Higgs mass is what it is not as a result of a *fortuitous* cancellation between $\delta m_{H,top}^2$ and

the counterterm or any other effect. For this to be the case, a physical mechanism must prevent $\Delta m_H^2(E)$ from growing. In turn, this physical mechanism, not included in the SM, should give effects observable at the LHC.

Ideally, one would like to be quantitative here and be able to estimate at which energy these new phenomena should show up, since this is likely to be crucial for a test at the LHC. This is possible in specific models and, if it is the case, for a given amount of fine tuning that one is willing to tolerate. On general grounds, however, it is possible to say that the physical mechanism that dumps $\Delta m_H^2(E)$ has to pass a nontrivial test: it must not disturb the agreement of the SM with the EWPT, as summarized in the previous Section. To find such a mechanism, involving a minimum amount of tuning among otherwise uncorrelated parameters, is to solve the "little hierarchy" problem [7]. The hierarchy is between the Higgs mass, supposedly close to $100\,GeV$, and the lower limit on the cut-off of the SM, as elaborated upon always in the previous Section.

1.3 Some ideas on EWSB with a light Higgs

In this overview, I find it useful to comment on two proposal for EWSB, since they may accommodate or even imply a light Higgs, while addressing at the same the "little hierarchy" problem. One is EWSB triggered by supersymmetry breaking, as in the Minimal Supersymmetric Standard Model (MSSM). The other assumes some new strong force which gives rise to a pseudo-Goldstone boson with the quantum numbers of the standard Higgs, supplemented with the mechanism that goes under the name of "little Higgs". A general overview of these two schemes goes beyond the scope of these lectures. Rather I concentrate my attention on the way they tackle the "little hierarchy" problem.

The Minimal SuperSymmetric Standard Model. The MSSM neatly implies a light Higgs and a perturbative dynamics of EWSB. The Higgs is light because the quartic Higgs coupling is a gauge coupling. The perturbative dynamics gives, in principle, the easiest way to solve the little hierarchy problem. The MSSM has indeed no special problem in preserving the success of the SM in describing the EWPT. The effort that has gone and will possibly go in MSSM studies appears therefore well justified. The most important question nevertheless remain. Where are the supersymmetric particles? Even more concretely, which are the chances that they will be seen at the LHC? The possible answer, at the present state of knowledge, goes inevitably back to the fine tuning problem in the curvature of the Higgs potential [8].

As well known, the MSSM involves two Higgs doublets. For the present purposes, however, there is no essential loss in assuming that the vev taken by one of them, the one that couples to the up quarks, is far bigger than the vev of the other, i.e., in the familiar notation, $tan\beta \gg 1$. In this case the previous discussion applies unaltered and we can focus, as before, on $\Delta m^2_{H,top}(E)$.

The way taken by the MSSM in dumping $\Delta m^2_H(E)$ is almost too well known to be recalled here. In the case of $\Delta m^2_{H,top}(E)$, the exchange of the stops freeses it when E crosses their masses and keeps it constant up to the scale - call it Λ_{SB} - where these masses are generated by the specific mechanism of supersymmetry breaking and supersymmetry breaking transmission. In this way, in logarithmic approximation, $\delta m^2_{H,top}$ is modified into [2]

$$\delta m^2_{H,top}(MSSM) \sim \frac{6G_F m_t^2}{\sqrt{2}\pi^2} m^2_{ST} log \frac{\Lambda_{SB}}{m_{ST}} \sim 0.15\, m^2_{ST} log \frac{\Lambda_{SB}}{m_{ST}}, \quad (9)$$

where m_{ST} is a suitable average of the stop masses. and can be sizeable if m_{ST} is big enough. But a big m_{ST} brings back the problem in eq. (9), since $\delta m^2_{H,top}$ grows quadratically with it. Recent studies [9] in the complex space of the MSSM parameters in its supergravity version [14] (but the situation in gauge-mediated models [11] is not better) confirm this simple argument and set in at least a factor of 20 the cancellation required against the term in eq. (9) - or its counterpart - with a favorable choice of various parameters (A_t, $tan\beta$, etc.). Ref. [12], as updated in [13], has a rather striking pictorial representation of this problem.

The current limits on generic sparticle masses from direct searches or from the EWPT, as mentioned, are somewhat less constraining. The portion of parameter space of the MSSM explorable at the LHC by direct sparticle production, a primary goal of the machine, is still sizeable. Nevertheless, in the spirit of the argument discussed above, the need of a significant cancellation in the curvature of the Higgs potential due to the present limit on the Higgs mass weakens any statement about where supersymmetry should be found, which is unfortunate. This may be irrelevant, with the MSSM around the corner of the parameter space explored so far. Alternatively it may indicate the necessity to go beyond

[2] This, as the one that follows, is a symplified discussion. To the least, when the energy range of interest is large compared to the Fermi scale, the influence on the stop masses of the gluino exchange is significant and has to be taken into account.

the MSSM [3] or, more drastically, the need of some variation on the current picture of supersymmetry breaking.

The little Higgs A neat way to have a light scalar, i.e. a candidate Higgs, from a strong dynamics is to arrange its symmetries and the pattern of symmetry breaking in such a way that this scalar comes out to be a pseudo-Goldstone boson, like a π in QCD [17]. The prefix *pseudo* is mandatory since a genuine Goldstone could neither get a mass nor it could have a Yukawa coupling to the fermions. Both could arise, however, from a "weak" explicit breaking of the original global symmetry.

Let us concentrate on the top Yukawa coupling which is the major source of the growing $\Delta m_H^2(E)$, as defined in Sect. 1.2. If this coupling between the top and the Goldstone boson were directly introduced in the Lagrangian, no change would have occurred in $\Delta m_{H,top}^2(E)$ relative to the SM up to the scale Λ of the strong dynamics. Based on Sect. 1.1, this could hardly be considered, however, as a satisfactory solution of the "little hierarchy" problem: $\Delta m_{H,top}^2(E)$ needs to be cut-off well before getting to such a scale. In "little Higgs" models [18] this is achieved by adding another top-like quark which also couples to the Goldstone-Higgs in a globally symmetric way but has a non-symmetric, although gauge invariant, mass term. In this way, and without any adjustment of uncorrelated parameters, when E crosses the mass of the heavy top, m_T, $\Delta m_{H,top}^2(E)$ gets frozen and stays constant up to Λ, where it eventually dies out. Consequently, up to non-logarithmic terms, eq. (7) is replaced by

$$\delta m_{H,top}^2(LH) \sim \frac{6G_F m_t^2}{\sqrt{2}\pi^2} m_T^2 log\frac{\Lambda}{m_T}. \tag{10}$$

This is a significant step forward in the direction of solving the "little hierarchy" problem. With some caveats, though. Le me denote by f the "pion decay constant" of the strong dynamics at Λ, so that $\Lambda \sim 4\pi f$. It is $m_T > 2\lambda_t f \sim 2f$, where λ_t is the usual top Yukawa coupling, so that, from eq. (10),

$$\delta m_{H,top}^2(LH) \gtrsim 1.2 f^2. \tag{11}$$

Now the minimal value that f can attain without undoing the EWPT is very much model dependent [19, 20]. However, taking $f > 1$ TeV as a minimum limit in suitably designed models gives already quite a big

[3] To correct the Higgs mass prediction at tree level by suitable F-term [14] or D-term [15] effects or by higher dimensional operators [16] are examples in this direction.

contribution to the Higgs mass, that has to be cancelled away. I would care less about this problem if the model could be tested at the LHC even for f significantly greater than 1 TeV, but this may not be the case [21, 20]. And I do not see any strong reason for insisting on $f \sim 1$ TeV, given that quite a significant cancellation needs anyhow to be swallowed.

Needless to say, in spite of this warning, I believe that possible signals of the "little Higgs" should be carefully looked for.

2. Relating G_F to an extra dimension at the weak scale

2.1 Motivations and the basic setup

To contemplate the possibility that there exist one or more extra space dimensions may undoubtedly give a new twist to several of the basic problems in fundamental physics. As remarkable examples it suffices to mention the possible connection between the weakness of gravity and a large extra dimension [22] or the interpretation of the "big" hierarchy, the one between M_W and M_{Pl}, as a gravitational blue-shift by a warp factor with a non trivial dependence on a 5th coordinate [23]. Here I am specifically concerned with the EWSB problem and with the proposition that the Fermi scale be related to the inverse radius of a compactified extra dimension [24–26]. Even this qualification, however, leaves open several possibilities whose description goes beyond the scope of these lectures. In the following I concentrate on a proposal that aims at pushing as far as possible the calculability of the Higgs potential. As it should be clear by now, I look for an optimal solution of the "little hierarchy" problem. To the least, I expect that this will maximize the chances of having this solution tested at the LHC.

The aspect of extra-dimensional theories that interests me most and could have phenomenological applications is the possibility, intrinsically extra-dimensional, of breaking symmetries by boundary conditions on the various fields at the borders of the extra dimension(s). It is in fact even possible to break the same electroweak symmetry in such a way [27], but it is doubtful whether one would improve over 4-dimensional (4D) theories [28]. On the contrary, breaking supersymmetry by boundary conditions on an extra dimension [29] looks more promising for reasons that will appear shortly.

All the SM fields are meant to depend also on an extra coordinate y, $\Psi(x_\mu, y)$, ranging from 0 to $L \equiv \pi R/2$, and are extended to incorporate supersymmetry in the full 5D space. The physics will depend upon the conditions given to the various fields at the boundaries of the y-segment

$(0, \pi R/2)$. Take $(even, even)$ boundary conditions for all the SM fields[4], so that each of them has a constant mode in y, with vanishing momentum in the fifth direction, which would otherwise appear as a mass term from the 4D point of view. These "zero modes" are massless. On the contrary, give $(even, odd)$, $(odd, even)$ or (odd, odd) boundary conditions to the extra fields implied by Poincar invariance and supersymmetry in 5D, consistently with the parities at the two boundaries. In this way all these extra fields give massive excitations: from the 4D point of view all their modes have masses which, in the simplest case, are multiple integers of $1/R$.

This simple construction [26] has a remarkable property. While supersymmetry is manifestly broken in a global way, since there is no supersymmetry in the spectrum - which makes it closer to experiment -, supersymmetry is locally unbroken. This will be spelled out more clearly in the next Section, but is already now intuitively clear. At each boundary all fields can be grouped into doublets of 4D N=1 supersymmetry (one boson and one fermion) with the same boundary conditions, even or odd, inside each supermultiplet. This is unavoidable, given the parities at the two boundaries. On the other hand, the N=1 supermultiplets at the two boundaries are different from each others because of our choice of boundary conditions. At every point of the segment $(0, \pi R/2)$ there is at least one 4D supersymmetry, but this supersymmetry changes from point to point since it is different at the two boundaries. This is the basis for the finiteness properties of the loop calculations that have been performed and we are about to describe. All the divergent *local* counterterms have to respect this residual symmetry. The change of supersymmetry at the two boundaries - and, as such, of the corresponding supermultiplets - also allows to describe the Yukawa couplings of the up and down quarks to a single Higgs field as in the SM and unlike in the MSSM, which is not a negligible simplification.

2.2 The $5D \to 4D$ projection and the residual symmetries

To realize the setup outlined above the fields, organized in N=1 supermultiplets in 5D, are the gauge fields $(A_\mu, \lambda_1, \lambda_2, \phi_\Sigma)$, the matter fields $(\psi_M, \phi_M, \psi^c_M, \phi^c_M)$ and the Higgs fields $(\phi_H, \psi_H, \phi^c_H, \psi^c_H)$. The indices of the $SU(3)xSU(2)xU(1)$ gauge group are left understood, as the flavor indices for matter fields. The boundary conditions assigned to each of

[4]The y-segment $(0, L)$ is fictitiously extended at $y < 0$ and $y > L$ and the equations of motion are meant to be solved in the whole covering space.

ψ_M, ϕ_H, A^μ	ϕ_M, ψ_H, λ_1	ϕ_M^c, ψ_H^c, λ_2	ψ_M^c, ϕ_H^c, ϕ_Σ
$(+,+)$	$(+,-)$	$(-,+)$	$(-,-)$

Table 2. Boundary conditions for gauge, matter and Higgs fields at $y = 0$ and $y = L$.

these fields, $(+ \equiv even, - \equiv odd)$, as already indicated, are given in Table 2. Only one Higgs supermultiplet (in 5D) appears in Table 2, since only one Higgs scalar will get a vev. The model defined in this way has no gauge anomaly (See App. A). The opportunity to introduce a second Higgs supermultiplet, without a vev, to cancel a Fayet Iliopoulos (FI) term (See App. B) will be discussed later on in Sect. 2.6.

The supersymmetry transformations for the vector fields, splitted in A_μ, $\mu = 1, 2, 3, 4$, and A_5, are, as an example,

$$\delta A_\mu = i\xi_i^+ \sigma_\mu \lambda_i + h.c., \;\; \delta A_5 = \epsilon^{ij} \xi_i \lambda_j + h.c., \tag{12}$$

where ξ_i, $(i, j = 1, 2)$ are the two spinorial transformation parameters associated with the 2 supersymmetries in 4D implied by the N=1 supersymmetry in 5D. It is manifest from eqs. (12) that the boundary conditions in Table 2 are, in general, not compatible with the supersymmetry transformations, since fields with different boundary conditions transform into each other under (12). Hence we cannot expect that supersymmetry remains unbroken. Nevertheless the consistency between boundary conditions and supersymmetry transformations is kept if the y-dependent ξ_i are also restricted to satisfy appropriate boundary conditions, i.e. $(+,-)$ and $(-,+)$ respectively for ξ_1 and ξ_2. This defines the residual supersymmetry that remains intact after the $5D \to 4D$ projection, as alluded to in the previous section. The same restricted supersymmetry transformations determine the appropriate supermultiplets at the two boundaries. For instance, the two N=1 supermultiplets that contain the Higgs field ϕ_H are $\widehat{H} = (\phi_H, \psi_H)$ at $y = 0$ and $\widehat{H'_c} = (\phi_H^+, \psi_H^c)$ at $y = L$. Accordingly, the total 5D Lagrangian is

$$\mathcal{L} = \mathcal{L}_5 + \delta(y)\, \mathcal{L}_4 + \delta\left(y - \frac{\pi R}{2}\right) \mathcal{L}'_4, \tag{13}$$

where $\mathcal{L}_5$, the so called "bulk" Lagrangian, respects a full N=1 supersymmetry in 5D, whereas $\mathcal{L}_4$ and $\mathcal{L}'_4$ are 4D Lagrangians invariant under the (different) relevant supersymmetries. Among other things, $\mathcal{L}_4$ and $\mathcal{L}'_4$ will contain, the first, the Yukawa coupling of the up quarks to $\widehat{H}$ and, the second, the Yukawa couplings of the down quarks and of the lep-

tons to $\widehat{H'_c}$. These couplings could not be placed anywhere else without spoiling the residual supersymmetry.

2.3 Spectrum of the Kaluza Klein modes

The solution of the free equations of motion, with the assigned boundary conditions, gives the spectrum of the various modes. This spectrum is particularly simple when one ignores possible kinetic terms localized at the boundaries or gauge invariant "bulk" masses for the matter (or Higgs) supermultiplets. Corresponding to the wave functions the spectrum is shown in Fig. 2.

$$\begin{aligned}
(+,+) &: \quad \cos \tfrac{2n}{R} y; \\
(+,-) &: \quad \cos \tfrac{2n+1}{R} y; \\
(-,+) &: \quad \sin \tfrac{2n+1}{R} y; \\
(-,-) &: \quad \sin \tfrac{2n}{R} y,
\end{aligned} \tag{14}$$

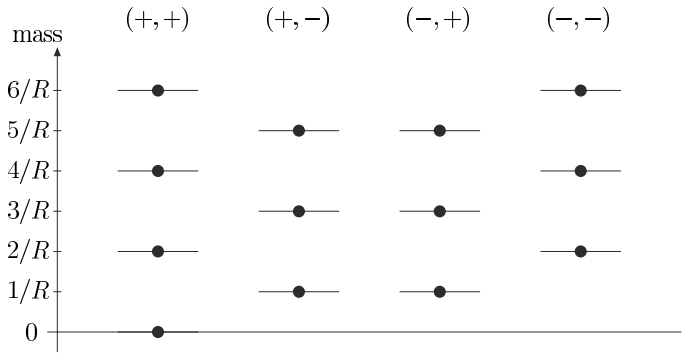

Figure 2. Tree-level KK mass spectrum of a multiplet (vector, matter or Higgs) with the boundary conditions as indicated.

Of special interest for the following is the deformation of the spectrum of matter supermultiplets, and of the corresponding wave functions, in presence of a bulk mass term, consistent both with gauge invariance and with the residual supersymmetry. This mass term, in general, can have a y-dependence and, for consistency with the parities at the two boundaries, must be (odd, odd). Considering a constant mass term M inside the

segment $(0, L)$, all the KK modes become heavier, with a mass growing like M for $MR >> 1$, except for the fermionic zero mode, which remains massless, and one of the two lightest states in the towers of scalars, a sfermion, whose mass decreases from $1/R$ and tends asymptotically to zero as MR increases [30]. The wave functions are modified accordingly. For the massless fermion, taking, e.g., $M > 0$, the corresponding wave function is

$$\xi_0(y) = [\frac{2}{M}(1 - \exp(-\pi MR))]^{-\frac{1}{2}} \exp(-My), \tag{15}$$

giving rise to a partial localization towards one of the boundaries ($y = 0$ for $M > 0$), as it happens similarly for the lightest scalar. The physical interpretation is the following: when M becomes larger than $1/R$, supersymmetry is progressively recovered in the spectrum, with a light quasi-supersymmetric multiplet at one boundary and a tower of massive supermultiplets in the bulk. From a theoretical point of view, the bulk mass M is a new parameter, but with a special status: it does not undergo any renormalization.

2.4 A first attempt: the Constrained Standard Model

What happens of the EWSB with the simplest setup described in Sect. 2.2 ignoring, for the time being, bulk masses and boundary kinetic terms? Needless to say, we are mostly interested in the loop corrections to the Higgs mass from the towers of top-stop states. There are several ways of doing this calculation, either involving a sum over the towers of intermediate KK states [26] or by working with propagators in mixed (p_μ, y) space for the different components of the superfields [31]. With a notation appropriate to this second case, the relevant diagrams are shown in Fig. 3, where $h^0(x)$ is the y-independent component of the Higgs field and F_u, F_t are the auxiliary components of the $\widehat{u}$ and $\widehat{t}$ supermultiplets, which we have not eliminated.

The result can best be seen by means of the function $\Delta m^2_{H,top}(E)$ introduced in Sect. 1.2. It is [26, 31]

The result can best be seen by means of the function $\Delta m^2_{H,top}(E)$ introduced in Sect. 1.2. It is [26, 31]

$$\Delta m^2_{H,top}(E) = \frac{3\sqrt{2}}{4\,R^2} G_F m_t^2\, x^4 \left[\coth^2\left(\frac{\pi x}{2}\right) - \tanh^2\left(\frac{\pi x}{2}\right)\right] \tag{16}$$

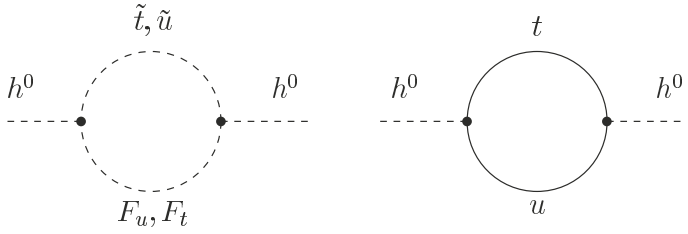

Figure 3. One-loop irreducible diagrams contributing to the mass squared of the Higgs boson.

where $x = ER$. For $E << 1/R$, $\Delta m^2_{H,top}(E)$ goes to the SM expression, as it should, whereas in the opposite limit, above the compactification scale, $1/R$, $\Delta m^2_{H,top}(E)$ gets exponentially dumped to

$$\Delta m^2_{H,top}(E) \sim 6\sqrt{2} G_F m_t^2 E^4 R^2 \exp(-\pi E R). \tag{17}$$

In this way a finite $\delta m^2_{H,top}$ is obtained

$$\delta m^2_{H,top}(CSM) = \frac{63\zeta(3)}{\sqrt{2}\pi^4} \frac{G_F m_t^2}{R^2} = \frac{0.19}{R^2}, \tag{18}$$

where $\zeta(3) = 1.20$. I find this an interesting step forward relative to eq. (7) or even to eq. (9), which justifies a further exploration of the general idea.

It is possible to go over a complete calculation of the Higgs potential in this simplest case where the top has a flat wave function in y. This requires a straightforward extension of the previous calculation to the entire Higgs potential [26]. Such a calculation is relevant both because of the corrections to the quartic Higgs coupling and because of the higher order terms $(h^0)^4 (h^0 R)^n$ [5]. This corresponds to the minimal implementation of the idea, which we have called Constrained SM [26], since the relevant part of the full Higgs potential involves a single parameter, R, instead of two, as in the SM. Consequently, given the physical value of the Fermi scale, the Higgs mass is determined to be about 130 GeV and the inverse radius is in the 400 GeV range. Note in particular that, although eq. (18) gives the dominant term in the curvature of the Higgs

[5] With a single Higgs supermultiplet, a quadratically divergent tadpole term in the auxiliary field of the hypercharge vector supermultiplet, a FI term (See App. B) appears [32]. However, for a sensible cut-off, as determined below in Sect. 2.7, this term is of little numerical significance to the Higgs potential [36]. Alternatively (see Sect. 2.6) a second Higgs supermultiplet can be introduced which cancels the divergent FI term and is prevented from getting a vev by a bulk mass, small relative to $1/R$.

potential at zero field, the relation between the Higgs mass and $1/R$ cannot be simply read from eq. (18). The curvature of the potential has the standard relation with the Higgs mass only when the powers of h^0 higher than four in the potential are unimportant, which is not the case here.

2.5 Localizing the top: the physical picture

The properties of the top wave function in the fifth dimension are crucial to the numerical determination of the EWSB parameters [30, 33]. In the CSM the top wave function is y-independent. As discussed below in Sect. 2.7, consistency with the EWPT may require the top to be partly localized in y, which introduces an extra parameter. The Higgs mass becomes then a function of $1/R$ which is, in turn, not determined [34, 35].

As already mentioned, in the CSM the one loop top correction to the Higgs mass squared dominates over the gauge corrections, which contribute as [24]

$$\delta m^2_{H,ew}(CSM) = -\frac{7\zeta(3)(3g^2+g'^2)}{8\pi^4 R^2} = -1.5\frac{10^{-2}}{R^2}, \qquad (19)$$

i.e. one order of magnitude less than eq. (18) and with opposite sign. Note that the relative size between the top and the electroweak contributions is approximately maintained as in the SM with a universal cut-off, eq. (6).

Compared with this situation, what one should expect with a localized top, through a bulk mass term M, is intuitively clear. The progressive localization of the top at one boundary, as M increases, makes it feel less and less the global breaking of supersymmetry through appropriate conditions at the two boundaries. Consequently the top contribution to the curvature of the potential, which vanishes in the supersymmetric limit, decreases its numerical significance. Hence a cancellation takes place between the top and the electroweak contributions, which becomes more and more significant as M increases. In turn, for the physical value of the Fermi scale, this results in an increase of $1/R$ from the relatively low value of the CSM.

To clarify even further the physical picture, it is useful to see what happens in the extreme case of exact localization of the top at the $y = 0$ boundary, by which I mean that the left handed doublet, $\widehat{Q}$, and the right handed singlet, $\widehat{u}$, that contain the top quark, only appear as 4D N=1 supermultiplets in $\mathcal{L}_4$, eq. (13), with no KK towers. In this case, no supersymmetry breaking enters the diagrams of Fig. 3, which do not contribute, therefore, to the curvature of the potential. To see a

supersymmetry breaking effect one needs the exchange of a vector or a Higgs supermultiplet, through any of their components, which only occurs at two loops. An explicit two loop calculation gives in fact the following contribution to the Higgs mass, through the curvature of the potential, (LT for "localized top") [35]

$$\delta m^2_{H,top}(LT) = \frac{3G_F m_t^2}{\sqrt{2}\pi^2}(m_Q^2 \log\frac{c}{Rm_Q} + m_U^2 \log\frac{c}{Rm_U}), \qquad (20)$$

where

$$m_U^2 = \frac{7\zeta(3)}{24\pi}\frac{8\alpha_s + 6\alpha_t}{L^2}, \quad m_Q^2 = \frac{7\zeta(3)}{24\pi}\frac{8\alpha_s + 3\alpha_t}{L^2} \qquad (21)$$

are the stop squared masses induced at one loop [25], $\alpha_S = g_S^2/4\pi$, $\alpha_t = \lambda_t^2/4\pi$ and $c = 1.24$. Note again the finite result, as expected. Note also the connection between eq. (21), and the MSSM result, eq. (9). Numerically it is

$$\delta m^2_{H,top}(LT) = 1.4\frac{10^{-2}}{R^2}, \qquad (22)$$

quite close to $\delta m^2_{H,ew}(CSM)$, eq. (19), up to the opposite sign. The one loop electroweak contribution goes through unchanged from the CSM to the top localized case. The near cancellation between eq. (19) and eq. (22) could be at the origin of a significant increase in $1/R$ from the CSM value of about 400 GeV without fine tunings in the Higgs potential. In turn, as we shall discuss, this may help addressing the "little hierarchy" problem.

2.6 The Higgs potential in detail

To complete the analysis of EWSB, we need the Higgs potential for an arbitrary top localization, as determined by the continuous mass parameter M. There are at least two reasons for doing that, other than the logical one. At least the left handed doublets cannot be exactly localized at $y = 0$ if one wants to use only one Higgs with a nonzero vev to describe both the Yukawa couplings to the up and to the down quarks, necessarily sitting at two different boundaries (See Sect. 2.2). Furthermore, the sum of eq. (19) and eq. (22) gives an overall negative contribution which, literally taken, means no EWSB.

The Higgs potential with a partial top localization has been studied in Refs. [34, 35]. There we consider the $\widehat{Q}$ and $\widehat{u}$ supermultiplets in 5D with an equal bulk mass M. We also take two Higgs supermultiplets in 5D with opposite boundary conditions and an equal bulk mass M_H, so as to cancel any FI term. A mass term M_H is necessary to stabilize

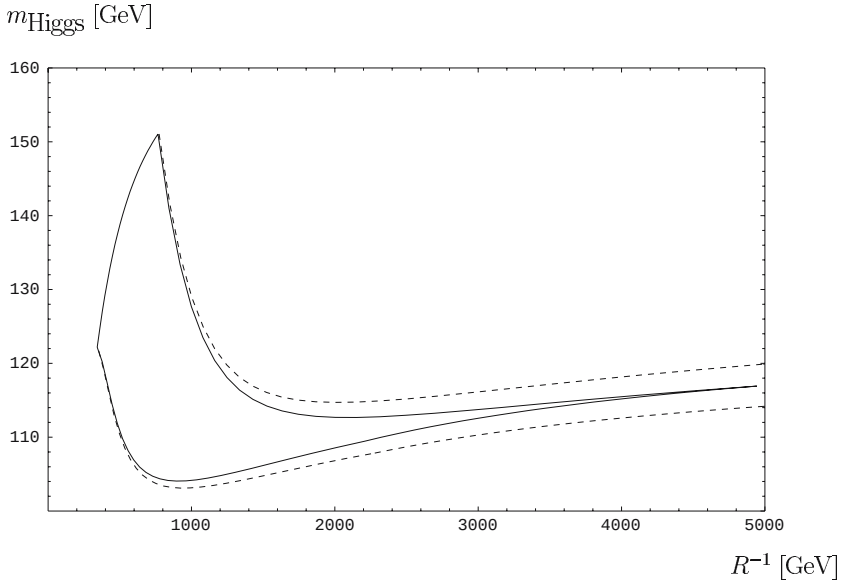

Figure 4. Range of the Higgs mass for a partially localized top as function of $1/R$. The full line is for $m_t^{pole} = 174.3\,GeV$. The dotted line is for $m_t^{pole} = 174.3 \pm 5.1 GeV$

the second Higgs scalar when MR becomes small. For $MR = 0$ this is therefore a variant with respect to the CSM, although still with a non localized top. On the contrary, for a sizable MR, M_H is progressively negligible.

As already mentioned, the introduction of the parameter M and the near cancellation noticed in the previous Section between the top and the electroweak contributions to the slope of the potential for $MR >> 1$ allow $1/R$ to vary as a free parameter in a wide range, from a few hundreds of GeV for low MR up to several TeV for $MR > 1$, without any significant fine tuning. In Fig 4 I show the range of the physical Higgs mass for $1/R$ below 5 TeV and M_H limited by a maximum cancellation in the slope of the potential at about 10% level. The raise of m_H at low values of $1/R$ is an effect of the KK modes of the top-stop towers. In the upper range of values for $1/R$, M_H is irrelevant. Values of $1/R$ above 5 TeV would imply a seemingly accidental cancellation between the top and the electroweak contributions by more than a factor of 10. Note that, for $1/R$ above about 1.5 TeV, m_H must be always close to the experimental lower limit of 115 GeV. It must also be said that the calculation of the Higgs potential in Ref. [35], upon which Fig. 4 is based, is accurate for low $1/R$, below about 1 TeV, where the one loop top contribution dominates, and for high $1/R$, above $3 \div 4$ TeV, where

a quasi-localized approximation for the top allows an analytic two loop calculation of the potential. In the transition region a better calculation might be necessary.

2.7 Constraints from Electroweak Precision Tests

Dominant effects. In general one expects effects on the EWPT from three different sources:

- Tree level effects from exchanges of heavy KK modes;
- Tree level effects from supersymmetric operators localized on the boundaries;
- Calculable effects from radiatively-generated non-supersymmetric operators.

In the following, for concreteness, I do not consider any localization parameter for the first two generations of matter[6]. In this case, if one neglects the operators on the boundaries, there is no tree level effect on the EWPT from the exchanges of heavy KK modes, because discrete momentum conservation on the 5th direction forbids any coupling bilinear in the SM particles (the zero modes) to a single KK state [26]. There are, on the contrary, calculable loop effects which are there only to the extent that supersymmetry is broken. A contribution to ϵ_3 would be the best example, as there are several other similar effects on observables not traditionally included among the EWPT, but not less important, like $b \to s + \gamma$ or the muon $g - 2$. Explicit calculations done in the last two cases (but not yet for ϵ_3 [37]) show that these effects are below the current sensitivity for $1/R$ above $400 \div 500$ GeV [38]. Let me therefore concentrate on the effects produced by supersymmetric operators localized on the boundaries.

The coefficients of these operators, all with negative dimension in mass, are pure parameters, as would be pure parameters the coefficients of the higher dimensional operators added to the SM Lagrangian. The most important difference among the two cases is that the SM Lagrangian is perturbative up to very high energy whereas the physics

[6]When the third generation is localized, this is a source of flavor violation, mostly through a difference between the coupling of the KK gluons to the first two generations and to the third [39]. If one assumes mixing angles and phases of the down-quark Yukawa coupling matrix comparable to those of the Cabibbo-Kobayashi-Maskawa matrix, the strongest bound arises from the CP-violating ϵ-parameter in K physics and is about $1/R > 1.5$ TeV.

described by $\mathcal{L}_5$ in eq. (13) is not. As a consequence, while it is consistent to take arbitrarily small coefficients for the higher dimensional operators in the SM, this is not naturally the case for the theory under consideration. The operators allowed by the symmetries would at least be generated by quantum corrections, with maximal strenght of their coefficients, at the scale where perturbation theory breaks down. Therefore I consider the operators in the sector where perturbation theory breaks down first, at the lowest energy, i.e. the operators generated by loops involving the top Yukawa coupling, and I require that their coefficients, as estimated by means of naive dimensional analysis adapted to 5D, saturate perturbation theory at the scale Λ. A theory constructed in this way is trustable and predictive, although only at energies sufficiently lower than Λ.

With this assumption, the operators of interest to the EWPT are (the supersymmetric extension of) the Higgs kinetic term and (of) the operator O_H in Table 1, both localized at $y = 0$,

$$\delta\mathcal{L}_4 = Z_H|D_\mu H|^2 + C_H|H^+ D_\mu H|^2 + \dots \tag{23}$$

where the dots stand for their supersymmetric completion. Note that H in eq. (23), although taken at $y = 0$, is a 5D field with a canonically normalized kinetic term in $\mathcal{L}_5$. As such, since H has mass dimension $3/2$, Z_H and C_H have dimension -1 and -4 respectively. From naive dimensional analysis, taking into account the angular integration factors

$$\int \frac{d^4p}{(2\pi)^4} = \frac{1}{16\pi^2}\int p^3 dp, \quad \int \frac{d^5p}{(2\pi)^5} = \frac{1}{24\pi^3}\int p^4 dp \tag{24}$$

in 4D and 5D respectively, one has

$$Z_H \sim \frac{24\pi^3}{16\pi^2}\frac{1}{\Lambda}, \quad C_H \sim \frac{(24\pi^3)^2}{16\pi^2}\frac{1}{\Lambda^4}. \tag{25}$$

What is the impact of these operators on the EWPT? The answer is immediate for O_H, being

$$H = \frac{1}{\sqrt{2\pi R}}h^0 + \dots, \tag{26}$$

where h^0 is the standard Higgs field, canonically normalized in 4D. In this way, from the 5D coefficient C_H, one obtains the 4D coefficient $c_H/\Lambda^2 = C_H/(2\pi R)^2$ in eq. (3) and therefore, from eq. (5)

$$\Delta\epsilon_1(O_H) \sim \frac{9\pi^2}{(\Lambda R)^2}\frac{v^2}{\Lambda^2}. \tag{27}$$

The effect of Z_H on the EWPT is more involved since it gives rise both to a mixing of the W and the Z with their KK modes and to a nonzero vev of the scalar partner of the W, i.e. an $SU(2)$ triplet, in the 5D vector supermultiplet (generically denoted by ϕ_Σ in Table 2). The net overall effect on the EWPT of this term is again only a contribution to the parameter ϵ_1, which can be read from Ref. [39],(but note the different definition of R)

$$\Delta\epsilon_1(Z_H) \sim z_H^2 \frac{\pi^2}{12} m_Z^2 R^2 \sim \frac{3\pi^2}{64} \frac{m_Z^2}{\Lambda^2}, \tag{28}$$

valid for small $z_H = Z_H/(2\pi R)$.

Where is the cut-off?. To determine the numerical significance of these effects on the EWPT, but not only for this reason, one needs to know where the cut-off is, or rather its connection with the compactification scale $1/R$. As already said, this is related to the energy scale at which perturbation theory breaks down. In the case of a 5D gauge coupling, g_5, taking into account the angular integration factor as before, perturbation theory is approximately lost at $\Lambda_g \sim 24\pi^3/g_5^2$, i.e., from the relation between the 4D and the 5D couplings $g_4^2 = g_5^2/2\pi R$,

$$\Lambda_g R \sim \frac{12\pi^2}{g_4^2}. \tag{29}$$

Even in the strong gauge sector therefore, Λ_g is higher than $1/R$ by about two orders of magnitude. When the top quark is non localized, however, perturbation theory in its Yukawa coupling at the $y = 0$ boundary

$$\delta\mathcal{L}_{4,top} = \lambda_{t,5}\, QuH \tag{30}$$

gets saturated at a lower scale Λ, since, from naive dimensional analysis,

$$\lambda_{t,5} \sim \frac{(24\pi^3)^{3/2}}{16\pi^2} \frac{1}{\Lambda^{3/2}} \tag{31}$$

and $\lambda_t = \lambda_{t,5}/(2\pi R)^{3/2}$, so that

$$\Lambda R \sim \frac{4.5}{\lambda_t^{2/3}} \sim 4.5. \tag{32}$$

With the progressive localization of the top, ΛR increases up to $\Lambda R \sim 12\pi^2/\lambda_t^2$ for an exactly localized top.

We can now get back to the estimate of the effects on the EWPT of $\Delta\epsilon_1(Z_H)$ and $\Delta\epsilon_1(O_H)$. We have

$$\Delta\epsilon_1(Z_H) \sim 2 \cdot 10^{-4} (\frac{4.5}{\Lambda R})^2 (R\,TeV)^2 \tag{33}$$

$1/R$	360 ± 70
h	133 ± 10
$\tilde{t}_1$, $\tilde{u}_1$	210 ± 20
$\chi^{\pm}$, χ^0, $\tilde{g}$, $\tilde{q}$, $\tilde{l}$	360 ± 70
$\tilde{t}_2$, $\tilde{u}_2$	540 ± 30
A_1, q_1, l_1, h_1	720 ± 140

Table 3. The particle spectrum and $1/R$ in the CSM. All entries are in GeV.

and

$$\Delta\epsilon_1(O_H) \sim 6 \cdot 10^{-3}(\frac{4.5}{\Lambda R})^4 (RTeV)^2. \tag{34}$$

Note incidentally that $z_H = Z_H/(2\pi R)$ is indeed small, since it is

$$z_H \sim 0.15(\frac{4.5}{\Lambda R}). \tag{35}$$

If we now require $|\Delta\epsilon_1| < 2 \cdot 10^{-3}$, the estimate in eq. (33) allows a low $1/R$, whereas eq. (34), taken at face value, wants $1/R$ above 1.5 TeV. Note in eq. (34), however, the high power of ΛR, which makes the estimate highly uncertain. Note also that $1/R$ increases from the 400 GeV range of the CSM by a progressive localization of the top, which, as we said, rapidly increases ΛR and therefore suppresses $\Delta\epsilon_1(O_H)$.

From all this I conclude that values of $1/R$ of about 1 TeV are compatible with the EWPT. Lower values, such as those required by the CSM, seem to need, at the present state of knowledge, an adjustment in $\Delta\epsilon_1$ by about one order of magnitude.

2.8 Spectrum and phenomenology

The most characteristic feature of the spectrum of the superpartners is the relative heaviness of gauginos and higgsinos, all approximately degenerate at about $1/R$. The Lightest Supersymmetric Particle (LSP) is therefore a sfermion. In the CSM the LSP is a stop. The relevant spectrum of the CSM is shown in Table 3 with an estimate of the uncertainties. The lightest stop is stable or quasi stable if a relevant $U(1)_R$ symmetry is slightly broken.

Particle localization alters the spectrum of sfermions but does not change the fact that one of them is the LSP. The masses of the sfermions of charge Q and hypercharge Y are given by

$$m^2(Q,Y) = m^2_{\rm tree} + m^2_{\rm rad}(Q,Y) + Ym_Z^2 - Qm_W^2 \tag{36}$$

	$\widetilde{Q}$	$\widetilde{u}$	$\widetilde{d}$	$\widetilde{L}$	$\widetilde{e}$
$R\,m_{\rm rad}$	0.24	0.24	0.20	0.09	0.05

Table 4. Radiative masses in units of $1/R$ for the different sfermions. $\widetilde{Q} = (\widetilde{u}_L, \widetilde{d}_L)$, $\widetilde{L} = (\widetilde{\nu}_L, \widetilde{e}_L)$.

where $m_{\rm tree}$ is the tree level mass, including the Yukawa contribution, and $m_{\rm rad}$ is the one loop contribution, as in eq. (21). Both $m_{\rm tree}$ and $m_{\rm rad}$ depend upon the corresponding localization parameter MR [30, 35]. For $MR = 0$, $m_{\rm tree} = 1/R$, up to the Yukawa term which is important for the stops. For $MR > 1$, $m_{\rm rad}$ dominates and rapidly approaches the localized limit given in Table 4.

Who is the LSP depends therefore upon the localization parameters of the different multiplets. Exact cancellation of the FI term (See App. B) is guaranted by taking equal localization parameters for the quarks within one generation, M_q, and, independently, equal localization parameters for the leptons, M_l. Remember that these masses are unrenormalized parameters, so it make sense to set them equal. Suppose that only the third generation quarks are partly localized. Then the LSP is a stop for low $1/R$ and a sbottom for higher $1/R$, due to the difference in $m_{\rm rad}$ ($m_{\rm rad}(\widetilde{d}) < m_{\rm rad}(\widetilde{Q}), m_{\rm rad}(\widetilde{u})$). The LSP may also be a slepton for a sizable $M_l R$. The difference in $m_{\rm rad}$ ($m_{\rm rad}(\widetilde{e}) < m_{\rm rad}(\widetilde{L})$)) makes it definitely more likely to be a charged slepton (in spite of the D-term effects in eq. (36)).

A stable or metastable charged LSP is a striking phenomenological feature. For definiteness, take it to be a stop (but a sbottom would not make any difference in the experimental signal). By picking up a quark, once pair produced in a high energy collision, it would make any of the two super-hadrons $T^+ = \widetilde{t}\overline{d}$ or $T^0 = \widetilde{t}\overline{u}$ (and their charge conjugates $T^-, \overline{T}^0$), which should be detectable as stable particles, since their possible decay into one another is slow enough to let them both cross the detector. $T^\pm$ could appear as a stiff charge track with little hadron calorimeter activity, hitting the muon chambers and distinguishable from a muon via $\mathrm{d}E/\mathrm{d}x$ and time-of-flight. The neutral states, on the contrary, could be identified as missing energy since they would traverse the detector with little interaction [40].

3. Summary and conclusions

The mechanism of EWSB is one of the greatest mysteries in particle physics. Not the only one, but the one with the greatest chances of

being clarified by the experimentation at the LHC. For the time being, the SM provides a successful phenomenological description of EWSB in terms of a Higgs doublet with a mexican hat potential. This is in fact an understatement. The experiments of the ninethies have shown that there must be some fundamental truth in the Higgs description of EWSB. A Higgs boson is very likely to exist. It is likely to be a weakly interacting, narrow state. It is pretty likely, although not certain, that its mass is in the $100 \div 200$ GeV range. The LHC will tell if this is true or not.

As I said in Sect. , this goal of the LHC, although clearly important, is not, however, the very reason for expecting that the LHC will clarify the mechanism and the dynamics of EWSB. There are two aspects to this statement. First there is the uneasiness with the fact that the curvature at zero field of the Higgs potential, crucially negative, is just an input to the theory without any deeper understanding. Discovering the Higgs would not shed light on this problem. Second, there is the extreme ultraviolet sensitivity of the curvature of the potential: in the SM the Higgs mass depends on the "Higgs mass spectral function" Δm_H^2, defined in Sect. 1.2, which is rapidly growing with energy and contributes to the Higgs mass mostly where it cannot be trusted. While the first aspect is generic, the second one is the basis for thinking that the LHC will reveal new phenomena related to EWSB and not included in the SM. At the LHC we should see what dumps Δm_H^2 at energies well within the range explored by the machine. In this way we might also understand what triggers EWSB by giving the Higgs a negative mass squared.

If this view is right and we are not misguided in a way or another, we are faced with a problem. Why haven't we seen yet any manifestation of this dumping mechanism anywhere in direct or indirect experiments? The straigthest interpretation of the data, mostly from the EWPT, points to a significant gap between the Higgs mass and the effective scale that parametrizes a generic deviation from the SM in terms of higher dimensional operators, taken maximally symmetric. This is the "little hierarchy"problem. More precisely the tension is between this scale and the energy at which the dumping mechanism advocated above should become operative. Finding a theory that dissolves this tension is to solve the "little hierarchy"problem. As I have tried to make it clear, there are assumptions (and "judgements") in the line of reasoning that leads to the formulation and the very existence of this problem. To the point that I am sometimes asked why do I care about it at all. To me the answer looks obvious. If these assumptions are valid, the expectation that the LHC will clarify the mechanism and the dynamics of EWSB

is well justified. Since they are also reasonable, I prefer to take them seriously.

At least in part, this problem is at the basis of the revival of theoretical interest in EWSB. Without pretending to be exhaustive, I have summarized two attempts at addressing this problem, the MSSM and the "little Higgs", and I have described more at lenght a third one, based on supersymmetry breaking by extra dimensions. Needless to say, there is a great physical difference between the first and the two other cases. While the first can be extended up to very high energy, like the GUT or even the Planck scale, this in not true for the little Higgs or the extradimensional theories. Quite on the contrary, the gap between the dumping scale of Δm_H^2 and the cut-off of little Higgs and extradimensional models is not more than a decade or so. Is this a step backward? Conceptually it may look to be the case. Incidentally this motivates to search for possible "ultraviolet completions" of these models. On the issue, however, I prefer to keep an open mind and leave it to the LHC experiments to decide what is relevant and what is not. From a theoretical point of view, I find that looking for neat solutions of the "little hierarchy" problem is a well motivated (and difficult) task. I suspect that the search will continue.

4. Gauge anomalies

A gauge theory of fermions in 5D is vector-like. Hence no gauge anomaly is expected. To make contact with phenomenology, however, chiral fermionic zero modes are needed, which is obtained by suitable boundary conditions (BC) on the segment associated with the extra dimension[41, 42]. In turn, this can generate anomalies localized on the boundaries.

In Sect. 2.2 I have considered two types of BC for the fermions in a 5D supermultiplet:

- Matter-like BC: $\Psi(+,+)$ and $\Psi^c(-,-)$;

- Higgsino-like BC: $\Psi(+,-)$ and $\Psi^c(-,+)$.

In the first case, equal FI terms are generated in general at the two boundaries. The condition for their vanishing is equivalent to the condition $TrY = 0$, restricted to the fermionic zero modes in the associated supermultiplets. Again, this condition is fulfilled by the fermions of the SM. In the case of Higgs-like BC, the FI terms at the two boundaries are equal in magnitude and opposite in sign [32, 45]. As such, they are of a particular nature, only possible in 5D. With a 4D superfield notation,

their contribution to the action may be written as

$$\delta S_{FI} = \xi \int d^4x \int_0^L dy \int d^4\theta (\delta_5 V - \Phi - \bar{\Phi}), \tag{37}$$

where V is the N=1 vector multiplet and Φ is the chiral multiplet in the same 5D hypermultiplet. Eq. (37) and the gauge transformations of the various fields make explicit the gauge invariance of the integrand factor $(\delta_5 V - \Phi - \bar{\Phi})$, unlike what happens for a 4D FI term, where the θ-integration is essential to achieve gauge invariance. This is the basis for the property of a FI term like in eq. (37) of maintaining both gauge invariance and supersymmetry in the vacuum [45].

A sufficient condition for avoiding a divergent FI term in the case of 5D multiplets with Higgs-like BC is to have $TrY = 0$ for the $(+,+)$ components of the 5D supermultiplets. In this case even a finite FI term is absent at all if the 5D multiplets have equal bulk masses.

Acknowledgements

I thank Lawrence Hall, Yasunori Nomura, Riccardo Rattazzi, Alessandro Strumia, Roberto Contino, Michele Papucci, Guido Marandella, Giacomo Cacciapaglia, Marco Cirelli, Paolo Creminelli, Claudio Scrucca, Takemichi Okui, Steven Oliver for many useful discussions. I am indebted to Alessandro Strumia for providing me with Fig. 1 and Table 1. It is a pleasure to thank the organizers of this School, especially Dimitri Kazakov and Gerard Smadja. This work has been partially supported by MIUR and by the EU under TMR contract HPRN-CT-2000-00148.

References

[1] LEP Higgs Working Group, LHWG Note 2001-04.

[2] The ElectroWeak Working Group, http:// lepewwg.web.cern.ch/LEPEWWG/

[3] M. E. Peskin and T. Takeuchi, Phys. Rev. Lett. **65** (1990) 964 ; M. E. Peskin and T. Takeuchi, Phys. Rev. **D46** (1992) 381.

[4] G. Altarelli and R. Barbieri, Phys. Lett. **B253** (1991) 161 ; G. Altarelli, R. Barbieri and S. Jadach, Nucl. Phys. **B369** (1992) 3. Erratum-ibid.**B376** (1992) 444.

[5] R. Barbieri and A. Strumia, Phys.Lett. **B462** (1999) 144.

[6] L. J. Hall and C. F. Kolda, Phys. Lett. **B459** (1999) 213.

[7] R. Barbieri, talk given at "Frontiers Beyond the Standard Model", Minneapolis, October 10-12, 2002

[8] R. Barbieri and G. F. Giudice, Nucl. Phys. **B306** (1988) 63.

[9] G. Kane , B. Nelson , Lian-Tao Wang and T. Wang, hep-ph/0304134; J. Casas, J. Espinosa and I. Hidalgo, hep-ph/0310137.

[10] R. Barbieri, S. Ferrara and C. Savoy, Phys.Lett. **B119** (1982) 343; A. Chamseddine, R. Arnowitt and P. Nath, Phys. Rev. Lett. **49** (1982) 970 ; L. Hall, J. Lykken and S. Weinberg, Phys. Rev. **D27** (1983) 2359.

[11] M. Dine and W. Fischler, Nucl. Phys. **B204** (1982) 346 ; L. Alvarez-Gaume, M. Claudson and M. B. Wise, Nucl. Phys. **B207** (1982) 96 ; M. Dine and A. E. Nelson, Phys. Rev. **D48** (1993) 1277.

[12] L. Giusti, A. Romanino and A. Strumia, Nucl. Phys. **B550** (1999) 3.

[13] G. F. Giudice, hep-ph/0311344.

[14] H. E. Haber and M. Sher, Phys. Rev. **D35** (1987) 2206.

[15] P. Batra, A. Delgado, D. E. Kaplan and T. M. P. Tait, hep-ph/0309149.

[16] J. Casas, J. Espinosa and I. Hidalgo, hep-ph/0310137.

[17] H. Georgi and A. Pais, Phys. Rev. D **D10** (1974) 539 ; D. Kaplan and H. Georgi, Phys.Lett. **B136** (1984) 183.

[18] N. Arkani-Hamed, A. Cohen and H. Georgi, Phys.Lett. **B513** (2001) 232 ; N. Arkani-Hamed, A. Cohen and E. Katz and A. Nelson, JHEP **07** (2002) 034 ; N. Arkani-Hamed, A. G. Cohen, E. Katz, A. E. Nelson, T. Gregoire and J. G. Wacker, JHEP **0208** (2002) 021.

[19] C. Csaki, J. Hubisz, G. D. Kribs, P. Meade and J. Terning, Phys. Rev. **D67** (2003) 115002 ; J. L. Hewett, F. J. Petriello and T. G. Rizzo, JHEP **0310** (2003) 062 ; C. Csaki, J. Hubisz, G. D. Kribs, P. Meade and J. Terning, Phys. Rev. **D68** (2003) 035009.

[20] M. Perelstein, M. E. Peskin and A. Pierce, hep-ph/0310039.

[21] G. Burdman, M. Perelstein and A. Pierce, Phys. Rev. Lett. **90** (2003) 241802.

[22] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. **B429** (1998) 263; I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. **B436** (1998) 257.

[23] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**(1999) 3370.

[24] I. Antoniadis, S. Dimopoulos, A. Pomarol and M. Quiros, Nucl. Phys. **B544** 503 (1999).

[25] A. Delgado, A. Pomarol, M. Quiros, Phys. Rev. **D60** (1999) 095008.

[26] R. Barbieri, L. J. Hall and Y. Nomura, Phys. Rev. **D63** (2001) 105007.

[27] C. Csaki, C. Grojean, H. Murayama, L. Pilo and J. Terning, hep-ph/0305237; C. Csaki, C. Grojean, L. Pilo and J. Terning, hep-ph/0308038; R. Sekhar Chivukula, D. A. Dicus and H. J. He, Phys. Lett. **B525** (2002) 175.

[28] R. Barbieri, A. Pomarol and R. Rattazzi, hep-ph/0310285.

[29] J. Scherk and J. H. Schwarz, Phys. Lett. **B82** (1979) 60 ; Nucl. Phys. **B153** (1979) 61.

[30] R. Barbieri, G. Marandella and M. Papucci, Phys. Rev. **D66** (2002) 095003.

[31] N. Arkani-Hamed, L. Hall, Y. Nomura, D. Smith and N. Weiner, Nucl. Phys. **B605** (2001) 81.

[32] D. M. Ghilencea, S. Groot Nibbelink and H. P. Nilles, Nucl. Phys. **B619** (2001) 385.

[33] D. Marti and A. Pomarol, Phys. Rev. **D66** (2002) 125005.

[34] R. Barbieri, L. J. Hall, G. Marandella, Y. Nomura, T. Okui, S. J. Oliver and M. Papucci, Nucl. Phys. **B663** (2003) 141.

[35] R. Barbieri, G. Marandella and M. Papucci, Nucl. Phys. **B668** (2003) 273.

[36] R. Barbieri, L. J. Hall and Y. Nomura, hep-ph/0110102.

[37] G. Marandella and M. Papucci, work in progress.

[38] R. Barbieri, G. Cacciapaglia and A. Romito, Nucl. Phys. **B627** (2002) 95 ; G. Cacciapaglia, M. Cirelli and G. Cristadoro, Nucl. Phys. **B634** (2002) 230.

[39] A. Delgado, A. Pomarol and M. Quiros, JHEP **0001** (2000) 030.

[40] M. B. Chertok, G. D. Kribs, Y. Nomura, W. Orejudos, B. Schumm and S. Su, in *Proc. of the APS/DPF/DPB Summer Study on the Future of Particle Physics (Snowmass 2001)* ed. N. Graf, eConf **C010630** (2001) 310.

[41] L. J. Dixon, J. A. Harvey, C. Vafa and E. Witten, Nucl. Phys. **B261**(1985) 678.

[42] L. J. Dixon, J. A. Harvey, C. Vafa and E. Witten, Nucl. Phys. **B274** (1986) 285.

[43] N. Arkani-Hamed, A. G. Cohen and H. Georgi, Phys. Lett. **B516**(2001) 395.

[44] C. A. Scrucca, M. Serone, L. Silvestrini and F. Zwirner, Phys. Lett. **B525** (2002) 169.

[45] R. Barbieri, R. Contino, P. Creminelli, R. Rattazzi and C. A. Scrucca, Phys. Rev. **D66**(2002) 024025.

CURRENT STATUS OF THE CKM MATRIX AND THE CP VIOLATION

Achille Stocchi
Laboratoire de l'Accélérateur Linéaire, IN2P3-CNRS et Université de Paris-Sud, BP 34, F-91898 Orsay Cedex
stocchi@lal.in2p3.fr

Abstract

These lectures give an introduction and the current status of flavour physics in the quark sector, with special attention to the CKM matrix and CP violation. We describe the measurements which contribute to the determination of the CKM matrix elements and how, together with important theoretical developments, they have significantly improved our knowledge on the flavour sector of the Standard Model. These lectures are complemented by the seminar of U. Mallik (see these proceedings) which describes in more details the most recent CP-violating related measurements by the B-factories.
The results presented are up-to-date till winter 2004.

Keywords: CKM matrix, CP violation, Beauty (B) hadrons, B decays, Unitarity Triangle

1. Introduction

Accurate studies of the production and decay properties of beauty and charm hadrons are exploiting a unique laboratory for testing the Standard Model in the fermion sector, for studying QCD in the non-perturbative regime and for searching for New Physics through virtual processes.

In the Standard Model, weak interactions among quarks are encoded in a 3×3 unitary matrix: the CKM matrix. The existence of this matrix conveys the fact that quarks, in weak interactions, act as linear combinations of mass eigenstates [1, 2].

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 31–82.

The CKM matrix can be parametrized in terms of four free parameters which are measured in several physics processes.

In a frequently used parametrization, these parameters are named: λ, A, $\overline{\rho}$ and $\overline{\eta}$[1]. The Standard Model predicts relations between the different processes which depend upon these parameters. CP violation is accommodated in the CKM matrix and its existence is related to $\overline{\eta} \neq 0$. The unitarity of the CKM matrix can be visualized as a triangle in the $(\overline{\rho}, \overline{\eta})$ plane. Several quantities, depending upon $\overline{\rho}$ and $\overline{\eta}$ can be measured and they must define compatible values for the two parameters, if the Standard Model is the correct description of these phenomena. Extensions of the Standard Model can provide different predictions for the position of the apex of the triangle, given by the $\overline{\rho}$ and $\overline{\eta}$ coordinates. The most precise determination of these parameters is obtained using B decays, $B^0 - \overline{B^0}$ oscillations and CP asymmetry in the B and in the K sectors.

Many additional measurements of B meson properties (mass, branching fractions, lifetimes...) are necessary to constrain the Heavy Quark theories [Operator Product Expansion (OPE) /Heavy Quark Effective Theory (HQET) /Lattice QCD (LQCD)] to allow for precise extraction of the CKM parameters. In addition, to be able to extract the Standard Model parameters, it is also necessary to control and measure the backgrounds, and to acquire a detailed understanding of the experimental apparatus. All these aspects are important because they propagate as systematic errors attached to the values extracted for the CKM parameters. For instance, the values and the uncertainties of the B hadron lifetimes enter in many important quantities. Experimental progress in various B physics measurements has been crucial in the determination of the CKM matrix elements. These last aspects are not treated in these lectures.

2. Short story: from strangeness to the CKM Matrix

The discovery of the strange particle in 1947 was totally unexpected and can be seen as the beginning of a new era in particle physics which has not ended yet. Just after the pion discovery by C.M.G. Lattes, H. Muirhead, C.F. Powell and G.P. Occhialini [4], in 1947, the same year, C.C. Butler and G.D. Rochester [5] reveal, having exposed a cloud chamber to cosmic rays, the existence of a still-heavier unstable particle decaying in a typical V-topology; this earlier name could be ascribed

[1] $\overline{\rho} = \rho(1 - \frac{\lambda^2}{2})$; $\overline{\eta} = \eta(1 - \frac{\lambda^2}{2})$[3].

to the characteristic topology of the tracks which were produced when a neutral particle decays into two charged particles. At the same time there were also events in which a charged particle trajectory had a sharp break indicating a decay ($V^{\pm}$) (corresponding to the decay $K^+ \to \mu^+\nu_\mu$). In fact, the first example of such particles was reported by L. Leprince-Ringuet and M. Lhéritier in 1944 [6]. They observed a secondary cosmic ray particle, in a cloud chamber placed at the Laboratoire de l'Argentière (Hautes-Alpes), producing a recoil electron (energetic delta ray). From the measured curvatures of the ongoing and outgoing particles and using the value of the scattering angle of the electron it was possible to determine the mass of the incident particle which was found to be of 495 $\pm$ 60 MeV/c^2. It is today clear that this particle is the charged Kaon, nevertheless this discovery came too early, since even the pion was not discovered at that time!

It took two years to confirm the result of Rochester and Butler. These experiments were continued at higher altitude and with high degree of precision [7]. The results unambiguously established the existence of two states: $\Lambda \to p\pi^-$ and $K^0 \to \pi^+\pi^-$. In 1953 it became possible to produce those V-particles in accelerators[2] and it was then clear that they were produced in reactions mediated by the strong interaction; furthermore those particles were always produced in pairs (*associated production*). On the other hand their typical lifetime was of about 10^{-10} s which is a typical time scale of the weak interaction[3]. These particles are then "strange", as they are produced through the strong interaction whereas they decay through weak interaction processes. The solution was proposed, after several unfruitful tentative, by M. Gell-Mann [8], to introduce a new additive quantum number: the *strangeness*[4]. The strangeness was assigned to be -1 for the Λ, the $\overline{\mathrm{K}}^0$ and the K^- (and +1 for the corresponding antiparticles), -2 for the Ξ^- and 0 for all non-strange particles and making the hypothesis that this new quantum number is conserved by strong and electromagnetic interactions and is not conserved by the weak interaction. This allows

[2]The Brookhaven 3 GeV Cosmotron was the first accelerator delivering strange particles, followed by the Berkeley 6 GeV Bevatron.

[3]The scattering cross section of events like $\pi^- p \to K^0\Lambda$ corresponds to the geometrical cross section of hadrons ($\simeq 10^{-13}$ cm)^2 which indicates that the Λ and the K^0 are produced through strong interactions. The natural lifetime of the strong interaction can be estimated using the relationship $\tau_{st} = (had.radius)/c \simeq 10^{-23}s$.

[4]The observation of events such as $\Xi^- \to \Lambda\pi^-$, the so-called cascade events, and the non-observation of events such as $\Xi^- \to n\pi^-$, closed up the option that the *strangeness* could be a multiplicative quantum number (a kind of "strange parity") being +1 for strange particles and -1 for non-strange ones. It was in fact the first indication of the existence of double strange particles.

to "explain", a posteriori, why strange particles are always produced in pairs (by strong interactions ΔS=0) and have a relatively long lifetime (decay through weak interactions ΔS=1).

In the decay, the strangeness changes by one unit and these transitions were classified as $\Delta S = 1$. An intense experimental activity on strange particles shown, in the fifties, that the absolute decay rate for these transitions was suppressed by a factor of about 20 as compared with the corresponding rate measured in $\Delta S = 0$ decays.

In the early 60's the existence of new constituents of matter was postulated: they were called quarks. They were independently introduced by M. Gell-Mann[9] and G. Zweig [10] and they should transform according to the fundamental representation of SU(3). They were supposed to have spin 1/2 and to exist in three varieties: the quark u with charge +2/3, the quarks d and s with charge -1/3. By analogy with leptons it was suggested that the quarks were also organized into doublets and the existence of a new quark of charge 2/3 was proposed [11].

In 1963, N. Cabibbo proposed [1] a model to account for the suppression of ΔS=1 transitions. In this model the d and s quarks, involved in weak processes, are rotated by a mixing angle θ_c: the Cabibbo angle. The quarks are organized in a doublet:

$$\begin{pmatrix} u \\ d_c \end{pmatrix} = \begin{pmatrix} u \\ d\ cos\theta_c + s\ sin\theta_c \end{pmatrix} \tag{1}$$

the small value of sin θ_c ($\simeq 0.22$) is responsible for the suppression of strange particle decays (the coupling being proportional to $sin^2\theta_c$). In this picture the slight suppression of $n \to p\ e^-\overline{\nu_e}$ with respect to the rate of $\mu^- \to e^-\nu_\mu\overline{\nu_e}$ is also explained by the fact that the coupling in the neutron decay is proportional to $cos^2\theta_c$.

In this model, the neutral current coupling can be written[5]:

$$u\overline{u} + d\overline{d}\ cos^2\theta_c + s\overline{s}\ sin^2\theta_c + (s\overline{d} + d\overline{s})\ cos\theta_c sin\theta_c. \tag{2}$$

The presence of the $(s\overline{d} + d\overline{s})$ term implies the existence of a flavour changing neutral current (FCNC). This was a serious problem for the Cabibbo model, since these couplings would produce contributions to Δm_K and $K_L \to \mu^+\mu^-$ decays which are larger by several order of magnitude.

[5]More formally. The charged currents are described by the operators $J_\mu^{+(-)}$. The existence of the neutral current is needed to complete the group algebra (obtained by commuting the operators J_μ^+ and J_μ^-) and necessarily contains ΔS=± 1 terms.

In 1970 S. Glashow, J. Iliopoulos and L. Maiani [12] (GIM) proposed the introduction of a new quark, named c, of charge 2/3 and the introduction of a new doublet of quarks formed by the c quark and by a combination of the s and d quarks orthogonal to d_c (eq. 1):

$$\begin{pmatrix} c \\ s_c \end{pmatrix} = \begin{pmatrix} c \\ s\ cos\theta_c - d\ sin\theta_c \end{pmatrix}. \tag{3}$$

In this way the $(s\overline{d} + d\overline{s})$ term (in Eq. 2), in the neutral current, is cancelled.

The discovery of the charm quark in the form of $c\overline{c}$ bound states [13] and the observation of charmed particles decaying into strange particles [14] (the $c\overline{s}$ transitions which are proportional to $cos^2\theta_c$ dominate over the $c\overline{d}$ transitions which are proportional to $sin^2\theta_c$) represent a tremendous triumph of this picture.

It should be reminded that a candidate event for the decay of a charm hadron was first observed in 1971, in Japan, in an emulsion detector exposed to cosmic rays [15]: $X^{\pm} \rightarrow h^{\pm}\pi^0$. The lifetime of $h^{\pm}$ and its mass were found to be 10^{-14} sec and 1.8 GeV respectively ! (see [16] and [17] for more details).

The charge current, mediated by the emission of a W boson, can then be written:

$$(\overline{uc})\gamma^{\mu}(1-\gamma_5)V \begin{pmatrix} d \\ s \end{pmatrix} \tag{4}$$

where $\gamma^{\mu}(1-\gamma_5)$ is the $V-A$ current, which accounts also for parity violation, u, d, s, c are the mass eigenstates and V is defined as:

$$V = \begin{pmatrix} cos\theta_c & sin\theta_c \\ -sin\theta_c & cos\theta_c \end{pmatrix}. \tag{5}$$

V is the Cabibbo unitary matrix which specifies the quark states which are involved in weak interactions. In 1975 the Mark I group at SPEAR discovered the third charged lepton: the τ [18]. Two years later the fifth quark, the b, was found at FNAL [19]. The indirect existence for the top quark t from the observation of $B^0_d - \overline{B^0_d}$ oscillations [20] suggested the existence of an heavier version of the doublets (u,d) and (c,s)[6]. The t quark was finally discovered in 1995 at Fermilab [21] in $p\overline{p}$ collisions.

The existence of three quark doublets was already proposed by M. Kobayashi and K. Maskawa in 1973 [2] as a possible explanation for CP

[6]Another indirect piece of evidence for the existence of top was the measurement of the Z^0 coupling to a $b\overline{b}$ pairs, which shown that the b quark is a member of an doublet with partner that cannot be a u or a c quark.

violation. Their proposal is a generalization of the Cabibbo rotation and implies that the weak flavour changing transitions are described by a 3×3 unitary matrix:

$$\begin{pmatrix} u \\ c \\ t \end{pmatrix} \rightarrow V \begin{pmatrix} d \\ s \\ b \end{pmatrix} , \quad V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} . \qquad (6)$$

This matrix conveys the fact that there is an arbitrary rotation, usually applied to the -1/3 charged quarks, which is due to the mismatch between the strong and the weak eigenstates. This matrix can be parametrized using three real parameters and one phase which cannot be removed by redefining the quark field phases. This phase leads to the violation of the CP symmetry. In fact since CPT is a good symmetry for all quantum field theories, the complexity of the Hamiltonian implies that the time reversal invariance T and thus CP is violated[7]. In this picture the Standard Model includes CP violation in a simple way.

3. The Standard Model in the fermion sector and the CKM matrix

The Standard Model is based on the $SU(2)_L \times U(1)_Y$ gauge symmetry, where the index L stands for left, since only the $left$ handed particles are implied in the charged weak current. In a Lagrangian the mass term of a fermionic field is of the type[8]

$$m\overline{\psi}\psi = m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R). \qquad (7)$$

Thus the mass implies a $left - right$ coupling[9] which is not gauge invariant. An economical approach for introducing fermion masses, in

[7]The time is an anti-linear operator: $T(\lambda_1|\psi_1 > +\lambda_2|\psi_2 >) = \lambda_1^* < T\psi_1| + \lambda_2^* < T\psi_2|$. It can be simply understood, recalling that $\psi(x,t)$ and $\psi^*(x,-t)$ (and not $\psi(x,-t)$) obey to the same Schrodinger equation. If the operator T is applied to the Standard Model Lagrangian and thus to the CKM matrix: $T\ V(CKM)|..>= V^*(CKM)\ < T...|$. If $V(CKM)$ is complex, $V(CKM) \neq V(CKM)^*$. In this case the Hamiltonian does not commute with T, thus T is not conserved and, since CPT is conserved, CP is violated.

[8]The Euler-Lagrange equation implies the following correspondence between motion equations and Lagrangians:
$\mathcal{L} = \partial^\mu\psi\delta_\mu\psi - m^2\psi^2 \rightarrow (\partial_\mu\partial^\mu + m^2)\psi \rightarrow (E^2 = p^2 + m^2)$ (Einstein equation).
$\mathcal{L} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - m\overline{\psi}\psi \rightarrow (i\gamma^\mu\partial_\mu - m)\psi = 0$ (Dirac equation).

[9]It simply follows from the properties of projection operators and using the equalities $\overline{\psi}_R = \overline{\psi}P_L$ and $\overline{\psi}_L = \overline{\psi}P_R$. It follows that:
$m\overline{\psi}\psi = m\overline{\psi}(P_L + P_R)\psi = m\overline{\psi}(P_LP_L + P_RP_R)\psi = m[(\overline{\psi}P_L)(P_L\psi) + (\overline{\psi}P_R)(P_R\psi)] = m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R)$

a gauge-invariant way, is to consider Yukawa couplings in which contributes a weak iso-doublet field ϕ:

$$\overline{\psi}_L \phi \psi_R \ ; \ \text{with} \qquad \phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \ , \ \ \mathrm{I}_\phi = 1/2, \ \mathrm{Y}_\phi = 1 \ , \tag{8}$$

where the quantum numbers of the new doublet (ϕ) are exactly those needed to restore the gauge invariance of the interaction vertex giving the mass to fermions.

In the Standard Model, the Lagrangian corresponding to charged weak interactions can be written as:

$$\mathcal{L}_W = \frac{g}{2} \overline{Q}^{Int.}_{L_i} \gamma^\mu \sigma^a Q^{Int.}_{L_i} W^a_\mu \ ; \ Q^{Int.}_{L_i} = \begin{pmatrix} u_{L_i} \\ d_{L_i} \end{pmatrix}, L^{Int.}_{L_i} = \begin{pmatrix} \nu_{L_i} \\ \ell_{L_i} \end{pmatrix}, \tag{9}$$

the index $Int.$ indicates the weak interaction basis, σ^a are the Pauli matrices (a=1,2,3), W^a_μ are the SU(2)$_L$ gauge bosons and i is the quark index. It can be noted that: $\overline{Q}^{Int.}_{L_i} Q^{Int.}_{L_i} = \overline{Q}^{Int.}_{L_i} \mathbf{1}_{ij} Q^{Int.}_{L_j}$. The charged weak interactions are family blind (the quantum numbers of the Standard Model are I_3 and Y which do not "feel" the family index).

In the interaction basis the Yukawa interaction is:

$$\mathcal{L}_Y = Y^d_{ij} \overline{Q}^{Int.}_{L_i} \phi d^{Int.}_{R_j} + Y^u_{ij} \overline{Q}^{Int.}_{L_i} \tilde{\phi} u^{Int.}_{R_j} + Y^\ell_{ij} \overline{L}^{Int.}_{L_i} \phi \ell^{Int.}_{R_j} \ + \ H.C. \ , \tag{10}$$

where $\tilde{\phi} = i\sigma_2 \phi^*$. In the most general case the matrices Y_{ij} are complex. The presence of two independent matrices Y_{ij}, for the *up*-type and *down*-type quark, is due to the behaviour of the Yukawa coupling itself.

After spontaneous symmetry breaking (SSB):

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \ , \tag{11}$$

and the Yukawa interaction can be written:

$$\mathcal{L}_M = \overline{d}^{Int.}_{L_i} M^d_{ij} d^{Int.}_{R_j} + \overline{u}^{Int.}_{L_i} M^u_{ij} u^{Int.}_{R_j} + \overline{\ell}^{Int.}_{L_i} M^\ell_{ij} \ell^{Int.}_{R_j} \ + \ H.C. \tag{12}$$

where $M^u = (v/\sqrt{2}) Y^u$ and $M^d = (v^*/\sqrt{2}) Y^d$. Physical masses are obtained by finding transformations of the fields such that the corresponding mass matrices become real and diagonal:

$$M^f(diag) = V^f_L M^f (V^f_R)^\dagger \ . \tag{13}$$

Therefore, the mass eigenstates are

$$\begin{aligned} d_{L_i} = (V^d_L)_{ij} d^{Int.}_{L_j} \ ; \ d_{R_i} = (V^d_R)_{ij} d^{Int.}_{R_j} \\ u_{L_i} = (V^u_L)_{ij} u^{Int.}_{L_j} \ ; \ u_{R_i} = (V^u_R)_{ij} u^{Int.}_{R_j} \\ \ell_{L_i} = (V^\ell_L)_{ij} \ell^{Int.}_{L_j} \ ; \ \ell_{R_i} = (V^\ell_R)_{ij} \ell^{Int.}_{R_j} \\ \nu_{L_i} = (V^\nu_L)_{ij} \nu^{Int.}_{L_j} \end{aligned} \tag{14}$$

In this basis the Lagrangian for the weak interaction can be written as:

$$\mathcal{L}_W = \frac{g}{2}\overline{u}_{L_i}\gamma^\mu \left[V_L^u (V_L^d)^\dagger\right] d_{L_j} W_\mu^a + h.c. \tag{15}$$

where

$$V(CKM) = V_L^u (V_L^d)^\dagger \ . \tag{16}$$

V(CKM) is the CKM matrix. The phenomenon of flavour changing can be appreciated in two different ways (different basis). If we use the basis in which the mass matrices are diagonal, the Lagrangian for the interactions is not anymore family blind. The interaction among quarks belonging to different families are possible and the couplings are encoded in the CKM matrix.

If the same procedure is applied in the lepton sector it follows that:

$$V(leptons) = (V_L^\nu (V_L^\ell)^\dagger) = (V_L^\ell (V_L^\ell)^\dagger) = 1 \ , \tag{17}$$

since the mass matrix of the neutrinos is arbitrary (the neutrinos are massless in the SM), we can always choose $V_L^\nu = V_L^\ell$.
There is freedom in parametrising the CKM matrix:
- a permutation between the different generations (it is normally chosen to order the quarks by increasing value of their mass (u, c, t and d, s, b),
- the presence of phases in the CKM matrix. It is clear that M($diag$) is unchanged if the matrices $V_{L(R)}$ are multiplied by a matrix containing only phases: $\tilde{V}_{L(R)}^f = P^f V_{L(R)}^f$, it follows $V(CKM) = P^u V(CKM) P^{*d}$.

As long as some of these phases are not observable, one has to require that the CKM matrix contains the minimal number of phases, all the others being absorbed in the definition of quark wave functions.

The 2×2 matrix can be used to illustrate the contribution of these arbitrary phases in the CKM matrix:

$$\begin{aligned} V &= \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \rightarrow \begin{pmatrix} e^{-i\phi_1} & 0 \\ 0 & e^{-i\phi_2} \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \times \cdots \\ &\cdots \times \begin{pmatrix} e^{+i\chi_1} & 0 \\ 0 & e^{+i\chi_2} \end{pmatrix} = \begin{pmatrix} V_{11} e^{-i(\phi_1 - \chi_1)} & V_{12} e^{-i(\phi_1 - \chi_2)} \\ V_{21} e^{-i(\phi_2 - \chi_1)} & V_{22} e^{-i(\phi_2 - \chi_2)} \end{pmatrix} \end{aligned} \tag{18}$$

It can be noted that:

$$(\phi_2 - \chi_2) = (\phi_2 - \chi_1) + (\phi_1 - \chi_2) - (\phi_1 - \chi_1) \tag{19}$$

Among the four phases, corresponding to the four quark flavours, only three can be chosen in an arbitrary way, since one phase difference is obtained as a linear sum of the other three. In the general case, the number of arbitrary phases is: $2n(families) - 1$.

The CKM matrix is a rotation matrix and, in a complex plane, can be parametrized in terms of a given number of angles (real numbers) and phases (complex numbers) as indicated in Table 1.

Family of quarks	num. of Angles	num. Phases	Irreducible Phases
n	$n(n-1)/2$	$n(n+1)/2$	$n(n-1)/2-(2n-1)$ $=(n-1)(n-2)/2$
2	1	3	0
3	3	6	1
4	6	10	3

Table 1. Numbers of angles and phases parametrising a complex rotation matrix. The last column gives the number of phases which cannot be reabsorbed into the quark fields.

It results that a 2× 2 matrix (the Cabibbo matrix) is parametrized in terms of one real parameter and contains no phase. The 3× 3 matrix (CKM) is parametrised in terms of three real parameters and one irreducible phase. The presence of this complex number in the Lagrangian, as explained at the end of Section 2 (footnote 7), is responsible, and it is the only one, of the fact that the CP symmetry is violated in the Standard Model.

4. The CKM Matrix

Many parametrizations of the CKM matrix have been proposed in the literature. The most popular are the standard parametrization [22] recommended by [23] and a generalization of the Wolfenstein parametrization [24] as presented in [3].

With $c_{ij}=\cos\theta_{ij}$ and $s_{ij}=\sin\theta_{ij}$ $(i,j=1,2,3)$, the standard parametrization is given by:

$$V_{\rm CKM}=\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23}-c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23}-s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23}-c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12}-s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1}$$

where δ is the phase necessary for CP violation. c_{ij} and s_{ij} can all be chosen to be positive and δ may vary in the range $0\leq\delta\leq 2\pi$. However, measurements of CP violation in K decays force δ to be in the range $0<\delta<\pi$. s_{13} and s_{23} are small numbers: $\mathcal{O}(10^{-3})$ and $\mathcal{O}(10^{-2})$, respectively. Consequently, from phenomenological applications, the four independent parameters are taken to be:

$$s_{12}=|V_{us}|,\quad s_{13}=|V_{ub}|,\quad s_{23}=|V_{cb}|,\quad \delta. \tag{2}$$

The first three quantities can be extracted from tree level decays mediated by the $s \to u$, $b \to u$ and $b \to c$ transitions, respectively. The phase δ can be obtained from CP violating or loop processes sensitive to the V_{td} matrix element.

The absolute values of the elements of the CKM matrix show a hierarchical pattern with the diagonal elements being close to unity: $|V_{us}|$ and $|V_{cd}|$ being of order 0.2, $|V_{cb}|$ and $|V_{ts}|$ of order $4 \cdot 10^{-2}$ and $|V_{ub}|$ and $|V_{td}|$ of order $5 \cdot 10^{-3}$. The Wolfenstein parametrization is useful to illustrate this structure. It shows that the matrix is almost diagonal, namely that the coupling between quarks of the same family is close to unity, and is decreasing as the separation between families increases:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4). \tag{3}$$

The set (2) is replaced by:

$$\lambda, \quad A, \quad \rho, \quad \eta \ , \tag{4}$$

known as the Wolfenstein parameters. To obtain the exact expression of the CKM parameters in the Wolfenstein parametrization, it is convenient to go back to the standard parametrization and to make the following change of variables in (1) [3, 25]:

$$s_{12} = \lambda \ , \qquad s_{23} = A\lambda^2 \ , \qquad s_{13}e^{-i\delta} = A\lambda^3(\rho - i\eta). \tag{5}$$

At order λ^5, the obtained CKM matrix in the extended Wolfenstein parametrization is:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} - \frac{\lambda^4}{8} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{A^2\lambda^5}{2}(1 - 2\rho) - iA^2\lambda^5\eta & 1 - \frac{\lambda^2}{2} - \lambda^4(\frac{1}{8} + \frac{A^2}{2}) & A\lambda^2 \\ A\lambda^3\left[1 - (1 - \frac{\lambda^2}{2})(\rho + i\eta)\right] & -A\lambda^2(1 - \frac{\lambda^2}{2})\left[1 + \lambda^2(\rho + i\eta)\right] & 1 - \frac{A^2\lambda^4}{2} \end{pmatrix} + \mathcal{O}(\lambda^6). \tag{6}$$

By definition, the expression for V_{ub} remains unchanged relative to the original Wolfenstein parametrization and the corrections to V_{us} and V_{cb} appear only at $\mathcal{O}(\lambda^7)$ and $\mathcal{O}(\lambda^8)$, respectively. The advantage of this generalization of the Wolfenstein parametrization, over other generalizations found in the literature, is the absence of relevant corrections in V_{us}, V_{cd}, V_{ub} and V_{cb}. It can be noted that the element V_{td} can be re-expressed as:

$$V_{td} = A\lambda^3(1 - \overline{\rho} - i\overline{\eta})$$

where [3]

$$\overline{\rho} = \rho(1 - \frac{\lambda^2}{2}), \qquad \overline{\eta} = \eta(1 - \frac{\lambda^2}{2}). \tag{7}$$

This elegant change in V_{td}, with respect to the original Wolfenstein parametrization, allows a simple generalization of the so-called unitarity triangle to higher orders in λ [3] as discussed below.

4.1 The Unitarity Triangle

From the unitarity of the CKM matrix ($VV^\dagger = V^\dagger V = 1$), non diagonal elements of the matrix products corresponding to six equations relating its elements can be written. In particular, in transitions involving b quarks, the scalar product of the third column with the complex conjugate of the first row must vanish:

$$V_{ud}^* V_{ub} \; + \; V_{cd}^* V_{cb} \; + \; V_{td}^* V_{tb} \; = \; 0 \tag{8}$$

$$\begin{array}{ll}
V_{ud}^* V_{us} + V_{cd}^* V_{cs} + V_{td}^* V_{ts} = 0 & \lambda\,\lambda\,\lambda^5 \\
V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0 & \lambda^3\,\lambda^3\,\lambda^3 \\
V_{us}^* V_{ub} + V_{cs}^* V_{cb} + V_{ts}^* V_{tb} = 0 & \lambda^4\,\lambda^2\,\lambda^2 \\
V_{ud}^* V_{td} + V_{us}^* V_{ts} + V_{ub}^* V_{tb} = 0 & \lambda^3\,\lambda^3\,\lambda^3 \\
V_{td}^* V_{cd} + V_{ts}^* V_{cs} + V_{tb}^* V_{cb} = 0 & \lambda^4\,\lambda^2\,\lambda^2 \\
V_{ud}^* V_{cd} + V_{us}^* V_{cs} + V_{ub}^* V_{cb} = 0 & \lambda\,\lambda\,\lambda^5
\end{array}$$

Figure 1. The six triangle equations from the unitarity condition of the CKM matrix.

Using the parametrization given in Equation (6), and neglecting contributions of order $\mathcal{O}(\lambda^7)$, the different terms, in this expression, are respectively:

$$\begin{aligned}
V_{ud}V_{ub}^* \; &= \; A\lambda^3(\overline{\rho} + i\overline{\eta}), \\
V_{cd}V_{cb}^* \; &= \; -A\lambda^3, \\
V_{td}V_{tb}^* \; &= \; A\lambda^3(1 - \overline{\rho} - i\overline{\eta})
\end{aligned} \tag{9}$$

The three expressions are proportional to $A\lambda^3$, which can be factored out, and the geometrical representation of Eq. (8), in the $(\overline{\rho}, \overline{\eta})$ plane, is a triangle with summit at C(0, 0), B(1, 0) and A($\overline{\rho}$, $\overline{\eta}$).

- The lengths CA and BA, to be denoted respectively by R_b and R_t, are given by

$$\overline{AC} \equiv R_b \equiv \frac{|V_{ud}V_{ub}^*|}{|V_{cd}V_{cb}^*|} = \sqrt{\overline{\rho}^2 + \overline{\eta}^2} = (1 - \frac{\lambda^2}{2})\frac{1}{\lambda}\left|\frac{V_{ub}}{V_{cb}}\right|, \qquad (10)$$

$$\overline{AB} \equiv R_t \equiv \frac{|V_{td}V_{tb}^*|}{|V_{cd}V_{cb}^*|} = \sqrt{(1-\overline{\rho})^2 + \overline{\eta}^2} = \frac{1}{\lambda}\left|\frac{V_{td}}{V_{cb}}\right|. \qquad (11)$$

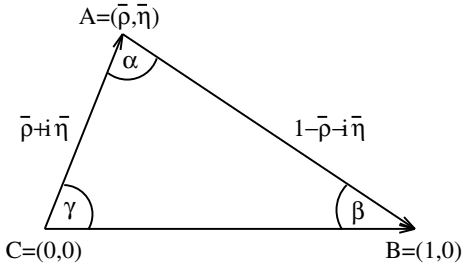

Figure 2. The Unitarity Triangle.

The angles β and $\gamma = \delta$ of the unitarity triangle are related directly to the complex phases of the CKM-elements V_{td} and V_{ub}, respectively, through

$$V_{td} = |V_{td}|e^{-i\beta}, \qquad V_{ub} = |V_{ub}|e^{-i\gamma}. \qquad (12)$$

Each of the angles is the relative phase of two adjacent sides (a part for possible extra π and minus sign) so that:

$$\beta = \quad arg(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}) \;=\; \text{atan}(\frac{\overline{\eta}}{1-\overline{\rho}}) \qquad (13)$$

$$\gamma = \quad arg(\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}) \;=\; \text{atan}(\frac{\overline{\eta}}{\overline{\rho}}) \qquad (14)$$

- The unitarity relation (Eq. 8) can be rewritten as

$$R_b e^{i\gamma} + R_t e^{-i\beta} = 1 \qquad (15)$$

- The angle α can be obtained through the relation $\alpha+\beta+\gamma = 180^\circ$ expressing the unitarity of the CKM-matrix.

The triangle shown in Figure 2 -which depends on two parameters $(\overline{\rho},\ \overline{\eta})$)-, plus $|V_{us}|$ and $|V_{cb}|$ gives the full description of the CKM matrix.

The Standard Model, with three families of quarks and leptons, predicts that all measurements have to be consistent with the point A$(\overline{\rho},\ \overline{\eta})$.

4.2 General Introduction to Oscillation and CP Violation

In this section we give a general introduction to the oscillation and CP violation formalism in view of their impact on the CKM matrix element determination.

A general system which satisfies the coupled Schrodinger equation can be written as:

$$i\frac{d}{dt}\begin{pmatrix} B^0 \\ \overline{B}^0 \end{pmatrix} = H\begin{pmatrix} B^0 \\ \overline{B}^0 \end{pmatrix} \qquad (16)$$
$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \; ; \; H_{ij} = M_{ij} - i\Gamma_{ij}/2$$

The B^0 and the $\overline{B}^0$ are the flavour eigenstates. Transitions between B^0 and $\overline{B}^0$ are then possible and the Hamiltonian has to be diagonalized to find the new eigenstates which are:

$$|B_L^0 >= p|B^0 > +q|\overline{B}^0 > \quad ; \qquad |B_H^0 >= p|B^0 > -q|\overline{B}^0 > \qquad (17)$$

where $|q|^2 + |p|^2 = 1$. Solving the eigenvalues equation (supposing that CPT is conserved) and defining $\Delta m = M_H - M_L$ and $\Delta\Gamma = \Gamma_H - \Gamma_L$ it follows:

$$\begin{aligned} \Delta m^2 - 1/4\Delta\Gamma^2 = &\quad 4|M_{12}|^2 - |\Gamma_{12}|^2 \\ \Delta m\Delta\Gamma = &\quad 4Re(M_{12}\Gamma_{12}^*) \end{aligned} \qquad (18)$$

In the Standard Model, $\mathrm{B}^0 - \overline{\mathrm{B}^0}$ transitions occur through a second-order process -a box diagram- with a loop that contains W and up-type quarks. The exchange of the top quark dominates the part relative to the mass difference (M_{12}), while light quarks contribute to the part relative to the decay (Γ_{12}) (only common states to B^0 and $\overline{B}^0$ contribute). It results that : $\Gamma_{12}/M_{12} = m_b^2/m_t^2 << 1$. The relations (18) simplify to:

$$\begin{aligned} \Delta m = &\quad 2|M_{12}| \\ \Delta\Gamma = &\quad \frac{2Re(M_{12}\Gamma 12^*)}{|M_{12}|} \\ \frac{q}{p} = &\quad -\frac{|M_{12}|}{M_{12}} \end{aligned} \qquad (19)$$

The last expression is valid at leading approximation[10]. We are now interested in the time evolution of the flavour eigenstates in the

[10]Beyond the leading approximation the expression becomes : $\frac{q}{p} = -\frac{\Delta m - i/2\Delta\Gamma}{2M_{12} - i\Gamma_{12}}$

hypothesis ($\Delta\Gamma << \Delta m$):

$$|B^0_{phys.}(t)>= e^{-imt}\ e^{-\Gamma/2\ t}(cos\ \Delta m/2t\ |B^0> +\ i\ \frac{q}{p}\ sin\Delta m\ t/2\ |\overline{B}^0>)$$

$$|\overline{B}^0_{phys.}(t)>= e^{-imt}\ e^{-\Gamma/2t}(cos\ \Delta m/2t\ |\overline{B}^0> +\ i\ \frac{q}{p} sin\Delta mt/2\ |B^0>)$$

It follows:

$$< f|H|B^0_{phys.}(t) > |^2 = \frac{e^{-\Gamma t}}{2}[(1+cos\Delta mt)| < f|H|B^0(t) > |^2 + (1-cos\Delta mt)|\frac{q}{p}|^2| < f|H|\overline{B}^0(t) > |^2 + 2\mathrm{Im}\ (\ |\frac{\mathrm{q}}{\mathrm{p}}|\ \sin\ \Delta\mathrm{mt}\ \ < \mathrm{f}|\mathrm{H}|\mathrm{B}^0(\mathrm{t}) >< \mathrm{f}|\mathrm{H}|\overline{\mathrm{B}}^0(\mathrm{t}) >^*)] \quad (20)$$

The probability that a meson B^0 produced (by strong interaction) at time $t = 0$ transforms (by weak interaction) into a $\overline{B}^0$ (or stays as a B^0) at time t is given by:

$$Prob(B^0_{phys.}(t) \to B^0_{phys.}(t)(\overline{B}^0_{phys.}(t))) = \frac{1}{2}e^{-\Gamma t}(1+(-)cos\Delta mt) \quad (21)$$

Defining:

$$\lambda = \frac{q}{p}\frac{< f|H|\overline{B}^0 >}{< f|H|B^0 >} = \frac{q}{p}\frac{\overline{A_f}}{A_f}\ \ ;\ \ \overline{\lambda} = \frac{p}{q}\frac{< \overline{f}|H|B^0 >}{< \overline{f}|H|\overline{B}^0 >} = \frac{p}{q}\frac{A_{\overline{f}}}{\overline{A_{\overline{f}}}} \quad (22)$$

the equation becomes:

$$| < f|H|B^0_{phys.}(t) > |^2 = \frac{e^{-\Gamma t}}{2}| < f|H|B^0(t) > |^2 [\ (1+cos\Delta mt) + (1-cos\Delta mt)|\lambda|^2 - 2Im(\lambda)sin\Delta mt\ ] \quad (23)$$

and similarly for $| < f|H|\overline{B}^0_{phys.}(t) > |, | < \overline{f}|H|B^0_{phys.}(t) > |$ and $| < \overline{f}|H|\overline{B}^0_{phys.}(t) > |$. We concentrate here on the two cases of CP violation which are the most relevant for CKM physics.

B^0 sector: direct CP Violation and CP Violation in the interference between mixing and decays.

CP violation can occur because $\mathrm{Im}\lambda \neq \mathrm{Im}\overline{\lambda}$ and/or when $|\lambda|$, $|\bar{\lambda}|$ are different from unity. In this case the four quantities $Prob(B^0_{phys.} \to f), Prob(\overline{B}^0_{phys.} \to f), Prob(B^0_{phys.} \to \overline{f})$ and $Prob(\overline{B}^0_{phys.} \to \overline{f})$ (see eq. 23) have to be studied and thus $|\lambda|^2, |\overline{\lambda}|^2, \mathrm{Im}\lambda, \mathrm{Im}\overline{\lambda}$ are determined. The simplest case is when the final state f is a specific CP state. In this case $\overline{\lambda} = 1/\lambda \equiv \lambda_f$ and the previous conditions simplify to $\mathrm{Im}\lambda_\mathrm{f} \neq 0$ and/or $|\lambda_f| \neq 1$. The following asymmetry can be studied:

$$\begin{aligned} A_{CP}(mixing - decay) = & \quad \frac{Prob(B^0_{phys}(\Delta t)\to f) - Prob(\overline{\mathrm{B}}^0_{phys}(\Delta t)\to f)}{Prob(\mathrm{B}^0_{phys}(\Delta t)\to f) + Prob(\overline{\mathrm{B}}^0_{phys}(\Delta t)\to f)} \\ = & \quad C_f \; cos\Delta m_d \; \Delta t \; + S_f \; sin\Delta m_d \; \Delta t \; , \end{aligned} \tag{24}$$

where

$$C_f = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \; ; \; S_f = -\frac{2 \; \mathrm{Im}\lambda_\mathrm{f}}{1 + |\lambda_f|^2} \tag{25}$$

C_f corresponds to direct CP violation, since it is related to differences in the decay amplitudes, while S_f is related to the interference between the mixing and decays, involving the imaginary parts of p/q and of the decay amplitudes. It is important to note that, also in case $|\lambda_f|$=1, CP violation is possible if $Im\lambda_f \neq 0$. This case is particularly interesting. When only one amplitude dominates the decay process, $|\lambda_f| = 1$, implying $C_f = 0$ and $S_f = -\mathrm{Im}\lambda_\mathrm{f}$. We will see in the following that $-\mathrm{Im}\lambda_\mathrm{f}$ is the sine of twice an angle of the unitarity triangle.

B^+ sector: the direct CP Violation.

The transition amplitudes can be written as:

$$\begin{aligned} |< f|H|B^+ >| &= v_1 A_1 e^{i\theta_1} + v_2 A_2 e^{i\theta_2} \\ |< \overline{f}|H|B^- >| &= v_1^* A_1 e^{i\theta_1} + v_2^* A_2 e^{i\theta_2} \end{aligned} \tag{26}$$

where $v_{1,2}$ are the weak-CKM couplings and $A_{1,2}(\theta_{1,2})$ are the modulus and the strong phase respectively. The weak phase changes sign under CP (the strong phase do not). It follows:

$$\begin{aligned} a_{CP}(direct) &= \frac{|< f|H|B^+ >|^2 - |< \overline{f}|H|B^- >|^2}{|< f|H|B^+ >|^2 + |< \overline{f}|H|B^- >|^2} \\ &= \frac{2 r_B sin(arg(v_1/v_2)) sin(\theta_1 - \theta_2)}{1 + r_B^2 + 2 r_B cos(arg(v_1/v_2)) cos(\theta_1 - \theta_2)} \quad ; \quad r_B = \frac{|v_1|A_1}{|v_2|A_2} \end{aligned} \tag{27}$$

The basic conditions to have direct CP violation are the presence of two competing amplitudes, $r_B \neq 0$, and of nonzero weak phase and strong phase differences.

The weak phase difference is related to one of the unitarity triangle angles. To be more explicit, if for instance we consider a process which can occur through V_{ub} and V_{cb} mediated transitions, $arg(v_1/v_2) = \gamma$.

CP Violation in the Kaon sector.

Historically the parameter characterizing CP violation was defined as ϵ given in:

$$|K_S>= \frac{|K_1> +\epsilon K_2>}{\sqrt{(}1+|\epsilon|^2)} \;\; ; \;\; |K_L>= \frac{|K_2> +\epsilon K_1>}{\sqrt{(}1+|\epsilon|^2)} \tag{28}$$

$K_{1,2}$ are the CP eigenstates and the previous equation can be written in terms of flavour eigenstates:

$$|K_S>= \frac{1}{\sqrt{2}\ \sqrt{(}1+|\epsilon|^2)}\ [(1+\epsilon)|K^0> +(1-\epsilon)|\overline{K}^0>]$$
$$|K_L>= \frac{1}{\sqrt{2}\ \sqrt{(}1+|\epsilon|^2)}\ [(1+\epsilon)|K^0> -(1-\epsilon)|\overline{K}^0>]$$

where ϵ is related to p, q parameters by: $\epsilon = \frac{p-q}{p+q}$.

Two CP violating quantities are measured in the neutral Kaon sector:

$$\eta_{00} = \frac{<\pi^0\pi^0|H|K_L>}{<\pi^0\pi^0|H|K_S>} \;\; ; \;\; \eta_{\pm} = \frac{<\pi^+\pi^-|H|K_L>}{<\pi^+\pi^-|H|K_S>} \tag{29}$$

Defining

$$A_{00} =<\pi^0\pi^0|H|K^0> \;\; ; \;\; \overline{A}_{00} =<\pi^0\pi^0|H|\overline{K}^0>$$
$$A_{+-} =<\pi^+\pi^-|H|K^0> \;\; ; \;\; \overline{A}_{+-} =<\pi^+\pi^-|H|\overline{K}^0>$$
$$\lambda_{00} = \frac{q}{p}\frac{\overline{A}_{00}}{A_{00}} \;\; ; \;\; \lambda_{+-} = \frac{q}{p}\frac{\overline{A}_{+-}}{A_{+-}}$$

which implies

$$\eta_{00} = \frac{1-\lambda_{00}}{1+\lambda_{00}} \;\; ; \;\; \eta_{+-} = \frac{1-\lambda_{+-}}{1+\lambda_{+-}} \tag{30}$$

The $\pi\pi$ final states can have isospin I=0,2. Experimentally it is observed that $A_{I=2}/A_{I=0} \simeq 1/20$ (known has the Δ I=1/2 rule). In the approximation that only the I=0 amplitude contributes -no direct CP violation- it follows:

$$\epsilon_K = \frac{<\pi^0\pi^0_{I=0}\ |H|K_L>}{<\pi^0\pi^0_{I=0}\ |H|K_S>} = \eta_{00} \tag{31}$$

and similarly for the η_{+-}.

Contrarily to B mesons, in case of Kaon physics $\Delta\Gamma \simeq \Delta M$. Using the expression of q/p (eq. 18) it follows that:

$$\epsilon_K = \frac{e^{i\pi/4}}{\sqrt{2}\Delta m_K} \left(\mathrm{ImM_{12}} + 2\zeta \mathrm{ReM_{12}}\right) \tag{32}$$

where $\zeta = Im(A(K \to \pi\pi)_{I=0})/Re(A(K \to \pi\pi)_{I=0})$. The contribution, proportional to ζ, which is of about 2% correction to $|\varepsilon_K|$ can be neglected.

4.3 Standard Model formulae relating $\overline{\rho}$ and $\overline{\eta}$ to experimental and theoretical inputs

Five measurements restrict, at present, the possible range of variation of the $\overline{\rho}$ and $\overline{\eta}$ parameters:

- B hadrons can decay through the $b \to c$ and $b \to u$ transitions. Semileptonic decays offer a relatively large branching fraction ($\simeq$ 10 %) and corresponding measurements can be interpreted using a well established theoretical framework. The relative rate of charmless over charmed b-hadron semileptonic decays is proportional to the square of the ratio:

$$\left|\frac{V_{ub}}{V_{cb}}\right| = \frac{\lambda}{1 - \frac{\lambda^2}{2}} \sqrt{\overline{\rho}^2 + \overline{\eta}^2} \ , \tag{33}$$

  and it allows to measure the length of the side AC of the triangle (Figure 3).

- In the Standard Model, $\mathrm{B^0}-\overline{\mathrm{B^0}}$ oscillations occur through a second-order process -a box diagram- with a loop that contains W and up-type quarks. The box diagram with the exchange of a *top* quark gives the dominant contribution. The oscillation probability is given in eq. (21) and the time oscillation frequency, which can be related to the mass difference between the light and heavy mass eigenstates of the $\mathrm{B}^0_d - \overline{\mathrm{B}^0_d}$ system (eq. 19, $\Delta m = 2|M_{12}|$), is expressed, in the SM, as[11]:

$$\Delta m_d \ = \ \frac{G_F^2}{6\pi^2} m_W^2 \ \eta_b S(x_t) \ A^2\lambda^6 \ [(1-\overline{\rho})^2 + \overline{\eta}^2] \ m_{B_d} \ f_{B_d}^2 \hat{B}_{B_d} \ , \tag{34}$$

[11] Δm_q is usually expressed in ps^{-1} unit. 1 ps^{-1} corresponds to 6.58 10^{-4}eV.

where $S(x_t)$ is the Inami-Lim function [26] and $x_t = m_t^2/M_W^2$, m_t is the $\overline{MS}$ top quark mass, $m_t^{\overline{MS}}(m_t^{\overline{MS}})$, and η_b is the perturbative QCD short-distance NLO correction. The value of $\eta_b = 0.55 \pm 0.01$ has been obtained in [27] and $m_t = (167 \pm 5)$ GeV is used, as deduced from measurements by CDF and D0 Collaborations [21]. The remaining factor, $f_{B_d}^2 \hat{B}_{B_d}$, encodes the information of non-perturbative QCD. Apart for $\overline{\rho}$ and $\overline{\eta}$, the most uncertain parameter in this expression is $f_{B_d}\sqrt{\hat{B}_{B_d}}$ (6.4).
In the vacuum saturation approximation the matrix element of the V-A current is calculated between the vacuum and the pseudoscalar meson and only the axial current contributes. The constant f_{B_d} translates the probability that the quark and the antiquark meet to decay or the size of the B meson wave function at the origin. Another parameter is also introduced: the bag factor $\hat{B}_{B_d}$ which is inserted to take into account all possible deviation from vacuum saturation approximation. The values of the bag factors are expected to be close of the unity.
The measurement of Δm_d gives a constraint on the length of the side AB of the triangle (Figure 3).

- The $\mathrm{B_s^0} - \overline{\mathrm{B_s^0}}$ time oscillation frequency, which can be related to the mass difference between the light and heavy mass eigenstates of the $\mathrm{B_s^0} - \overline{\mathrm{B_s^0}}$ system, is proportional to the square of the $|V_{ts}|$ element. Neglecting terms, having a small contribution, $\mid V_{ts} \mid$ is independent of $\overline{\rho}$ and $\overline{\eta}$. The measurement of Δm_s would then give a strong constraint on the non-perturbative QCD parameter $f_{B_s}^2 \hat{B}_{B_s}$. In any case, the ratio between the values of the mass difference between the mass-eigenstates, measured in the $\mathrm{B_d^0}$ and in the $\mathrm{B_s^0}$ systems can be used:

$$\frac{\Delta m_d}{\Delta m_s} = \frac{m_{B_d} f_{B_d}^2 \hat{B}_{B_d}}{m_{B_s} f_{B_s}^2 \hat{B}_{B_s}} \left(\frac{\lambda}{1-\frac{\lambda^2}{2}}\right)^2 \frac{(1-\overline{\rho})^2 + \overline{\eta}^2}{\left(1 + \frac{\lambda^2}{1-\frac{\lambda^2}{2}}\overline{\rho}\right)^2 + \lambda^4 \overline{\eta}^2} \ . \quad (35)$$

The advantage in using the ratio $\frac{\Delta m_d}{\Delta m_s}$, instead of only Δm_d, is that the ratio $\xi = f_{B_s}\sqrt{\hat{B}_{B_s}}/f_{B_d}\sqrt{\hat{B}_{B_d}}$ is expected to be better determined from theory than the individual quantities entering into its expression. The measurement of the ratio $\Delta m_d/\Delta m_s$ gives a similar type of constraint as Δm_d, on the length of the side AB of the triangle.

- Indirect CP violation in the $\mathrm{K}^0 - \overline{\mathrm{K}^0}$ system is usually expressed in terms of the $|\varepsilon_K|$ parameter (as defined in Section 4.2) which is the fraction of CP violating component in the mass eigenstates. In the SM, the following equation is obtained

$$|\varepsilon_K| = C_\varepsilon \, A^2\lambda^6 \, \overline{\eta} \times \left[-\eta_1 S(x_c)\left(1-\frac{\lambda^2}{2}\right) + \eta_2 S(x_t) A^2\lambda^4 (1-\overline{\rho}) + \eta_3 S(x_c, x_t)\right] \hat{B}_K \tag{36}$$

where $C_\varepsilon = \frac{G_F^2 f_K^2 m_K m_W^2}{6\sqrt{2}\pi^2 \Delta m_K}$.
$S(x_i)$ and $S(x_i, x_j)$ are the appropriate Inami-Lim functions [26] depending on $x_q = m_q^2/m_W^2$, including the next-to-leading order QCD corrections [27, 28]. The most uncertain parameter is $\hat{B}_K$ (6.4).
The constraint brought by the measurement of $|\varepsilon_K|$ corresponds to an hyperbola in the $(\overline{\rho}, \overline{\eta})$ plane (Figure 3).

- The measurement of CP violation in the B sector.

The mixing induced CP asymmetry, $a_{J\psi K_S}$, in $\mathrm{B}^0_\mathrm{d} \to \mathrm{J}/\psi \mathrm{K_S}$ or $(\to J/\psi K_L)$ decays allows to determine the angle β of the Unitarity Triangle essentially without any hadronic uncertainties. As explained before, a possible manifestation of the CP asymmetry could appear in the interference between amplitudes describing decays with and without mixing. The process $B^0 \to J/\Psi K^0$ is dominated by tree diagram[12] and it follows that:

$$\begin{aligned} \frac{q}{p} &= \left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right) \quad \text{from B mixing} \\ \frac{< J/\Psi K^0 | H | \overline{B^0} >}{< J/\Psi K^0 | H | B^0 >} &= \left(\frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*}\right) \quad \text{from B decay amplitudes} \\ &\quad \left(\frac{V_{cd}^* V_{cs}}{V_{cd} V_{cs}^*}\right) \quad \text{from K mixing} \\ |\lambda_f|^2 = 1 \quad ; \quad & \mathrm{Im}\ \lambda_\mathrm{f} = \eta_{\mathrm{CP}} \sin 2\beta \end{aligned}$$

[12]The same process could be described by a Penguin diagram (with a ts transition) and a J/Ψ emitted from gluons. This process is proportional to $V_{ts}V_{tb}^*$. It is important to note that the amplitude associated to this process has the same phase, at order $\mathcal{O}(\lambda^2)$, as the dominant tree-level one. At order $\mathcal{O}(\lambda^4)$, V_{ts} is complex and differs from V_{cb}. Thus the correction to β is suppressed by a factor $\mathcal{O}(\lambda^4)$ and by an extra factor because the J/Ψ must be emitted by at least three gluons.

where η_{CP} is the CP eigenvalue of the final state. The asymmetry defined in eq. 24 gives:

$$A_{CP}(J/\Psi K_S) = -2\ \sin(2\beta)\ sin\Delta m_d\ \Delta t\ , \qquad (37)$$

The measurement of $A_{CP}(J/\Psi K^0)$ gives a constraint corresponding to $\sin(2\beta)$, in the $(\overline{\rho}, \overline{\eta})$ plane (Figure 3).

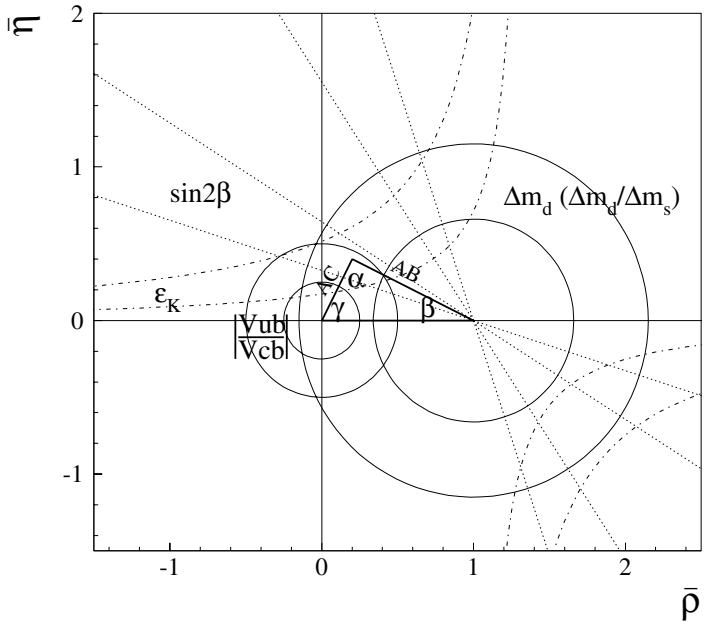

Figure 3. Unitarity Triangle. Constraints from $|V_{ub}|/|V_{cb}|$, $|\varepsilon_K|$, Δm_d or $\Delta m_d/\Delta m_s$ and $\sin(2\beta)$ are shown.

Constraints on $\overline{\rho}$ and $\overline{\eta}$ are obtained by comparing present measurements with theoretical expectations using the expressions given above and taking into account the different sources of uncertainties. In addition to $\overline{\rho}$ and $\overline{\eta}$, these expressions depend on other quantities. Additional measurements or theoretical determinations have been used to provide information on the values of these parameters; details are given in the next sections.

To illustrate the different constraints described in the present section, in Figure 3, the uncertainty bands for the quantities, obtained using Eqs. (33)–(37), are presented. Each band, corresponds to only one of

the constraints and contains 95% of the events obtained by varying the input parameters.

In the first column of Table 2 the different measured quantities are listed, with their explicit dependence on $\overline{\rho}$ and $\overline{\eta}$ given in the third column.

Measurement	$CKM\times$other	Constraint
$Br(b\to u\ell\overline{\nu})/\ Br(b\to c\ell\overline{\nu})$	$\lvert V_{ub}/V_{cb}\rvert^2$	$\overline{\rho}^2+\overline{\eta}^2$
Δm_d	$\lvert V_{td}\rvert^2 f_{B_d}^2 B_{B_d} f(m_t)$	$(1-\overline{\rho})^2+\overline{\eta}^2$
$\frac{\Delta m_d}{\Delta m_s}$	$\left\lvert\frac{V_{td}}{V_{ts}}\right\rvert^2 \frac{f_{B_d}^2 B_{B_d}}{f_{B_s}^2 B_{B_s}}$	$(1-\overline{\rho})^2+\overline{\eta}^2$
$\lvert\varepsilon_K\rvert$	$f(A,\overline{\eta},\overline{\rho},B_K)$	$\propto\overline{\eta}(1-\overline{\rho})$
$A_{CP}(J/\Psi K^0)$	$\sin(2\beta)$	$2\overline{\eta}\overline{\rho}/\left[(1-\overline{\rho})^2+\overline{\eta}^2\right]$

Table 2. Different measurements contributing in the determination of $\overline{\rho}$ and $\overline{\eta}$, with their functional dependences.

The values and errors of the relevant quantities used in the fit of the CKM parameters are summarized in Table 3.

For the extraction of the CKM parameters we use the Bayesian approach [37]. In these lectures we do not enter into any details related to the statistical method. Here we want just to explain the splitting of the errors as given in Table 3. We take a Gaussian distribution ($\mathcal{G}(x-x_0)$) when the uncertainty is dominated by statistical effects, or when there are several contributions of similar importance to systematic errors, so that the central limit theorem applies. We take a uniform p.d.f. if the parameter value is believed to be (almost) certainly within a given interval, and the points inside this interval are considered as equally probable. The second model is used for theoretical uncertainties. $\mathcal{U}(x)=1/2\sigma_{\rm theo}$ for $x\in[x_0-\sigma_{\rm theo},x_0+\sigma_{\rm theo}]$ and $\mathcal{U}(x)=0$ elsewhere. The combined p.d.f. ($\mathcal{P}$) is obtained by convoluting the Gaussian p.d.f. ($\mathcal{G}$) with the uniform p.d.f. ($\mathcal{U}$): $\mathcal{P}=\mathcal{G}\otimes\mathcal{U}$. When several determinations of the same quantity are available the final p.d.f, in the Bayesian approach, is obtained by taking the product of individual p.d.f.s (and normalizing the obtained distribution to unity). For more details on statistical methods see [29].

5. B Physics at different facilities

In this chapter we will discuss B physics at different machines. The main contributors in B hadron studies are:

- the e^+e^- colliders

Parameter	Value	Gaussian (σ)	Uniform (half-width)	Ref.
λ	0.2241	0.0036	-	[30]
$\|V_{cb}\|$(excl.)	42.1×10^{-3}	2.2×10^{-3}	-	Section 6.1
$\|V_{cb}\|$(incl.)	41.4×10^{-3}	0.7×10^{-3}	0.6×10^{-3}	Section 6.1
$\|V_{ub}\|$(excl.)	33.0×10^{-4}	2.4×10^{-4}	4.6×10^{-4}	Section 6.2
$\|V_{ub}\|$(incl.)	40.9×10^{-4}	4.6×10^{-4}	3.6×10^{-4}	Section 6.2
Δm_d	0.502 ps^{-1}	0.007 ps^{-1}	-	Section 6.3
Δm_s	> 14.5 ps^{-1} at 95% C.L.	sensitivity 18.3 ps^{-1}		Section 6.3
m_t	167 GeV	5 GeV	-	[21]
$f_{B_d}\sqrt{\hat{B}_{B_d}}$	223 MeV	33 MeV	± 12 MeV	Section 6.4
$\xi = \frac{f_{B_s}\sqrt{\hat{B}_{B_s}}}{f_{B_d}\sqrt{\hat{B}_{B_d}}}$	1.24	0.04	± 0.06	Section 6.4
η_b	0.55	0.01	-	[27]
$\hat{B}_K$	0.86	0.06	0.14	Section 6.4
$\|\varepsilon_K\|$	2.280×10^{-3}	0.019×10^{-3}	-	[23]
η_1	1.38	0.53	-	[28]
η_2	0.574	0.004	-	[27]
η_3	0.47	0.04	-	[28]
f_K	0.159 GeV	fixed		[23]
Δm_K	0.5301 $\times 10^{-2}$ ps^{-1}	fixed		[23]
$\sin(2\beta)$	0.739	0.048	-	Section 6.5
m_b	4.21 GeV	0.08 GeV	–	Section 6.1
m_c	1.3 GeV	0.1 GeV	–	Section 6.1
α_s	0.119	0.03	–	[28]
G_F	1.16639 $\times 10^{-5}$GeV^{-2}	fixed		[23]
m_W	80.23 GeV	fixed		[23]
$m_{B_d^0}$	5.2794 GeV	fixed		[23]
$m_{B_s^0}$	5.3696 GeV	fixed		[23]
m_K	0.493677 GeV	fixed		[23]

Table 3. Values of the relevant quantities used in the fit of the CKM parameters. In the third and fourth columns the Gaussian and the flat parts of the uncertainty are given (see text), respectively. The central values and errors are those adopted at the end of the "CKM Unitarity Triangle" Workshops ([29],[30]) and by HFAG [31] and are given and explained in the following sections (as indicated in the last column). The averages for the non perturbative QCD parameters are made by the CKM-LDG group [32]

– the symmetric-B factories operating at $\Upsilon(4S)$ (ARGUS/ CRYSTAL BALL and CLEO/CUSB experiments running at DORIS and CESR, respectively, from 1979 to 2002)

  - the asymmetric-B factories operating at $\Upsilon(4S)$ (Belle at KEK and BaBar at PEP experiments running from 1999)
  - the Z^0 resonance experiments (the LEP collaborations which run from 1989 to 1995 and the SLD collaboration at SLC which run from 1989 to 1998).

- the $p\overline{p}$ collider
  - the TeVatron collider, operating at $\sqrt{s} = 1.8$ TeV - phase I (D0 and CDF experiment from 1987 to 2000). They are presently running with an improved luminosity at $\sqrt{s} \simeq 1.9$ TeV -phase II.

An overview of these experiments, operating at different facilities, is given in Table 4.

At the $\Upsilon(4S)$, pairs of $B^{\pm}$ and B^0_d $\left(\overline{B^0_d}\right)$ mesons are produced on top of the hadronic background continuum from lighter $q\overline{q}$ pairs. The two B mesons are created simultaneously in a L=1 coherent state, such that before the first decay the final state contains a B and a $\overline{B}$; at the time of the decay of the first B meson, the second one is in the opposite flavour eigenstate. The production cross section is about 1.2 nb. Because of the energy available, only $B^{\pm}$ and B^0_d mesons are emitted. In symmetric B-factories B particles are produced almost at rest while at the asymmetric factories they have a boost of $\beta\gamma = 0.56$ (0.44) (for BaBar (Belle)). It is important to note that, in both cases, the average B momentum in the $\Upsilon(4S)$ rest frame is of the order of about 350 MeV/c.

Considering that the B lifetime is of the order of 1.6 ps, the flight distance of a B hadron, defined as $L = \gamma\beta c\tau$ is, on average, at asymmetric B-factories, of the order of 250 μm. This distance is measurable and highlights the greatest advantage of asymmetric B-factories where time dependent analyses, necessary for CP violation studies, are possible.

The B decay products are the only tracks produced in the events, there is no accompanying additional hadron. As a consequence the energy taken by each B meson is equal to the half the total energy in the e^+e^- center-of-mass frame; this constraint is, for instance, very important in rejecting the non-B events. The decay products of the two B particles are spread isotropically over the space and such events can be distinguished from the continuum which are more jetty-like.

At the Z^0 resonance, B hadrons are produced from the coupling of the Z^0 to a $b\overline{b}$ quark pair. The production cross section is of ~ 6 nb, which is five times larger than at the $\Upsilon(4S)$. Hadronic events account for about 70 % of the total production rate; among these, the fraction of $b\overline{b}$ events

is $\sim 22\%$[13], which is rather similar to the one observed when running at the $\Upsilon(4S)$ energy ($\sim$ 25 %). B hadrons are thus copiously produced[14]. The produced $b\bar{b}$ pair picks up from the vacuum other quark-antiquarks pairs and hadronizes into B hadrons plus few other particles. Therefore, not only $B^{\pm}$ and B^0_d mesons are produced, but also B^0_s mesons or b-baryons can be present in the final state. The b and $\bar{b}$ quarks hadronize independently. b quarks fragment differently from light quarks, because of their high mass as compared with Λ_{QCD}. As a result, B hadrons carry, on average, about 70% of the available beam energy, whereas the rest of the energy is distributed among the other particles emitted in the fragmentation process. As a consequence, the two B hadrons fly in opposite directions and their decay products belong to jets situated in two different hemispheres.

The hard fragmentation and the long lifetime of the b quark make that the flight distance of a B hadron at the Z pole, defined as $L = \gamma\beta c\tau$, is on average of the order of 3 mm.

At $p\overline{p}$ colliders, the situation is rather different. Here b quarks are produced mainly through the gluon-gluon fusion process $gg \rightarrow b\bar{b}$. At the Fermilab Collider ($\sqrt{s} = 1.8$ TeV), the differential b-production cross section depends on the rapidity and on the transverse momentum. In total, it is typically of the order of $50\mu b$, which is large. B decay products are situated inside events having an average multiplicity which is much larger than the multiplicity at the Z pole. Furthermore the ratio $\sigma_{b\bar{b}}/\sigma_{tot}$ is of the order of a few per mill. As a consequence, only specific channels e.g. with fully reconstructed final states, or semileptonic decays, can be studied with a reasonable signal to background ratio.

Registered data sets from experiments operating at different facilities are summarized in Table 4.

6. Evaluation of the parameters entering in the determination of the CKM parameters.

This section gives a short summary on the determination of the quantities entering in Unitarity Triangle fits. The discussion on the central values and attributed errors for these quantities has been extensively done and agreed values were adopted during the First Workshop on the

[13] whereas the fraction of $c\bar{c}$ events is $\sim 17\%$.

[14] In the intermediate energy region ("continuum") where the annihilation through one photon is dominant (V-coupling) the cross section scales with the energy available in the center of mass (squared), being of the order of 30 pb at 30 GeV and of about 10 pb at 60 GeV. In this energy range the fraction of $b\bar{b}$ events is $\sim 9\%$ whereas the fraction of $c\bar{c}$ events is $\sim 35\%$ (being the coupling proportional to the square of the electric charge.

Experiments	Number of $b\bar{b}$ events ($\times 10^6$)	Environment	Characteristics
LEP Coll.	~ 1 per expt. (4 expts.)	Z^0 decays ($\sigma_{b\bar{b}} \sim$ 6nb)	back-to-back 45 GeV b-jets, all B hadron produced.
SLD	~ 0.1	Z^0 decays ($\sigma_{b\bar{b}} \sim$ 6nb)	back-to-back 45 GeV b-jets, all B hadron produced, beam polarized.
ARGUS	~ 0.2	$\Upsilon(4S)$ decays ($\sigma_{b\bar{b}} \sim$ 1.2nb)	mesons produced at rest, B^0_d and B^+.
CLEO	~ 9	$\Upsilon(4S)$ decays ($\sigma_{b\bar{b}} \sim$ 1.2nb)	mesons produced at rest, B^0_d and B^+.
BaBar Belle	~ 130	$\Upsilon(4S)$ decays ($\sigma_{b\bar{b}} \sim$ 1.2nb)	asymmetric B-factories B^0_d and B^+.
CDF	$\sim$ several	$p\bar{p}$ collider-Run I $\sqrt{s} = 1.8$ TeV ($\sigma_{b\bar{b}} \sim 50\mu$b)	events triggered with leptons, all B hadron produced.

Table 4. Summary of recorded statistics by experiments operating at different facilities and main characteristics.

"Unitarity Triangle Parameters Determination" held at CERN from the 12-15 February 2002 [29]. More recent values are taken from the updates done during the Second Workshop on the "Unitarity Triangle Parameters Determination" held at Durahm from the 5-9 April 2003 [30]. Many of the experimental averages have been calculated by the HFAG (Heavy Flavour Averaging Group) and can be found in [31].

6.1 Determination of $|V_{cb}|$

The $|V_{cb}|$ element of the CKM matrix can be accessed by studying the decay rate of inclusive and exclusive semileptonic b-decays.

Determination of $|V_{cb}|$ using inclusive analyses. The first method to extract $|V_{cb}|$ makes use of B-hadrons inclusive semileptonic decays and of the theoretical calculations done in the framework of the OPE (Operator Product Expansion). The inclusive semileptonic width $\Gamma_{s.l.}$ is expressed as:

$$\Gamma_{s.l.} = \frac{BR(b \to cl\nu)}{\tau_b} = \gamma_{theory}|V_{cb}|^2;$$
$$\gamma_{theory} = f(\alpha_s, m_b, \mu_\pi^2, 1/m_b^3...). \quad (38)$$

From the experimental point of view the semileptonic width has been measured by the LEP/SLD and $\Upsilon(4S)$ experiments with a relative precision of about 2%:

$$\begin{aligned} \Gamma_{sl} = &\quad (0.431 \pm 0.008 \pm 0.007)10^{-10} MeV \quad \Upsilon(4S) \\ \Gamma_{sl} = &\quad (0.439 \pm 0.010 \pm 0.007)10^{-10} MeV \quad \text{LEP/SLD} \\ \Gamma_{sl} = &\quad (0.434 \times (1 \pm 0.018))10^{-10} MeV \quad \text{average} \end{aligned} \tag{39}$$

Using the theoretical determinations of the parameters entering into the expression of γ_{theory} in Eq. (38), the uncertainty on $|V_{cb}|$ comes out to be of the order of $5\%(2.0\ 10^{-3})$. Thus the precision on the determination of $|V_{cb}|$ is limited by theoretical uncertainties which are mainly related to the non perturbative QCD parameters.

These parameters can be experimentally determined using the fact that OPE gives expressions in terms of operators whose averaged values are universal when considering different aspects of the same reaction.

Moments of the hadronic mass spectrum, of the lepton energy spectrum and of the photon energy in the $b \to s\gamma$ decay are sensitive to the same non perturbative QCD parameters contained in the factor γ_{theory} of Eq. (38) and, in particular, to the mass of the b and c quarks and to the Fermi motion of the heavy quark inside the hadron, μ_π^2[15]. For more details, see for instance [34],[35].

First measurements have been done by CLEO and preliminary results have been obtained by BaBar and DELPHI.

As an example, DELPHI data have been used for the determination of these non perturbative QCD parameters and an illustration of the obtained results is given in Figure 4.

Using the experimental results on Γ_{sl}, Eq. (39), and on the determination of the non perturbative QCD parameters, the following value for $|V_{cb}|$ is obtained:

$$|V_{cb}| = (41.4 \pm 0.7 \pm 0.6_{\text{theo.}})\, 10^{-3} (\text{inclusive}) \tag{40}$$

This result brings an important improvement in the determination of the $|V_{cb}|$ element. The dominant part of the initial theoretical errors is now accounted for as experimental uncertainties, using the fitted non perturbative quantities (m_b, m_c, μ_π^2 and $1/m_b^3$ contributions) and the remaining theoretical error has been reduced by more than a factor three (previously the quoted theoretical error was $\pm$ 2.0 10^{-3}).

[15] In another formalism, based on pole quark masses, the $\overline{\Lambda}$ and λ_1 parameters are used, which can be related to the difference between hadron and quark masses and to μ_π^2, respectively.

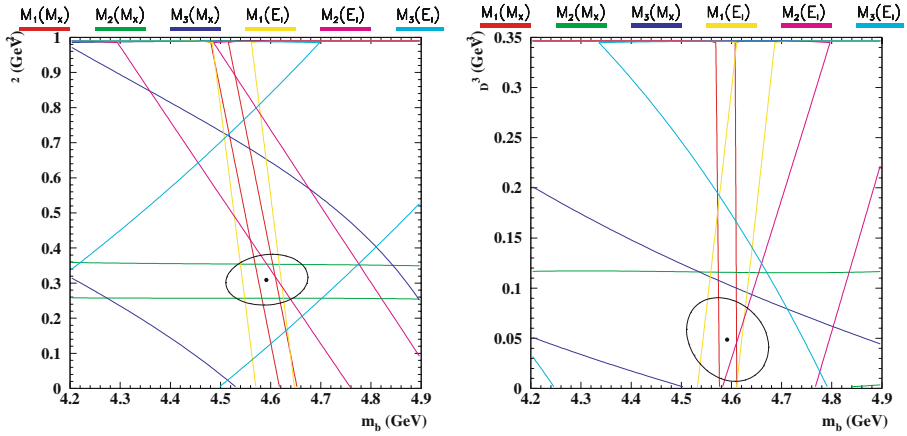

Figure 4. The moments analysis performed by DELPHI Collaboration [34]. The projection of the constraints, brought by six measured moments, over the $m_b - \mu_\pi^2$ (left) and $m_b - \rho_D^3$ (right) planes (ρ_D being related to the corrections corresponding to $1/m_b^3$ terms). The bands correspond to the total measurement accuracy and are given by keeping all other parameters fixed at their central values. The ellipses represent the 1σ contour.

Determination of $|V_{cb}|$ using B $\to$ D$^*\ell\nu$ analyses. An alternative method to determine $|V_{cb}|$ is based on exclusive $\overline{\mathrm{B}^0_\mathrm{d}} \to \mathrm{D}^{*+}\ell^-\overline{\nu_l}$ decays. Using HQET (Heavy Quark Effective Theory), an expression for the differential decay rate can be derived:

$$\frac{d\Gamma}{dw} = \frac{G_F^2}{48\pi^2}|V_{cb}^2||F(w)|^2 G(w) \ ; \ w = v_B.v_D \tag{41}$$

w is the 4-product of the B (v_B) and the D meson (v_D) velocities. G(w) is a kinematical factor and F(w) is the form factor describing the transition. At zero recoil (w=1) and for infinite quark masses, F(1) goes to unity. The strategy is then to measure $d\Gamma/dw$, to extrapolate at zero recoil and to determine $F(1) \times |V_{cb}|$.

The world average result (as given in PDG 2004) [31] is:

$$|V_{cb}| = (42.1 \pm 1.1 \pm 1.9_{F(1)})\ 10^{-3} \ = \ (42.1 \pm 2.2)\ 10^{-3} \quad \text{(exclusive)} \tag{42}$$

To evaluate $|V_{cb}|$, the value of F(1) = 0.91 $\pm$ 0.04 have been used [39, 40].

Determination of $|V_{cb}|$ using inclusive and exclusive methods. Combining these two determinations of $|V_{cb}|$ gives:

$$|V_{cb}| = (41.5 \pm 0.8)10^{-3} \quad (\text{exclusive} + \text{inclusive}) \tag{43}$$

The average has been obtained neglecting possible correlations between the two methods to determine of $|V_{cb}|$. This assumption is safe

from the experimental point of view, whereas detailed studies are still missing from theory side. It should be noted that the inclusive method is dominating the final precision on $|V_{cb}|$.

To conclude, it is important to remind that, as $|V_{cb}| = A\lambda^2$, the measurement of $|V_{cb}|$ allows the determination of A one of the four free parameters of the CKM matrix. Furthermore $|\mathrm{V}_{cb}|$ gives the scale of the Unitarity Triangle.

It is important to note also that $|\mathrm{V}_{cb}|$, today, is known with 2% accuracy. This achievement has to be considered as a legacy from LEP and CLEO experiments.

6.2 Determination of $|V_{ub}|$

The measurement of $|V_{ub}|$ is rather difficult because one has to suppress the large background coming from the more abundant semileptonic b to c quark transitions.

Several new determinations of the CKM element $|V_{ub}|$ are now available [31].

Determination of $|V_{ub}|$ using inclusive analyses. As for $|V_{cb}|$, the extraction of $|V_{ub}|$ from inclusive semileptonic decays is based on HQET implemented through OPE.

By using kinematical and topological variables, it is possible to select samples enriched in $b \to u\ell^-\overline{\nu_\ell}$ transitions. There are, schematically, three main regions in the semileptonic decay phase space to be considered:

- the lepton energy end-point region: $E_\ell > \frac{M_B^2 - M_D^2}{2M_B}$ (which was at the origin for the first evidence of $b \to u$ transitions)
- the low hadronic mass region: $M_X < M_D$ (pioneered by the DELPHI Coll. [41])
- the high q^2 region: $M_{\ell\nu}^2 = q^2 > (M_B - M_D)^2$.

in which the background from $b \to c\ell^-\overline{\nu_\ell}$ decays is small.

A summary of the different determinations of $|V_{ub}|$ is given in Figure 5. For the extraction of the CKM parameters we use the average calculated in [29], presented in [41] and given in Table 3 :

$$|\mathrm{V}_{ub}| \;=\; (40.9 \pm 4.6 \pm 3.6) \times 10^{-4} \quad \text{LEP-CLEO } (\textit{inclusive}) \quad (44)$$

Determination of $|V_{ub}|$ using exclusive analyses. The second method to determine $|V_{ub}|$ consists in the reconstruction of charmless semileptonic B decays: $\overline{\mathrm{B}} \to \pi(\rho)\ell\overline{\nu}$.

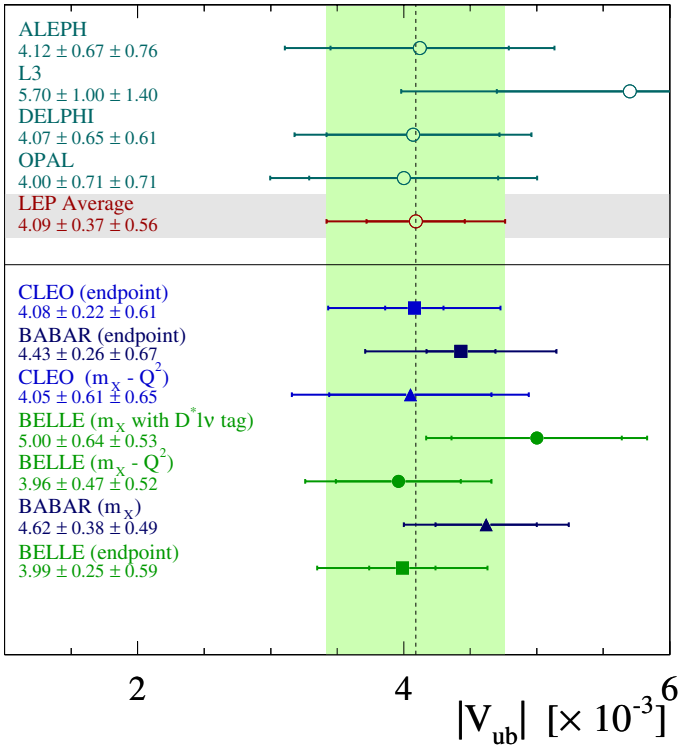

Figure 5. Summary of $|V_{ub}|$ inclusive measurements [31]. For the extraction of the CKM parameters we use the average calculated in [29], presented in [41] and given in Table 3.

Experimentally, the use of exclusive final states provides extra kinematical constraints for background suppression. Theoretically, the uncertainties are of a different nature as those already described in the inclusive analysis. The probability that the final state quarks form a given meson is described by form factors and, to extract $|V_{ub}|$ from actual measurements, the main problem rests in the determination of these hadronic form factors. As there is no heavy quark in the final state, symmetry arguments which were helpful to determine the form factor in $\overline{B} \to \mathrm{D}^* \ell \overline{\nu}$ decays cannot be invoked. Light-Cone Sum Rules can provide an evaluation at the 15-20% accuracy level. Lattice QCD calculations give a similar precision but these uncertainties are expected to be reduced in the near future. The main limitation in lattice calculations is that, at present, they can be used only in the high q^2 region.

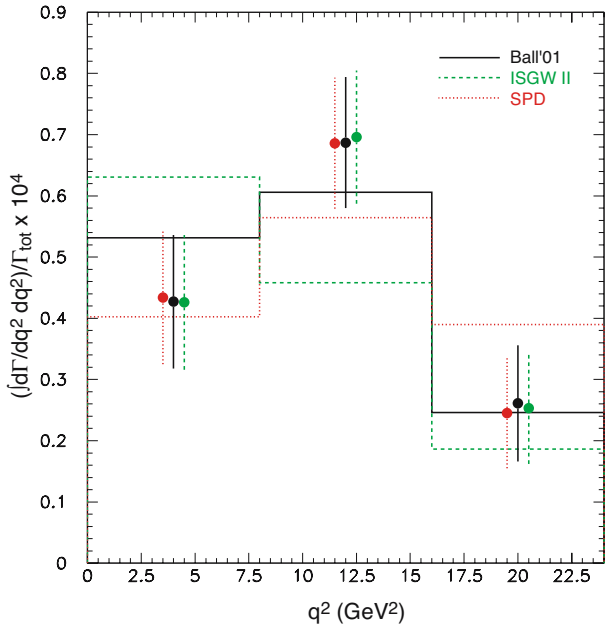

Figure 6. Differential branching fraction for $B^0 \to \pi^- \ell^+ \nu$ measured as a function of q^2, by the CLEO Coll., and compared with predicted values (histograms) for three models used to extract $|V_{ub}|$.

A summary of the different determinations of $|V_{ub}|$ can be found in [29] and [30]. The combined value of $|V_{ub}|$ is obtained by assuming that systematic uncertainties, attached to individual measurements, can be composed quadratically, for their uncorrelated components, and have correlated contributions, of similar size. This correlated part of the systematics arises mainly from the modelling of the $b \to u$ background. The relative theoretical error is similar for all measurements and, for the time being, the error from the BaBar measurement is used. The result is

$$\begin{aligned} |\mathrm{V}_{ub}| &= (33.8 \pm 2.4^{+3.7}_{-5.4}) \times 10^{-4} \\ &= (33.0 \pm 2.4 \pm 4.6) \times 10^{-4} \end{aligned} \tag{45}$$

The accuracy on the determination of $|\mathrm{V}_{ub}|$ using exclusive decays is limited by the theoretical uncertainty on hadronic form factor determination. An interesting analysis has been presented by the CLEO Collaboration at ICHEP02 [41], using the $\mathrm{B}^0_\mathrm{d} \to \pi^- \ell^+ \nu_\ell$ decay mode, which consists in extracting the signal rates in three independent regions of q^2. In this way it is possible to discriminate between models. The fit shows

that the ISGW II model is compatible with data at only 1% probability level. This approach could be used, in future, to reduce the importance of theoretical errors, considering that the ISGW II gave, at present, the further apart V_{ub} determination [41].

Determination of $|V_{ub}|$ using inclusive and exclusive methods. Combining the two determinations of $|V_{ub}|$ (44,45), we obtain, in practice, almost a Gaussian p.d.f. corresponding to:

$$|V_{ub}| = (35.7 \pm 3.1) \times 10^{-4}. \quad (46)$$

New and more precise results from Belle and Babar Collaborations will much improve the present situation.

6.3 Measurements of $B^0 - \overline{B^0}$ oscillations

Measurements of the $B^0_d - \overline{B^0_d}$ oscillation frequency: Δm_d. The probability that a B^0 meson oscillates into a $\overline{B}^0$ or remains as a B^0 is given in Eq. 21.

The measurement of Δm_d has been the subject of an intense experimental activity during the last ten years. Results are available which correspond to the combination of 27 analyses, using different event samples, performed by the LEP Coll./SLC/CDF/B-Factories experiments.

A typical proper time distribution is shown in Figure 8. The oscillating behaviour is clearly visible.

Figure 7 gives the results for Δm_d, obtained by each experiment and the overall average [31]:

$$\Delta m_d \quad = \quad (0.502 \pm 0.007)\ ps^{-1}. \quad (47)$$

The accuracy is of about 1%. The B-factories have the main contribution to this accuracy. Improvements can still be expected from these facilities and they are expected to reach a few per mill precision.

Search for $B^0_s - \overline{B^0_s}$ oscillations. As the B^0_s meson is expected to oscillate more than 20 times faster than the B^0_d ($\Delta m_s/\Delta m_d \propto 1/\lambda^2$) and as B^0_s mesons are less abundantly produced, the search for $B^0_s - \overline{B^0_s}$ oscillations is more difficult. The observation of fast oscillations requires the highest resolution on the proper time and thus on the B^0_s decay length.

for $B^0_s - \overline{B}^0_s$ oscillations has been observed so far.

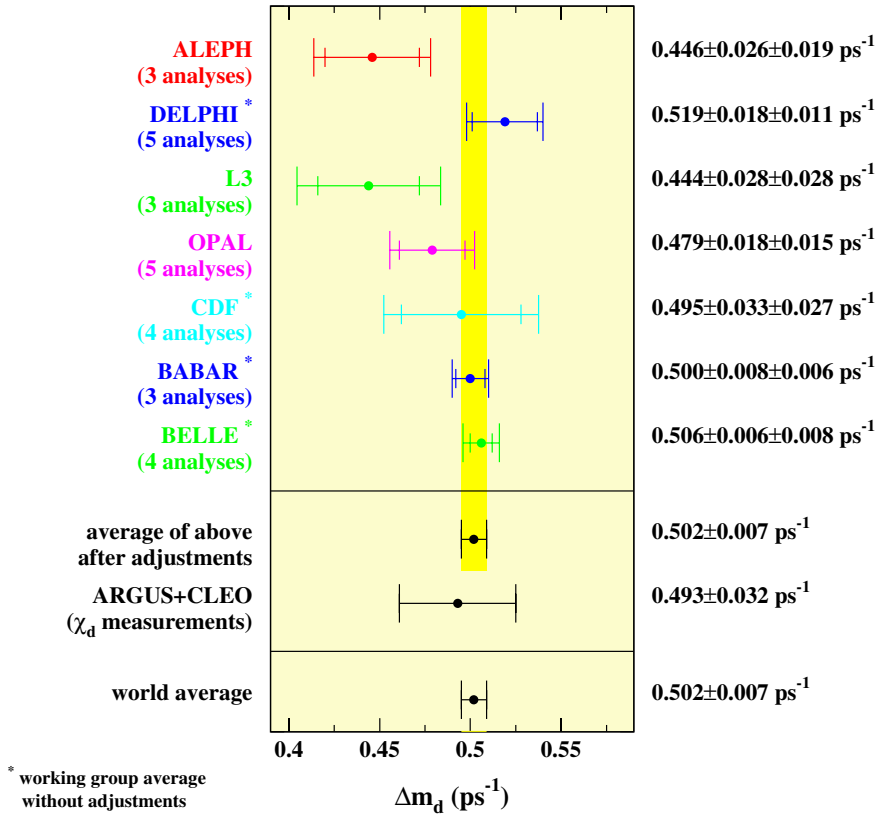

Figure 7. Summary of Δm_d measurements [31].

The method used to measure or to put a limit on Δm_s consists in modifying Eq. (21) in the following way [42]:

$$1 \pm \cos(\Delta m_s t) \rightarrow 1 \pm \mathcal{A} \cos(\Delta m_s t). \tag{48}$$

$\mathcal{A}$ and its error, $\sigma_{\mathcal{A}}$, are measured at fixed values of Δm_s, instead of Δm_s itself. In case of a clear oscillation signal, at a given frequency, the amplitude should be compatible with $\mathcal{A} = 1$ at this frequency. With this method it is easy to set a limit. The values of Δm_s excluded at 95% C.L. are those satisfying the condition $\mathcal{A}(\Delta m_s) + 1.645\ \sigma_{\mathcal{A}(\Delta m_s)} < 1$.

With this method, it is easy also to combine results from different experiments and to treat systematic uncertainties in the usual way since,

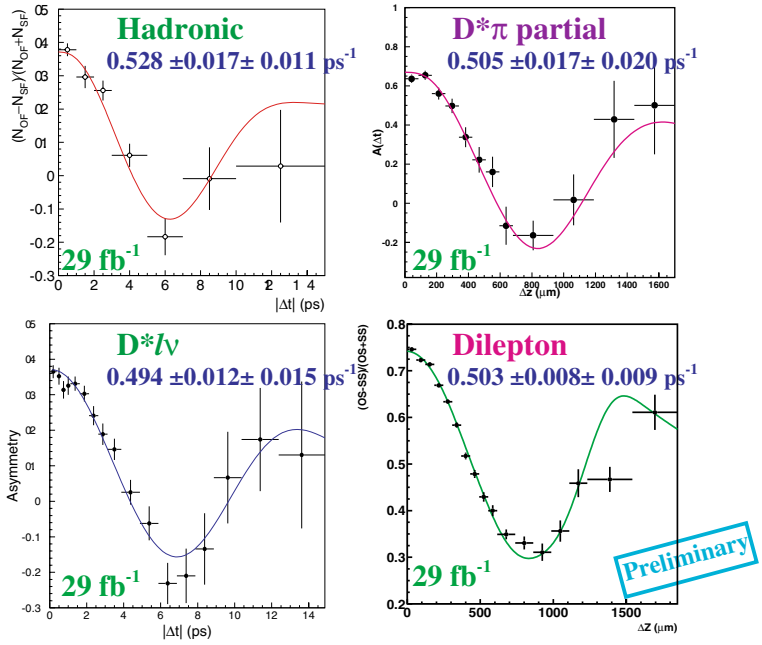

Figure 8. The plots show the $B^0_d - \overline{B}^0_d$ oscillations (Belle Coll.). The points with error bars are the data. The result of the fit gives the value for Δm_d.

for each value of Δm_s, a value for $\mathcal{A}$ with a Gaussian error $\sigma_{\mathcal{A}}$, is measured. Furthermore, the sensitivity of a given analysis can be defined as the value of Δm_s corresponding to $1.645\ \sigma_{\mathcal{A}}(\Delta m_s) = 1$ (using $\mathcal{A}(\Delta m_s) = 0$), namely supposing that the "true" value of Δm_s is well above the measurable value.

During last years, impressive improvements in the analysis techniques allowed to increase the sensitivity of the search for $\mathrm{B}^0_s - \overline{\mathrm{B}}^0_s$ oscillations. Figure 9 gives details of the different Δm_s analyses. The combined result of LEP/SLD/CDF analyses [31] (Figure 10) corresponds to:

$$\begin{aligned} \Delta m_s \quad &> 14.5\ ps^{-1} \quad at \quad 95\% \quad C.L. \\ \text{with a sensitivity :} \quad &\Delta m_s = 18.3\ ps^{-1}. \end{aligned} \tag{49}$$

The present combined limit implies that B^0_s oscillate at least 30 times faster than B^0_d mesons. Taking into account only the λ dependence of the ratio $\Delta m_d/\Delta m_s$ (eq. 35), this factor would be about 20. The present limit gives strong constraints on the $\overline{\rho}$ parameter whose value ends up to be about 0.2.

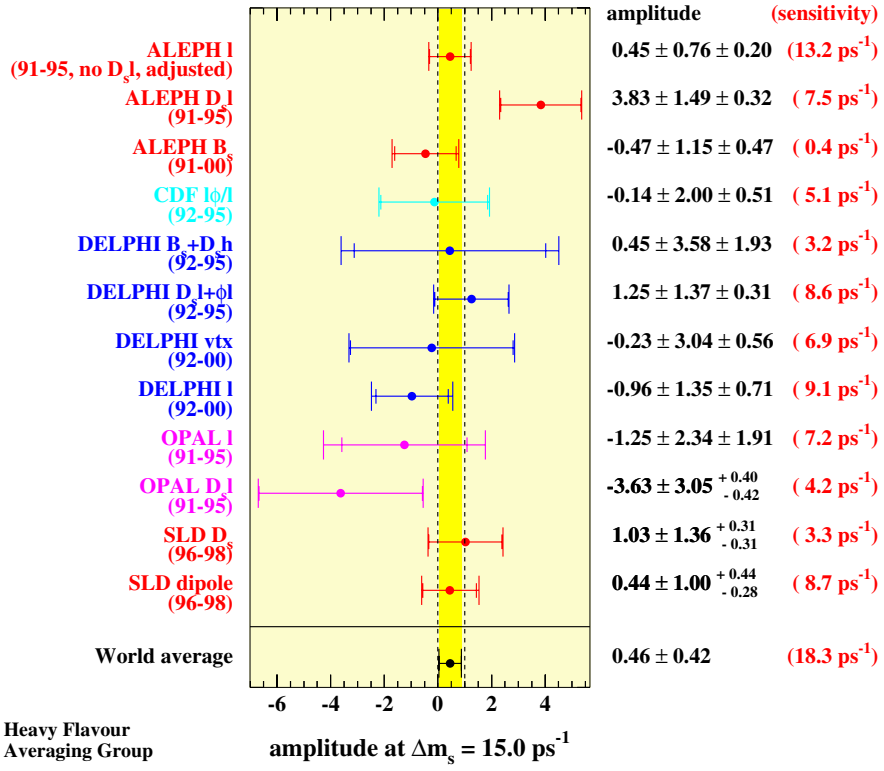

Figure 9. B_s^0 oscillation results. Values of the fitted amplitude at $\Delta m_s = 15$ ps^{-1} and of the sensitivity obtained by each experiment [31].

The significance of the "bump" appearing around 17 ps^{-1} is about 2.2 σ and no claim can be made for the observation of $B_s^0 - \overline{B_s^0}$ oscillations. Tevatron experiments are expected to measure soon these oscillations.

6.4 Some theoretical inputs: B_K, $f_B\sqrt{\hat{B}_B}$ and ξ

Constraints on $\overline{\rho}$ and $\overline{\eta}$ depend also upon three parameters which are related to the strong interaction operating in the non-perturbative regime: $f_B\sqrt{\hat{B}_B}$, ξ and B_K.

Expressions for these constraints have been given, respectively, in Eqs. (34), (35) and (36). Important improvements have been achieved during

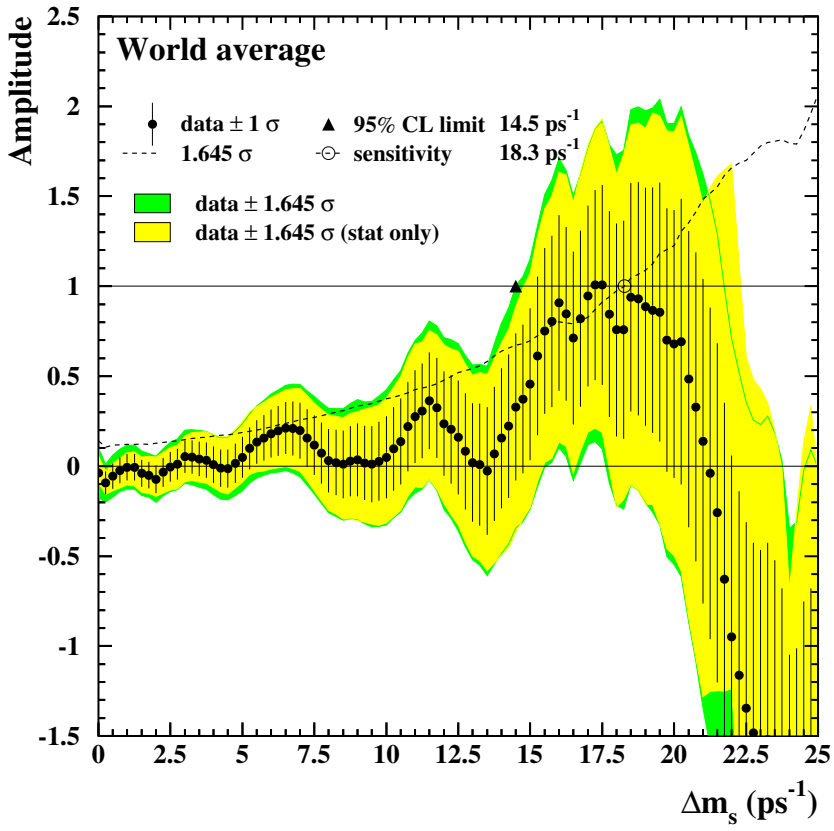

Figure 10. The plot [31] gives combined Δm_s results from LEP/SLD/CDF analyses shown as an amplitude versus Δm_s plot. The points with error bars are the data; the lines show the 95% C.L. curves (darker regions correspond to the inclusion of systematics). The dotted curve corresponds to the sensitivity.

the last few years in the evaluation of these parameters in the framework of Lattice QCD and a world-wide effort is organized in view of having precise determinations of these parameters. As a consequence, in this phenomenological analysis, only most recent results from Lattice QCD are used.

Brief introduction to Lattice QCD (LQCD) . Lattice QCD (LQCD) was invented about 25 years ago by K. Wilson [43].

Perturbation theory can be seen as a tool to perform functional integrals by which all vacuum expectation values of the quantum fields can be expressed. LQCD approach consists in a numerical evaluation of the functional integrals. It needs a discretization of the four-dimensional space-time by introducing a basic length, the lattice spacing (often indicated as a). So LQCD does not introduce new parameters or field variables in the discretization and it retains the same properties as QCD. In this sense, it is correct to say, that LQCD is not a model, as quark models for example, and therefore physical quantities can be computed from first principles without arbitrary assumptions. The only input parameters are the strong coupling constant and the six quark current masses.

Statistical errors.
Considering N points in each direction, the lattice will have a volume $(\mathrm{N\ a})^4$ (having so two natural cutoffs: a finite space resolution and a finite volume). The standard integrals are sampled over a finite net of points, whereas the functional integrals are sampled over a finite set of functions (or configurations). The vacuum expectation values are obtained by "averaging" over all the configurations. Those evaluations are done using MonteCarlo techniques. In this spirit, LQCD simulations are theoretical experiments carried out by numerical integration of the functional integral by MonteCarlo techniques. In this respect uncertainties on output quantities are evaluated following criteria which are very close to those used in experimental measurements. Results are obtained with "statistical errors", i.e. uncertainties originated by stochastic fluctuations, which may be reduced by increasing the sample of gluon-field configurations on which averages are performed. It is very reasonable to assume that the statistical fluctuations have a Gaussian distribution.
For several quantities statistical errors have been reduced to the percent level (or even less). However most of the results are affected by systematic effects.

Systematic errors.
Systematic uncertainties come from discretization effects, finite volume effects, the treatment of heavy quarks, chiral extrapolation and quenching. Errors coming from the discretization and from the finite volume can be addressed by brute-force improvements of numerical simulations or by improvements in the discretization procedures.
The quenched approximation is obtained by turning off virtual quark loops. An important consequence of this approximation is that the potential between a quark/antiquark pair depends on this approximation. In the full theory, at large distance, there is a screened potential between two hadrons because the string breaks by the creation of a $q\overline{q}$ pair. In quenched LQCD the string does not couple to such pairs and the long distance behaviour of the two theories is rather different. This problem is not so important since, for the long distance scale which matters in hadronic physics, and in which we are interested, there is a "natural" cutoff of about one Fermi due to confinement.
It is reasonable to expect that quenching corrections are lying between 10-20% for most of evaluated physical quantities.
Because of computing limitations, most numbers have been obtained in the quenched approximation. Theoretical estimates and some preliminary results in the (partially) unquenched case are also available and are used to estimate the corresponding systematic error of quenched results. These calculations are usually performed with two light quarks in the fermion loops, at values of the light-quark masses larger than the physical values and an extrapolation in these masses is required. Calculations are generally made at few values of the lattice spacing and thus contain discretization errors. An estimate of quenching errors is obtained by comparing quenched and unquenched results at similar values of the lattice spacing.
Another important issue is related to the chiral extrapolation. In fact it is difficult to simulate realistically light quarks, with their physical masses, and calculations are usually made for a set of (valence) quark masses, ranging from about $m_s/2$ to 2 m_s. The results need then to be interpolated or extrapolated. Similar extrapolation needs to be done, in partially quenched calculations, considering the range of sea quark masses used. The problem arises since there are logarithmic dependences in physical quantities as the valence and/or the sea quark mass are extrapolated to their physical values (divergences in some cases if masses vanish). In practice different extrapolations can be performed if one considers or not these terms. The JLQCD collaboration finds [44] that these different extrapolations tend to decrease the value of f_{B_d} relative to f_{B_s}. At present a reasonable view [45, 39] is to allow a decrease

of f_{B_d} by -10% and a negligible change in f_{B_s}.

For the present phenomenological analysis, the following values and errors have been used

$$f_{B_d}\sqrt{\hat{B}_{B_d}} = (223 \pm 33 \pm 12)\ MeV \tag{50}$$

$$\xi = \frac{f_{B_s}\sqrt{\hat{B}_{B_s}}}{f_{B_d}\sqrt{\hat{B}_{B_d}}} = 1.18 \pm 0.04 \pm 0.06$$

$$\hat{B}_K = 0.86 \pm 0.06 \pm 0.14$$

These estimates have to be considered as conservative, since they assume a maximal effect due to chiral extrapolation, reflected in the last error. These last errors are taken as flat distributions.

A detailed description on how these values have been obtained can be found in [29]. The CKM-LDG Group [32] is taking care of these averages.

6.5 Determination of $\sin(2\beta)$ from CP asymmetry in $J/\psi K^0$ decays.

BaBar and Belle collaborations have recently updated their measurements. The world average is [31]:

$$\sin(2\beta) = 0.739 \pm 0.048 \tag{51}$$

All details concerning the analyses techniques are described in these proceeding by the seminar corresponding to U. Mallik.

7. Determination of the Unitarity Triangle parameters

In this section we give the results for the quantities defining the Unitarity Triangle, assuming the validity of the Standard Model: $\overline{\rho}$, $\overline{\eta}$, $\sin(2\beta)$, $\sin(2\alpha)$ and γ as well as for other quantities as Δm_s, f_B and $\hat{B}_K$. The inputs used are summarised in Table 3 (see Section 4.3).
For more details and concerning latest results see [46].

7.1 Fundamental test of the Standard Model in the fermion sector

The most crucial test consists in the comparison between the region selected by the measurements which are sensitive only to the sides of the

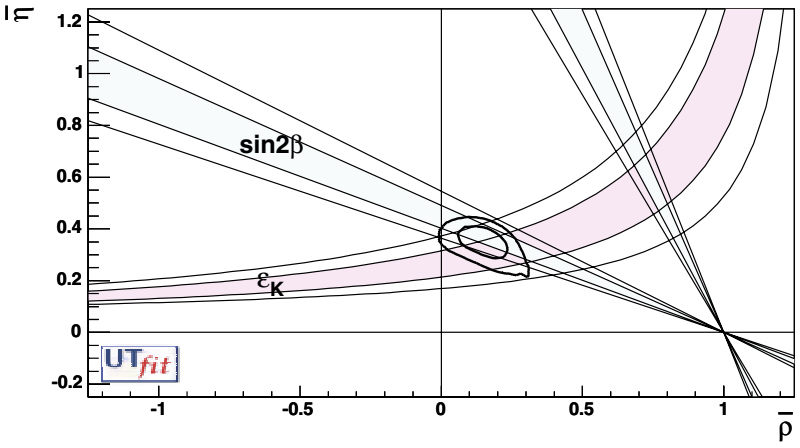

Figure 11. The allowed regions for $\overline{\rho}$ and $\overline{\eta}$ (contours at 68%, 95%) as selected by the measurements of $|V_{ub}|\,/\,|V_{cb}|$, ΔM_d, and by the limit on $\Delta M_s/\Delta M_d$ are compared with the bands (at 68% and 95% C.L.) from the measurements of CP violating quantities in the kaon ($|\varepsilon_K|$) and in the B ($\sin(2\beta)$) sectors.

Unitarity Triangle (semileptonic B decays and $\mathrm{B}^0 - \overline{\mathrm{B}^0}$ oscillations) and the regions selected by the direct measurements of CP violation in the kaon ($|\varepsilon_K|$) or in the B ($\sin(2\beta)$) sectors. This test is shown in Figure 11. It can be translated quantitatively through the comparison between the values of $\sin(2\beta)$ obtained from the measurement of the CP asymmetry in $J/\psi K^0$ decays and the one determined from "sides" measurements:

$$\begin{aligned} \sin(2\beta) = &\quad 0.685 \pm 0.047\ [0.547 - 0.770] \text{ at } 95\%\ C.L. \quad \text{sides only} \\ \sin(2\beta) = &\quad 0.739 \pm 0.048\ [0.681 - 0.787] \text{ at } 95\%\ C.L. \quad \mathrm{J}/\psi\mathrm{K}^0. \end{aligned} \tag{52}$$

The spectacular agreement between these values illustrates the consistency of the Standard Model in describing CP violation phenomenon in terms of one single parameter $\overline{\eta}$. It is also an important test of the OPE, HQET and LQCD theories which have been used to extract the CKM parameters.

It has to be noted that this test is significant provided the errors on $\sin(2\beta)$ from the two determinations are comparable.

Corresponding results, for the unitarity triangle parameters, are given in Table 5:

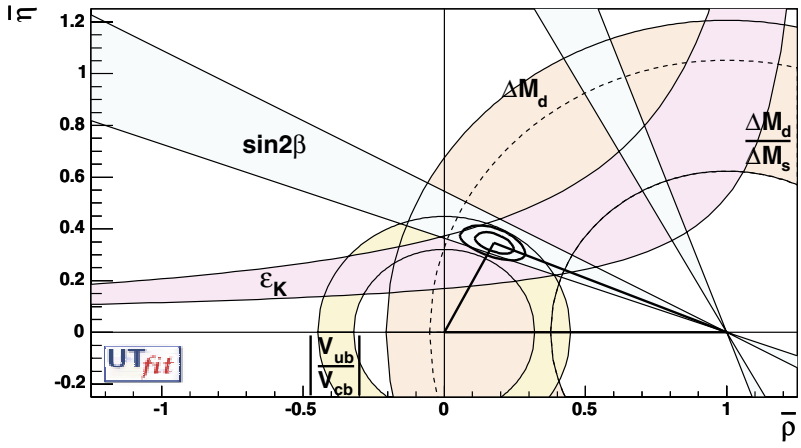

Figure 12. Allowed regions for $\overline{\rho}$ and $\overline{\eta}$ using the parameters listed in Table 3. The closed contours at 68% and 95% probability are shown. The full lines correspond to 95% probability regions for the constraints, given by the measurements of $|V_{ub}|\,/\,|V_{cb}|$, $|\varepsilon_K|$, Δm_d, Δm_s and $\sin(2\beta)$. The dotted curve corresponds to the 95% upper limit obtained from the experimental study of $B_s^0 - \overline{B_s^0}$ oscillations.

Parameter	68%	95%	99%
$\overline{\eta}$	$0.346\ ^{+0.039}_{-0.043}$	(0.227-0.416)	(0.099-0.437)
$\overline{\rho}$	0.153 ± 0.061	(0.030-0.325)	(-0.012-0.368)
$\sin(2\beta)$	0.685 ± 0.047	(0.547-0.770)	(0.280-0.806)
$\sin(2\alpha)$	-0.01 ± 0.35	(-0.85-0.83)	-
$\gamma[^\circ]$	65.3 ± 9.5	(38.9-84.8)	(15.8-90.0)

Table 5. Values and probability ranges for the unitarity triangle parameters when the constraints from $|\varepsilon_K|$ and $\sin(2\beta)$ measurements are not used.

7.2 Determination of the Unitarity Triangle parameters : $\overline{\eta}$, $\overline{\rho}$, $\sin(2\beta)$, $\sin(2\alpha)$, γ

By using all five available constraints ($|V_{ub}|\,/\,|V_{cb}|$, Δm_d, $\Delta m_s/\Delta m_d$, $|\varepsilon_K|$ and $\sin(2\beta)$), the results given in Table 6 are obtained.
Figures 12 and 13 show, respectively, the corresponding selected region in the ($\overline{\rho}$, $\overline{\eta}$) plane and the p.d.f. for the Unitarity Triangle parameters.

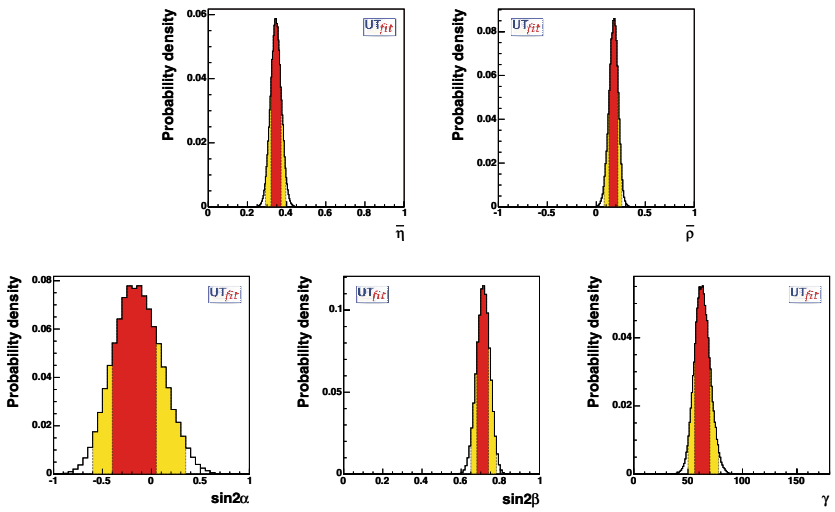

Figure 13. From top left to bottom, the p.d.f. for $\overline{\eta}$, $\overline{\rho}$, $\sin(2\alpha)$, $\sin(2\beta)$ and γ. The red (darker) and the yellow (clearer) zones correspond respectively to 68% and 95% of the normalised area. All available constraints have been used.

Parameter	68%	95%	99%
$\overline{\eta}$	0.342 ± 0.026	(0.291-0.396)	(0.272-0.415)
$\overline{\rho}$	0.174 ± 0.047	(0.076-0.260)	(0.045-0.293)
$\sin(2\beta)$	0.697 ± 0.035	(0.637-0.761)	(0.619-0.781)
$\sin(2\alpha)$	-0.15 ± 0.25	(-0.62-0.34)	(-0.73-0.50)
$\gamma[^\circ]$	61.1 ± 7.8	(48.6-76.0)	(43.2-82.9)

Table 6. Values and probability ranges for the unitarity triangle parameters obtained by using all five available constraints: $|V_{ub}|/|V_{cb}|$, Δm_d, $\Delta m_s/\Delta m_d$, $|\varepsilon_K|$ *and* $\sin(2\beta)$.

Indirect versus direct determination of the Unitarity Triangle angles. The value of $\sin(2\beta)$ was predicted, before its first direct measurement was obtained, by using all other available constraints, ($|V_{ub}|/|V_{cb}|$, $|\varepsilon_K|$, Δm_d and Δm_s). The "indirect" [16] determination has improved regularly over the years. Figure 14 shows this evolution for the "indirect" determination of sin2β which is compared with the recent determinations of $\sin(2\beta)$ from direct measurements.

[16] in the following, for simplicity, we will note as "direct" ("indirect"), the determination of $\sin(2\beta)$ from $A_{CP}(J/\psi K^0)$ (other constraints).

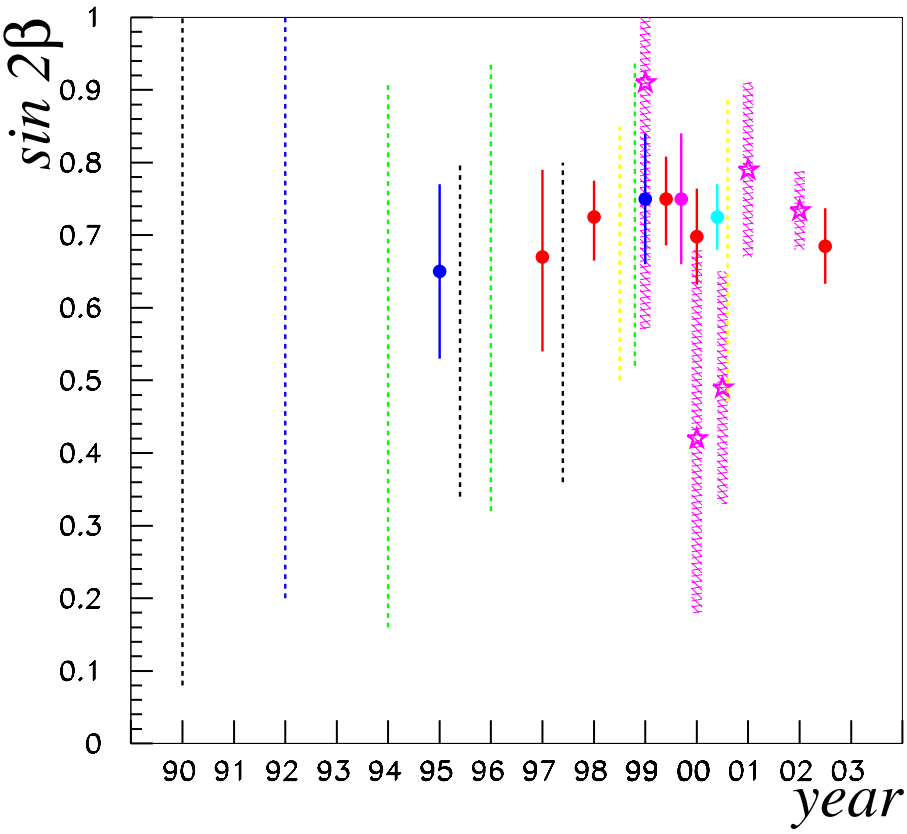

Figure 14. Evolution of the "indirect" determination of $\sin(2\beta)$ over the years. From left to right, they correspond to the following papers [38]: DDGN90, LMMR92, AL94, CFMRS95, BBL95, AL96, PPRS97, BF97, BPS98, PS98, AL99, CFGLM99, CPRS99, M99, CDFLMPRS00, B.et.al.00, HLLL00 and the value presented in this document. The dotted lines correspond to the 95% C.L. regions (the only information given in those papers). The larger bands (from year '99) correspond to values of $\sin(2\beta)$ from direct measurements ($\pm 1\sigma$).

This test should be repeated with other constraints.

The values for γ and $\sin 2\alpha$ given in Table 6 has to be taken as predictions for future measurements. A strong message is given for instance for the angle γ. The indirect determination of the angle γ is known

with an accuracy of about 10%. It has to be stressed that, with present measurements, the probability that γ is greater than 90° is only 0.003%.

7.3 Determination of other important quantities

In previous sections we have seen that we can get distributions for the different unitarity triangle parameters and how it can be instructive to remove from the fitting procedure the external information on the value of one (or more) of the constraints.

In this section we get the distributions for the values of other quantities, entering into the Standard Model expressions for the constraints, such as the hadronic parameters, or of a constraint as Δm_s. In case of the hadronic parameters, for instance, it is instructive to remove, from the fit, in turn, their external information. The idea is to compare the uncertainty on a given quantity, determined in this way, to its present experimental or theoretical error. This comparison allows to quantify the importance of present determinations of the different quantities to define the limits of the allowed region for the unitarity triangle parameters.

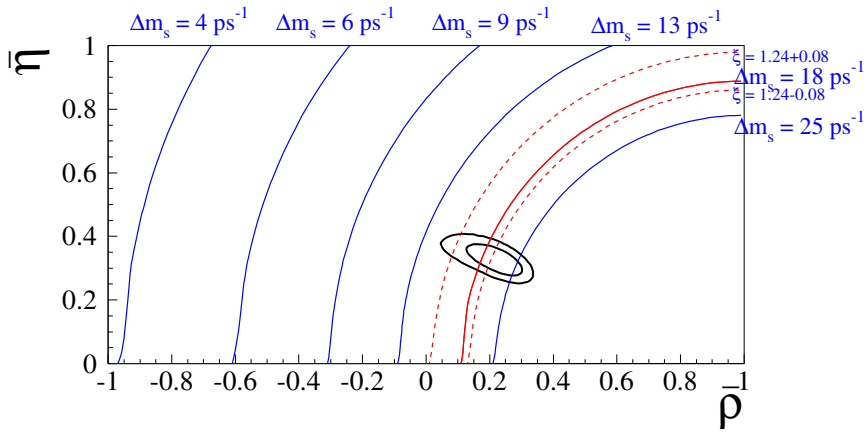

Figure 15. The allowed regions for $\overline{\rho}$ and $\overline{\eta}$ using the constraints given by the measurements of $|\varepsilon_K|$, $|V_{ub}|/|V_{cb}|$, Δm_d and $\sin(2\beta)$ at 68% and 95% probability are shown by the closed contour lines. The different continuous circles correspond to fixed values of Δm_s. Dashed circles, drawn on each side of the curve corresponding to $\Delta m_s = 18.0\ ps^{-1}$, indicate the effect of a variation by ± 0.08 on ξ.

The expected distribution for Δm_s. Figure 15 shows the allowed region for $\overline{\rho}$ and $\overline{\eta}$ obtained with all the constraints and how the constraint coming from the study of B_s^0–$\overline{B_s^0}$ mixing acts in this plane. A lower limit at 95% C.L. on Δm_s will exclude, at that degree of confidence, the $\overline{\rho}$-$\overline{\eta}$ region situated on the left of the corresponding curve.

Parameter	68%	95%	99%
Δm_s(including Δm_s) [ps^{-1}]	18.4±1.6	(15.4-21.2)	(14.6-25.4)
Δm_s(without including Δm_s) [ps^{-1}]	20.2 ± 3.0	(14.4-26.8)	(13.2-29.2)

Table 7. Δm_s central values and ranges corresponding to defined levels of probability, obtained when including or not the information from the experimental amplitude spectrum $\mathcal{A}(\Delta m_s)$.

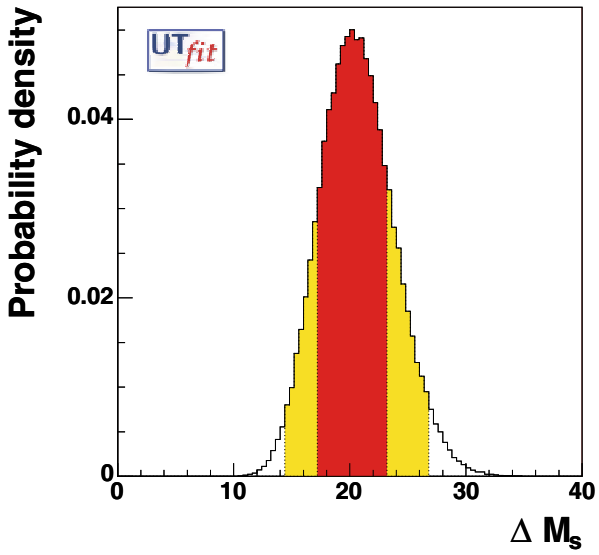

Figure 16. Δm_s probability distributions. The information from $B_s^0 - \overline{B_s^0}$ oscillations is not used.

It is also possible to extract the probability distribution for Δm_s, which is shown in Figure 16. Corresponding results are given in Table 7. Present analyses at LEP/SLD, with a sensitivity at 19.2 ps^{-1} are

situated in a high probability region for a positive signal (as the "signal bump" appearing around 17.5 ps^{-1}).

Accurate measurements of Δm_s are thus expected soon from the TeVatron.

Determination of $f_{B_d}\sqrt{\hat{B}_{B_d}}$ and $\hat{B}_K$. The value of $f_{B_d}\sqrt{\hat{B}_{B_d}}$ can be obtained by removing [17] the theoretical constraint coming from this parameter in the expression of the B^0_d–$\overline{B^0_d}$ oscillation frequency Δm_d. The main conclusion of this study is that $f_{B_d}\sqrt{\hat{B}_{B_d}}$ is measured with an accuracy which is better than the current evaluation from lattice QCD, given in Section 6.4. Results are summarized in Table 8. This shows that the present CKM fit, when all the available constraints are used, is, in practice, weakly dependent on the exact value assumed for the uncertainty on $f_{B_d}\sqrt{\hat{B}_{B_d}}$.

The parameter $\hat{B}_K$ can be also determined. Results are also summarized in Table 8. They indicates that values of $\hat{B}_K$ smaller than 0.5 (0.3) correspond to 0.6% (5×10^{-6}) probability while large values of $\hat{B}_K$ are compatible with the other constraints over a large domain. The present estimate of $\hat{B}_K$, from lattice QCD, with a 15% relative error (Table 3) has thus a large impact in the present analysis.

Parameter	68%	95%	99%
$f_{B_d}\sqrt{\hat{B}_{B_d}}$(MeV)	217 ± 12	(196-245)	(190-258)
$\hat{B}_K$	$0.69^{+0.13}_{-0.08}$	(0.53-0.96)	(0.49-1.09)

Table 8. Values and probability ranges for the non perturbative QCD parameters, if the external information (input) coming from the theoretical calculation of these parameters is not used in the CKM fits

7.4 Evolution on the precision on $\overline{\rho}$ and $\overline{\eta}$ over the last 15 years

The evolution of our knowledge concerning the allowed region in the ($\overline{\rho}$, $\overline{\eta}$) plane is shown in Figure 17. The reduction of the size of these regions, from years 1995 to 2000, is essentially due to the measurements of the sides of the Unitarity Triangle and to the progress in OPE, HQET

[17] Technically we assume a uniform distribution in a range which is much larger than the possible values taken by the parameters.

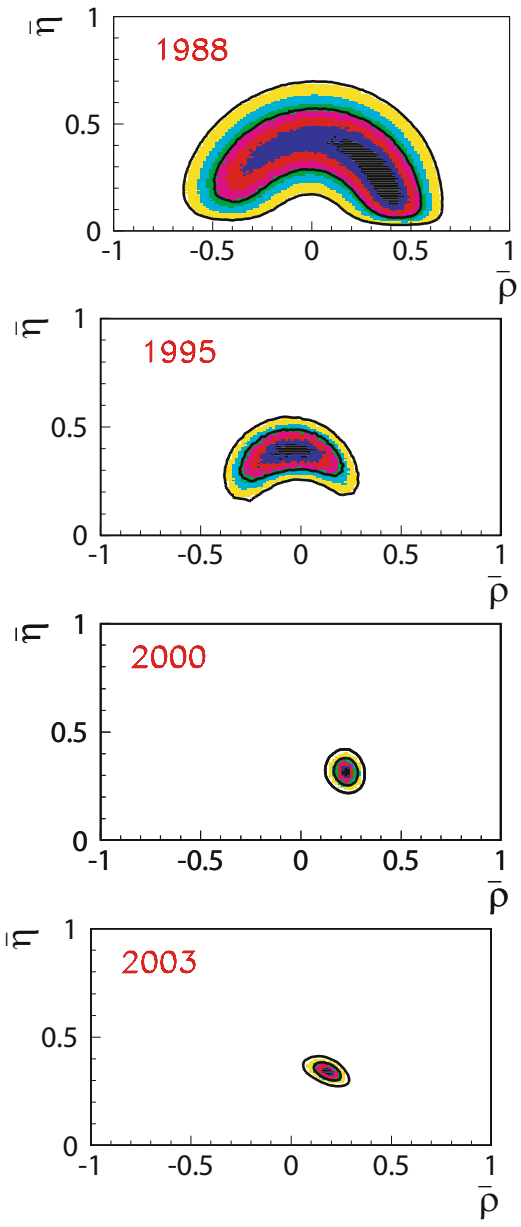

Figure 17. Evolution during the last 15 years of the allowed regions for $\overline{\rho}$ and $\overline{\eta}$ (contours at 68% and 95% probability are indicated). The very last results (updated till Winter 2004) are shown in Figure 12 and in Table 6.

and lattice QCD theoretical parameters determinations. The additional reduction, from years 2000 to 2003, which mainly concerns $\overline{\eta}$, is essentially driven by the measurement of $\sin(2\beta)$ through the CP violation asymmetry in J/ψ K^0 decays.

7.5 Dulcis in fundo : the new-comers

The huge statistics collected the B-factories allow the measurements of new CP-violating quantities. Direct measurements of γ, $\sin(2\beta+\gamma)$ and $\sin 2\alpha$ are now available :

- determination of $\sin 2\alpha$ using charmless $\pi\pi$ events,
- determination of $\sin(2\beta+\gamma)$ using $D^{(*)}\pi$ events,
- determination of γ using DK events.

We do not enter in any details for these analyses which are described in U. Mallik lectures.

The Figure 18 shows the impact of these new measurements to provide additional constraints in the $\overline{\rho}-\overline{\eta}$ plane. More details are given in [46].

These plots show the potentialities of B-factories, considering that additional measurements will be available, in a near future (about 2 years), with more than four times the statistics.

8. Conclusions

Flavour physics in the quark sector is entered in its mature age. Many and interesting results have been produced during the last 15 years. Traditional main players (LEP/SLD/CLEO) delivered results until this year, while B factories are moving B studies into the era of precision physics.

Many quantities have already been measured with a good precision. $|V_{cb}|$ is today known with a relative precision better than 2%. In this case, not only, the decay width has been measured, but also some of the non-perturbative QCD parameters entering into its theoretical expressions. It is a great experimental achievement and a success of the theory description of the non-perturbative QCD phenomena in the framework of the OPE. Many different methods, more and more reliable, are now available for determining the CKM element $|V_{ub}|$. The relative precision, today, is of about 10% and will be certainly improved in a near future at B-factories. The time dependence behaviour of $B^0-\bar{B}^0$ oscillations has been studied and precisely measured in the B_d^0 sector. The oscillation frequency Δm_d is known with a precision of about 1%. $B_s^0-\bar{B}_s^0$ oscillations have not been measured sofar, but this search has pushed the

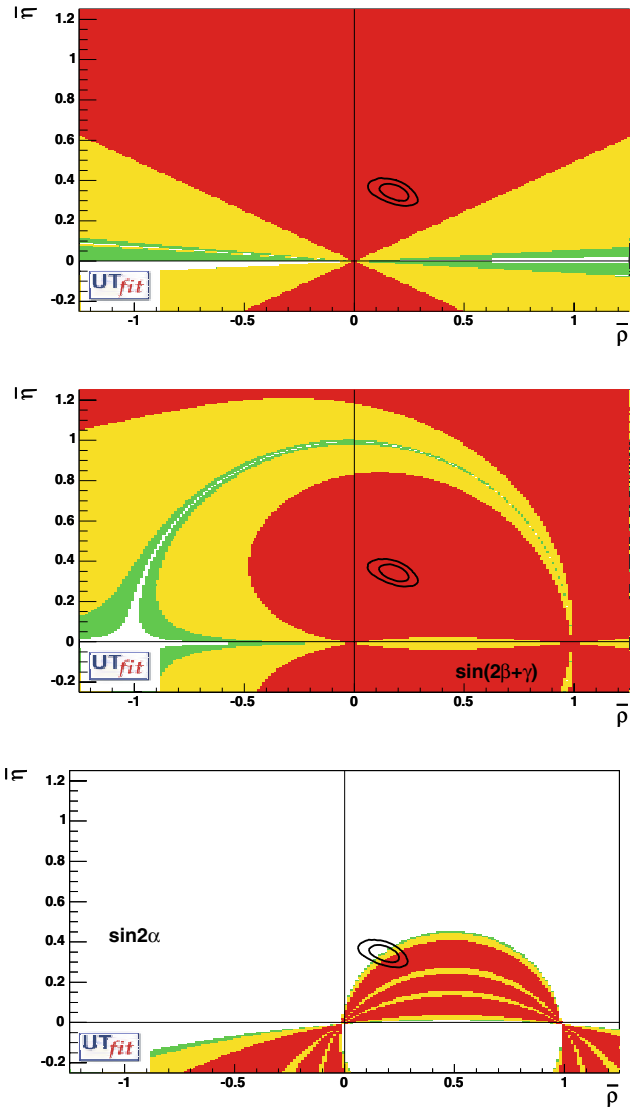

Figure 18. From top to bottom, the allowed region on the $\overline{\rho} - \overline{\eta}$ plane as selected by the direct measurement of γ, $\sin(2\beta + \gamma)$ and $\sin 2\alpha$ The red (darker), the yellow (clear) and green (clearer) zones correspond respectively to 68%, 95% and 99% of the normalised area. Contours at 68% and 95% probability selected using all the other available constraints are also shown.

experimental limit on the oscillation frequency Δm_s well beyond any initial prediction. Today we know that B_s^0 oscillate at least 30 times faster than B_d^0 mesons. The frequency of $B_s^0 - \bar{B}_s^0$ oscillations should be soon measured at the TeVatron. Nevertheless, the impact of the actual limit on Δm_s for the determination of the unitarity triangle parameters is crucial.

Many B decay branching fractions and relative CP asymmetries have been measured at B-factories. The outstanding result is the determination of sin 2β from B hadron decays into charmonium-K^0 final states. On the other hand many other exclusive hadronic rare B decays have been measured and constitute a gold mine for weak and hadronic physics, allowing to already extract different combinations of the unitarity triangle angles.

The unitarity triangle parameters are today known with a good precision. A crucial test has been already done: the comparison between the unitarity triangle parameters, as determined with quantities sensitive to the sides of the triangle (semileptonic B decays and oscillations), and the measurements of CP violation in the kaon (ϵ_K) and in the B (sin2β) sectors. The agreement is "unfortunately" excellent. The Standard Model is "Standardissimo": it is also working in the flavour sector. This agreement is also an important test of the OPE, HQET and LQCD theories which have been used to extract the CKM parameters.

The good news is that all these tests are at best at about 10% level. The current and the next facilities can surely push these tests to a 1% accuracy. It is important to note that charm physics can play an important role in this respect (providing a laboratory for LQCD) and the Charm-factory (CLEO-C) will play a central role for these issues.

9. Acknowledgements

I would like to thank the organizers, and especially Gerard Smadja, of the Cargèse School for the invitation and for having set up a very interesting school in a nice atmosphere.

I really would like to thank Vittorio Lubicz, Patrick Roudeau, Marie-Hélène Schune and Luca Silvestrini for the enlighting discussions and suggestions both in preparing these lectures and the document. Many thanks to Marcella Bona, Maurizio Pierini and Fabrizio Parodi, the "UT fitters" for providing me all the material for the determination of the CKM parameters.

References

[1] N. Cabibbo, *Phys. ReV. Lett.* **10** (1963) 351.

[2] M. Kobayashi and T. Maskawa,*Prog. Theor. Phys.* **49** (1973) 652.

[3] A.J. Buras, M.E. Lautenbacher and G. Ostermaier, *Phys. ReV.* **D50** (1994) 3433.

[4] C.M.G. Lattes, H. Muirhead, C.F. Powell and G.P. Occhialini, *Nature* **159**, 694 (1947).

[5] G.D. Rochester and C.C. Butler, *Nature* **160**, 855 (1947).

[6] L. Leprince-Ringuet and M. Lhéritier, *Comptes Rendues de l'Academie des Sciences* **219** (1944), 618. Presented à l'Académie de Sciences in the Séance du 17 Juin 1944.

[7] R.W. Thompson et al. *Phys. ReV.***83** (1951), 175.

[8] M. Gell-Mann, *Phys. ReV.* **92** (1953), 833.

[9] M. Gell-Mann, *Phys. Lett.* **8** (1964), 214.

[10] G. Zweig, CERN-Report TH-401 (1964) and TH-412 (1964): publised in *Developements in the Quark Theory of Hadrons, A Reprint Collection*, Vol I: 1964-1978, ed. by D.B. Lichtenberg and S.P. Rosen (Hadronic Press, Ma., 1980).

[11] B.J. Bjorken and S.L. Glashow, *Phys. Lett.* **11** (1964) 255.
Y. Hara, *Phys. ReV.* **B701** (1964) 134.
Z. Maki and Y. Ohnuki, *Prog. Theor. Phys.* **32** (1964) 144.

[12] S.L Glashow, J.Iliopoulos and L. Maiani, *Phys. ReV.* **D2** (1970) 1285.

[13] J.J. Aubert et al., BNL Spectrometer Coll., *Phys. ReV. Lett.* **33** (1974) 1404.
J.E. Augustin et al., SPEAR Coll., *Phys. ReV. Lett.* **33** (1974) 1406.

[14] V. Luth et al., SPEAR Coll., *Phys. ReV. Lett.* **34** (1975) 1125.
G. Goldhaber et al., SPEAR Coll., *Phys. ReV. Lett.* **37** (1976) 255.
I. Peruzzi et al., SPEAR Coll., *Phys. ReV. Lett.* **37** (1976) 569.

[15] K. Niu, E. Mikumo and Y. Maeda, *Prog. Theor. Phys.* **46** (1971) 1644.

[16] S. Bianco, F.L. Fabbri, D. Benson and I. Bigi hep-ex/0306039 to be published in *Il Nuovo Cimento*.

[17] The story of the charm discovery in Japan : K. Niu *Proceed. of the First Intern. Workshop on Nucl. Emulsion Techn.* June 12-14, 1998, Nagoya, Japan (DPNU-98-39).

[18] M.L. Perl et al., MARK I Coll.,*Phys. ReV. Lett.* **35** (1975) 1489.
M. L. Perl et al., MARK I Coll.,*Phys. Lett.* **63B** (1976) 466.

[19] S.W. Herb et al., Fermilab P.C. Coll., *Phys. ReV. Lett.* **39** (1977) 252.
W. R. Innes et al., Fermilab P.C. Coll., *Phys. ReV. Lett.* **39** (1977) 1240.

[20] C. Albjar et al., UA1 Coll.,*Phys. Lett.* **B186** (1987) 247.
H. Albrecht et al., ARGUS Coll.,*Phys. Lett.* **B192** (1987) 245.
A. Bean et al., CLEO Coll.,*Phys. ReV. Lett.* **58** (1987) 183.

[21] F. Abe *et al.*, CDF Collaboration, *Phys. ReV. Lett.* **74** (1995) 2626.
S. Abachi *et al.*, D0 Collaboration, *Phys. ReV. Lett.* **74** (1995) 2632.

[22] L.L. Chau and W.-Y. Keung, *Phys. ReV. Lett.* **53** (1984) 1802.

[23] K. Hagiwara et al., Particle Data Group, *Phys. ReV.* **66** (2002) 010001.

[24] L. Wolfenstein, *Phys. ReV. Lett.* **51** (1983) 1945.

[25] M. Schmidtler and K.R. Schubert, *Z. Phys.* **C 53** (1992) 347.

[26] T.Inami and C.S.Lim, *Prog.Theor.Phys.* **65** (1981) 297; *ibid.* **65** (1981) 1772.

[27] A.J. Buras, M. Jasmin and P.H. Weisz, *Nucl. Phys.* **B347** (1990) 491.

[28] S. Herrlich and U. Nierste, *Nucl. Phys.* **B419** (1994) 192;
G. Buchalla, A.J.Buras and M.E. Lautenbacher, *ReV. Mod. Phys.* **68**, (1996) 1125.

[29] "THE CKM MATRIX AND THE UNITARITY TRIANGLE". *Based on the First Workshop on "CKM Unitarity Triangle" held at CERN, 13-16 February 2002.* Edited by: M. Battaglia, A. Buras, P. Gambino and A. Stocchi. To be published as **CERN Yellow Book** (hep-ph/0304132).

[30] Proceedings of the Second Workshop on "CKM Unitarity Triangle" *held in Durham, 5-9 April 2003.* Edited by: P. Ball, P. Kluit, J. Flynn and A. Stocchi.

[31] http://www.slac.stanford.edu/xorg/hfag/

[32] http://www.hep.phys.soton.ac.uk/lwg-bmix/

[33] M. Ciuchini, E. Franco, V. Lubicz, F. Parodi, L. Silvestrini and A. Stocchi, "Unitarity Triangle in the Standard Model and sensitivity to new physics" (hep-ph/0307195)

[34] M. Battaglia, M. Calvi, P. Gambino, A. Oyanguren, P. Roudeau, L. Salmi, J. Salt, A. Stocchi, N. Uraltsev "Heavy Quark Parameters and $|V_{cb}|$ from Spectral Moments in Semileptonic B Decays" **Phys. Lett. B556** (2003) 41-49 (hep-ph/0210319).

[35] C. W. Bauer, Z. Ligeti, M. Luke, A. V. Manohar *Phys. ReV.* **D67** (2003) 054012. (hep-ph/0210027).

[36] A. J. Buras, F. Parodi, A. Stocchi, "The CKM Matrix and The Unitarity Triangle: another Look" **JHEP 0301** (2003) 029 (hep-ph/0207101).

[37] M. Ciuchini, G. D'Agostini, E. Franco, V. Lubicz, G. Martinelli, F. Parodi, P. Roudeau, A. Stocchi, "2000 CKM-Triangle Analysis A Critical Review with Updated Experimental Inputs and Theoretical Parameters", **JHEP 0107** (2001) 013. (hep-ph/0012308).

[38] C. Dib, I. Dunietz, F. Gilman, Y. Nir, *Phys. ReV.* **D41** (1990) 1522.
G. Buchalla, A.J. Buras and M.E. Lautenbacher, *ReV. Mod. Phys.* **68** (1996) 1125.
M. Lusignoli, L. Maiani, G. Martinelli and L. Reina, *Nucl. Phys.* **B369** (1992) 139.
A. Ali and D. London, in Proceeding of "ECFA Workshop on the Physics of a B Meson Factory", Ed. R. Aleksan, A. Ali (1993);
A. Ali and D. London, hep-ph/9405283;
A. Ali and D. London, in Proceeding of "27th International Conference on High Energy Physics (ICHEP95)", Glasgow, Scotland, 20-27 July 1994, hep-ph/9409399;
A. Ali and D. London, *Z. Phys.* **C65** (1995) 431. S. Herrlich and U. Nierste, *Phys. ReV.* **D52** (1995) 6505.
M. Ciuchini, E. Franco, G. Martinelli, L. Reina and L. Silvestrini, *Z. Phys.* **C68** (1995) 239.
A. Ali and D. London, *Nuovo. Cim.* **109A** (1996) 957;

A. Ali, *Acta Physica Polonica* **B 27** (1996) 3529;
A. Ali and D. London, *Nucl. Phys.* **54A** (1997) 297.
A.J. Buras, Invited talk at "7th International Symposium on Heavy Flavour Physics", Santa Barbara, CA, 7-11 July 1997, hep-ph/9711217;
A.J. Buras and R. Fleischer, *Adv. Ser. Direct. High. Energy Phys.* **15** (1998) 65; hep-ph/9704376.
R. Barbieri, L.J. Hall, S. Raby and A. Romanino, *Nucl. Phys.* **B493** (1997) 3.
A. Ali and B. Kayser, invited article in 'The Particle Century", Inst. of Physics Publ., Inc., Bristol and Philadelphia, 1998, Ed. Gordon Fraser, hep-ph/9806230.
P. Paganini, F. Parodi, P. Roudeau and A. Stocchi, *Phys. Scripta* **V. 58** (1998) 556.
F. Parodi, P. Roudeau and A. Stocchi, *Nuovo Cim.* **112A** (1999) 833.
F. Caravaglios, F. Parodi, P. Roudeau, A. Stocchi, hep-ph/0002171, talk at "B_{CP} 99", Taipei, Taiwan Dec. 3-7 1999, to appear in the Proceedings.
S. Mele, *Phys. ReV.* **D59** (1999) 113011.
A. Ali and D. London, *Eur. Phys J.* **C9** (1999) 687.
S. Plaszczynski, M.H Schune, Talk given at Heavy Flavours 8, Southampton, UK, 1999 (hep-ph/9911280).
M. Ciuchini, E. Franco, L. Giusti, V. Lubicz and G. Martinelli,*Nucl. Phys.* **B573** (2000) 201.
M. Bargiotti *et al.*, *La Rivista del Nuovo Cimento* **Vol. 23N3** (2000) 1.

[39] L. Lellouch, Plenary Talk given at the 31st ICHEP, Amsterdam, The Nederlands, July 24th-31st,2002.

[40] A. S. Kronfeld, P. B. Mackenzie, J. N. Simone, S. Hashimoto, S. M. Ryan, proceedings of FPCP, May 16–18, Philadelphia, Pennsylvania; *Phys. ReV.* **D66** (2002) 014503. (hep-ph/0110253).

[41] A. Stocchi "Experimental Results on Heavy Quarks" Plenary talk at ICHEP02 (26-31/7/2002 Amsterdam) *Nucl. Phys. Proc. Suppl.* **117** (2003) 145-163 (hep-ph/0211245).
M. Battaglia, talk given at ICHEP02 (26-31/7/2002 Amsterdam, The Nederlands)

[42] H.G. Moser and A. Roussarie, *Nucl. Instum. Meth.* **A384** (1997) 491.

[43] K. G. Wilson *Phys. ReV.* **D10** (1974), 2445.

[44] JLQCD Coll., S. Hashimoto et al., hep-lat/0209091.

[45] A.S. Kronfeld and S.M. Ryan, *Phys. Lett* **B543** (2002) 59, hep-ph/0206058.

[46] http://www.utfit.org

NEUTRINO MASSES

Stefan Pokorski
Institute for Theoretical Physics, University of Warsaw
and Institute of Physics, University of Bonn
Stefan.pokorski@fuw.edu.pl

Abstract After a review of masses in the Standard Model ν, a gauge invariant effective was term is introduced. The Dirac and Majorana contributions the see-saw mechanism, are discussed and the observable effects described.
A discussion of the quantum corrections is presented, with an implication on neutrinos masses.
A brief summary of neutrino masses in the context of GUT is then given.

Keywords: Electroweak symmetry, Dirac masses, Majorana masses, Effective interactions, Supersymmetry, See-saw mechanism

1. Introduction

It is instructive to discuss the theory of neutrino masses from the perspective of the Standard Model [1]. So let us have a look at the basic structure of the electroweak theory. Its underlying principles are:

1 local $SU(2)_L \times U(1)_Y$ gauge symmetry and electroweak unification

2 spontaneous breaking of $SU(2)_L \times U(1)_Y$ symmetry to $U(1)_{EM}$ gauge symmetry

3 matter content (chiral fermions, Higgs doublet (s), absence of the right-handed neutrino)

4 renormalizability

The principles (1) - (4) imply global U(1) symmetries of the theory: baryon and lepton number conservation $\Delta B = \Delta L = 0$. In fact, for leptons the implication is even stronger, namely $U_e(1) \times U_\mu(1) \times U_\tau(1)$ is

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 83–109.

a global symmetry of the electroweak lagrangian and the lepton flavour numbers are separately conserved: $\Delta L_e = \Delta L_\mu = \Delta L_\tau$=0. For quarks, quark mixing explicitly breaks quark flavour $U(1)'s$ and only the total baryon number is conserved. However, the structure of the theory assures that flavour changing neutral currents(FCNC) are absent at the tree level. They are generated at the one-loop level but suppressed with respect to the generic one-loop prediction $\mathcal{O}$ $(\alpha_{EM} G_F)$ due to the GIM mechanism (generalized to three families of quarks).
Renormalizability of the theory assures that quantum corrections can be calculated in terms of a small number of free parameters and they are in perfect agreement with the LEP precisions data. This is a very powerful test of the theory, and in particular of its basic structure (1)-(4), with precision $\mathcal{O}$ (one per mille). Equivalently, the theory is tested up to non-renormalizable corrections $\mathcal{O}(E^2/\Lambda^2)$ with the scale $\Lambda \sim \mathcal{O}(1\ \ TeV)$.
Renormalizability requires that only Lorentz and gauge invariant operators of at most dimension four can be present in the electroweak lagrangian. Only the theories with at most dimension four operators in the lagrangian can be renormalizable. This theorem is heuristically obvious but we shall not discuss it in more detail here. The important examples of dimension less than or equal to four operators are field kinetic terms, mass terms for bosons and fermions, the interaction terms for four scalar fields and the interaction term for a scalar with a fermion (Yukawa coupling). For instance, the Fermi interaction (four-fermion interaction) is an example of a (non-renormalizable) dimension six operator and cannot be present in a renormalizable theory.

2. Chirality and fermion masses in the electroweak theory

The particles of the Standard Model (SM) include the Higgs doublet $H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}$ with the hypercharge $Y = +1/2 \quad (Y = Q - T^3)$ and the chiral fermions. Massless chiral fermions are the fundamental objects of matter: left-handed, with helicity $\lambda = -1/2$, and right-handed, with helicity $\lambda = 1/2$. It is so because parity and charge conjugation are not the symmetries of our world. The left-handed fermions carry different weak charges from the right-handed fermions. Chiral fermion fields are two-component (Weyl) spinors (see the Appendix for some details on the two-component spinor notation):

$\underline{SU(2) \text{ doublets}}$

$$q_1 \equiv \begin{pmatrix} u \\ d \end{pmatrix} \qquad q_2 \equiv \begin{pmatrix} c \\ s \end{pmatrix} \qquad q_3 \equiv \begin{pmatrix} t \\ b \end{pmatrix} \tag{1}$$

$$l_1 \equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix} \qquad l_2 \equiv \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \qquad l_3 \equiv \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{2}$$

with the electric charge and the hypercharge

	u	d	ν	e
Q	2/3	-1/3	0	-1
Y	1/6	1/6	-1/2	-1/2

These are left-handed chiral fields in the representation $(0, 1/2)$ of the $SL(2, C)$ (see the Appendix), each describing two massless degrees of freedom: a particle with the helicity $\lambda = -1/2$ and its antiparticle with $\lambda = +1/2$. (The chiral fields can also be written as four-component spinors (see e.g. [1]) but in the following we shall be using the Weyl notation).

Right-handed fields [$(1/2, 0)$ of $SL(2, C)$] in the same representations of $SU(2) \times U(1)_Y$ as the left-handed fields (1) and (2) do not exist in Nature. Instead, we have

$\underline{SU(2) \text{ singlets}}$

$$\begin{array}{lll} u_R, & c_R, & t_R \\ d_R, & s_R, & b_R \\ e_R, & \mu_R, & \tau_R \end{array}$$

in $(1, +2/3)$, $(1, -1/3)$ and $(1, -1)$ of $SU(2) \times U(1)_Y$, respectively. These are right-handed chiral fields in the $(1/2, 0)$ representation of the group $SL(2, C)$. For constructing a Lorentz invariant Lagrangian, it is more convenient to take as fundamental fields only the left-chanded chiral fields. Thus, we introduce left-handed chiral fields, e.g

$$u^c, \quad c^c, \quad t^c \tag{3}$$

in $(1, -2/3)$ of $SU(2) \times U(1)_Y$, such that

$$\bar{u^c} \equiv CPu^c(CP)^{-1} = u_R \tag{4}$$

Indeed, CP transformation results in the simultaneous change of chirality and charges (representation $R \to R^*$ for internal symmetries). Moreover, we see that the electric charge $Q = T_3 + Y$ satisfies, e.g.

$$Q_{e^c} = +1 = -Q_e \tag{5}$$

and the two left-handed fields e and e^c become charge conjugate to each other when $U_{EM}(1)$ remains the only unbroken symmetry:

$$Ce^cC^{-1} = e \tag{6}$$

Only at this point we can define Dirac fermions

$$\Psi = \begin{pmatrix} e \\ \bar{e^c} \end{pmatrix} \quad \equiv \quad \begin{pmatrix} e_L \\ e_R \end{pmatrix} \tag{7}$$

We note that the matter chiral fields of the SM do not include a right-handed neutrino field ν_R in $(1,1)$ of $SU(2) \times U(1)_Y$ (such a charge assignement preserves the relation $Q = T_3 + Y$) or equivalently, a left-handed field ν^c such that

$$\nu_R = CP\nu^c(CP)^{-1} \tag{8}$$

However, since the neutrino electric charge $Q_\nu = 0$, after spontaneous breaking of $SU(2) \times U(1)_Y$ to $U(1)_{EM}$, we can construct a four-component spinor, namely a Majorana spinor Ψ_M, from the left-handed neutrino field ν alone:

$$\Psi_M = \begin{pmatrix} \nu \\ \bar{\nu} \end{pmatrix} \tag{9}$$

where $\bar{\nu} = CP\nu(CP)^{-1}$.

Once the only unbroken symmetry is $U(1)_{EM}$ the following mass terms are Lorentz and $U(1)_{EM}$ invariant:

Dirac masses

$$m(\lambda\lambda^c + \bar{\lambda}\bar{\lambda^c}) \equiv m_D\bar{\Psi}\Psi \tag{10}$$

where $\lambda = u, d, ..$ and $\lambda^c = u^c, d^c, ...$ They can be represented by the diagram,shown in Fig.1, which has a very simple interpretation (see the Appendix): in the $\lambda\lambda^c$ coupling, the field λ annihilates e.g. an incoming electron with helicity $\lambda = -1/2$ and momentum k and the field λ^c creates an outgoing electron with $\lambda = +1/2$ and momentum k. Thus, mass term means helicity flip. Note that the $\lambda\lambda^c$ coupling is not $SU(2) \times U(1)_Y$

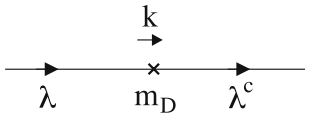

Figure 1.

invariant, so mass terms can appear only in the low energy effective theory. Majorana mass

$$m(\nu\nu + \bar{\nu}\bar{\nu}) \equiv m_M \bar{\Psi}_M \Psi_M \quad , \tag{11}$$

is possible because ν is an $U(1)_{EM}$ singlet; the corresponding diagram is shown in Fig. 2

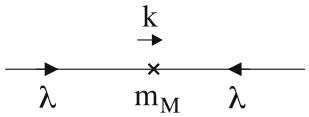

Figure 2.

and, again, the $\lambda\lambda$ coupling has the following interpretation: the first ν annihilates an incoming neutrino with $\lambda = -1/2$ and momentum k and the second ν creates an outgoing anti-neutrino (=neutrino) with $\lambda = +1/2$ and momentum k. Majorana neutrino is a particle with two helicity states and is invariant under charge conjugation.

Such mass terms, if present in the effective low energy theory, must originate in the SM, before the breaking of the gauge symmetry, from the Lorentz and $SU(2) \times U(1)_Y$ scalars. For the charged fermions, we can write down the following Yukawa couplings to the Higgs doublet:

$$\mathcal{L}_{Yukawa} = -Y_l^{BA} H_i^* l_{iA} e_B^c - Y_d^{BA} H_i^* q_{iA} d_B^c - Y_u^{BA} \epsilon_{ij} H_i q_{iA} u_B^c + hc \tag{12}$$

where i is the $SU(2)$ index and A, B are generation indices. We use the fact that the two-dim representation of $SU(2)$ is real and $i\tau_2 H$ transforms as H^*, i.e. as $2^*(\equiv 2)$ of $SU(2)$. Therefore, $(i\tau_2 Hq) = \epsilon_{ij} H_i q_j$ is also an invariant of $SU(2)$. After spontaneous breaking of $SU(2)\times U(1)_Y$ to $U(1)_{EM}$ by the Higgs boson vacuum expectation value v we obtain the earlier introduced Dirac masses

$$\mathcal{L}_{mass} = -v(Y_l^{BA} e_A e_B^c + Y_d^{BA} d_A d_B^c + Y_u^{BA} u_A u_B^c) + hc \qquad (13)$$

However, at the level of the full, $SU(2) \times U(1)_Y$ invariant theory, there is no renormalizable term that would give neutrino mass. It is so because ν^c is absent from the spectrum of the SM. Thus, in the SM, neutrinos are massless.

The interactions (12,13) are written in some "electroweak" basis defined by eigenvectors of the $SU(2) \times U(1)$ symmetry group. In such a basis, both the fermion masses and the Yukawa couplings are in general non-diagonal in the flavour indices (A, B). However, we can introduce another set of fields (say, primed fields) describing physical particles (mass eigenstates). The flavour of the primed fields is defined in the mass eigenstate basis. The two sets of fields are related to each other by unitary transformations:

$$\begin{gathered} u = U_L u' \qquad d = D_L d' \\ u^c = u'^c U_R^\dagger \quad d^c = d'^C D_R^\dagger \\ e = E_L e' \\ e^c = e'^c E_R^\dagger \end{gathered} \qquad (14)$$

which, of course, do not commute with the $SU(2) \times U(1)$ gauge transformations and can be performed only after the spontaneous breakdown of the gauge symmetry. In eq.(14), the fields u, d , e denote three-dimensional vectors in the flavour space.

The transformations (14) diagonalize the mass terms and the Yukawa couplings defined by (12). After diagonalization we can combine the chiral fields into Dirac fields, according to eq.(7), which are the mass eigenstates. The weak currents can be expressed in terms of the physical (mass eigenstates) fields:

$$\begin{aligned} J_\mu^- &= \sum_{A,B} \bar{u}'_A \bar{\sigma}_\mu (V_{CKM})_{AB} d'_B \\ &+ \sum_A \bar{\nu}'_A \bar{\sigma}_\mu e'_A \end{aligned} \qquad (15)$$

where the Cabibbo-Kobayashi-Maskawa matrix $V_{CKM} = U_L^\dagger D_L$. Note that the lepton current is diagonal in flavour (defined in the charged lepton mass eigenstate basis) because the massless neutrino field can be

redefined by the transformation $\nu_A = E^L_{AB}\nu'_B$ where E^L_{AB} is the transformation diagonalizing the charged lepton mass matrix (see(14)).

Thus, for the lepton current, $V_{CKM} = E^{L\dagger}E^L = 1$.

It is important to remember that in the SM (with only one Higgs doublet) the Yukawa couplings to the physical Higgs boson (and, in fact, also the couplings to the Z^0) and the mass terms are diagonalized by the same unitary rotations. So they are flavour diagonal. The only source of flavour non-conservation resides in V_{CKM}. In particular, not only the global lepton number but also each flavour lepton number is separately conserved.

The experimentally measured V_{CKM} matrix looks "natural": the bigger the differences between quark mass eigenvalues the smaller the mixing between them. The generic structure of 2×2 mass matrix with such properties is

$$\begin{pmatrix} 0 & \varepsilon \\ \varepsilon & 1 \end{pmatrix} \text{with} \;\; \lambda_{1,2} = \varepsilon^2, 1 \;\; \text{and} \;\; \theta \sim \varepsilon$$

3. Neutrino masses in the effective *SM*

There is at present strong experimental evidence for neutrino oscillations whose most obvious and most natural explanation is that neutrinos are massive and the mass eigenstates are different from the weak interaction eigenstates. This is the first experimental evidence for physics beyond the Standard Model and we must consider going beyond the principles sketched in the previous chapter.

A very natural attitude is to regard the Standard Model as the low energy effective theory which is an approximation to a deeper, still unknown and hopefully renormalizable theory. Then, non-renormalizable corrections to the Standard Model are acceptable as the remnants and traces of a deeper theory. They can be the most general higher dimension operators invariant under the Standard Model gauge symmetry and consistent with the Higgs doublet pattern, suppressed by the appropriate power of the scale Λ of the new (unknown) physics. Such corrections can describe various small effects absent in the (renormalizable) Standard Model but eventually established experimentally. This is then a window to discover before formulating in detail the deeper theory. Active neutrino Majorana masses are one example of such corrections.

In an extention of the SM that admits non-renormalizable corrections, gauge and Lorentz invariant neutrino mass terms may appear as dimension five operator

$$\frac{1}{M}(Hl_A)\lambda_{AB}(Hl_B) \tag{16}$$

where now cut-off scale to the SM is denoted by M (later on, in the see-saw mechanism, it will be identified with the right-handed neutrino Majorana mass) and we use the following notation: $(Hl) \equiv \epsilon_{ij} H^i l^j$ denotes $SU(2)$ contraction and $ll \equiv \epsilon^{\alpha\beta} l_\alpha l_\beta$ denote Lorentz contraction. After spontaneous SM gauge symmetry breaking by the Higgs boson vev the operator (11) gives indeed a Majorana mass matrix for the active neutrinos:

$$\nu_A m_{AB} \nu_B + hc, \qquad m_{AB} = \frac{v^2}{M} \lambda_{AB} \tag{17}$$

The matrix m is complex and symmetric. It can be diagonalized by a unitary matrix U:

$$\nu_B = U^{Ba} \nu'_a \tag{18}$$

where ν'_a are neutrino mass eigenstates. We get

$$\nu m \nu = \nu' U^T m U \nu' \tag{19}$$

where

$$U^T m U = diag(m_1, m_2, m_3) \tag{20}$$

The weak current can be written in the mass eigenstate basis (of charged leptons and neutrinos):

$$\bar{e}_A \bar{\sigma}_\mu \nu_A = \overline{E_L^{Ab}} e'_b \bar{\sigma}_\mu U^{Ac} \nu'_c \tag{21}$$

and the matrix

$$V_{MNS} = E_L^\dagger U \tag{22}$$

is the so-called Maki-Nakagava-Sakata neutrino mixing matrix. In the charged lepton mass eigenstate basis we get then

$$V_{MNS} = U \tag{23}$$

and the most general form of the 3×3 Majorana neutrino mixing matrix can be written as

$$U = V \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_2} & 0 \\ 0 & 0 & e^{i\alpha_3} \end{pmatrix} \tag{24}$$

where V can be parametrized in the same way as the V_{CKM} matrix and we use the convention that the neutrino Majorana mass eigenvalues are real. [One should keep in mind that the physical masses are defined by the poles of the propagators which are at $m^a_{ph} = |m^a|$, no matter if the mass matrix parameters are complex or real.] In our convention

the phases α_1, α_2 in eq. (24) are physical. They originate from the fact that, for the Majorana mass term $m\nu\nu$, there is no freedom of rotation of the neutrino fields ν by a phase because there is no right-handed fields rotation to compensate it. Indeed, for (we omit the primes)

$$(\nu_1\ \nu_2\ \nu_3) \begin{pmatrix} m_1 & & \\ & m_2 & \\ & & m_3 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{25}$$

we have

$$\nu_a \to e^{i\alpha_a}\nu \Rightarrow m_a \to m_a e^{i2\alpha_a} \tag{26}$$

in contradiction to our convention. We note that the Majorana mixing matrix is real (CP conserved) when the V phase $\delta = 0 \mod \pi$ and $\alpha_{1,2} = 0 \mod \pi/2$. However, for $\alpha_i = \pi/2$ we can absorb the phase into real mass eigenvalues with $m_a \to -m_a$. Thus, the sign of the Majorana mass has physical sense: different signs mean different CP parities. We recall that the physical mass defined by the pole in the propagator is always positive $m^a_{ph} = |m_a|$. The sign of the neutrino mass plays the role in its interactions.

It is worth putting that view at the Standard Model (or MSSM) as an effective low energy theory into the better known perspective. We know now that Quantum Electrodynamics (QED) is a renormalizable theory and at the same time it is the low energy approximation to the electroweak theory. Its renormalizability means calculability with arbitrary precision. But it is only an effective theory so we know that its predictions disagree with experiment at the level $\sim \mathcal{O}(E/M_{\rm W})$, where the energy E is the characteristic energy for a given process. For example, let us have a look at the lepton magnetic moment. It gets contributions from the diagrams depicted in Fig.3.

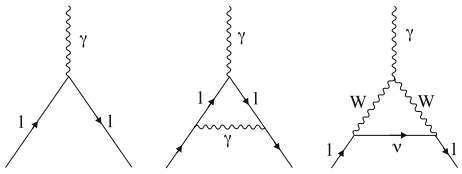

Figure 3.

Thus, for the effective interaction with the magnetic field we get

$$\frac{e}{2m_l}\bar{u}\vec{\sigma}\cdot\vec{H}u\ (1 + \frac{\alpha}{2\pi} + \mathcal{O}\left(\alpha\frac{m_l^2}{M_{\rm W}^2}\right) + \ldots) \tag{27}$$

where the role of the energy scale is played by the lepton mass m_l. The "weak" correction is calculable in the full electroweak theory, but at

the level of QED as an effective theory it has to be added as a new non-renormalizable (but $U(1)_{EM}$ invariant) interaction

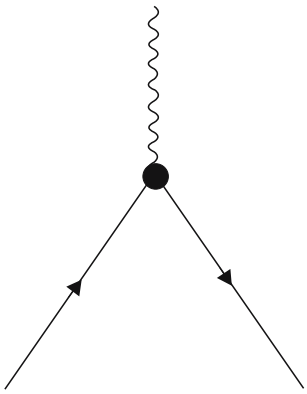

Figure 4.

$$\mathcal{L}^{eff} = \frac{m_l}{M_{\rm W}^2}\bar{\psi}\sigma_{\mu\nu}\psi F^{\mu\nu} \qquad \text{(dim 5)} \tag{28}$$

This would have been a way to discover weak interactions (and to measure the weak scale) in purely electromagnetic processes: we extend QED to a non-renormalizable theory by adding higher dimension operators and look for their experimental manifestation in purely electromagnetic processes once the experimental precision is high enough. Luckily enough for us, effective QED may also contain other than (28) non-renormalizable corrections, $U(1)_{EM}$ invariant but violating the conservation of quantum numbers that are accidentally conserved in QED, for instance flavour. Such corrections manifest themselves as different type of interactions - weak interactions - and were easy to discover experimentally. Next, insisting on renormalizability led us to the electroweak theory. The situation with the Standard Model can be expected to be quite analogous and we now hope to make a similar next step.

The main virtue of a renormalizable theory is that it can describe with good approximation the physics in some energy range, without any knowledge of the deeper theory. However, at the same time, once a renormalizable field theory has been formulated to describe certain phenomena, it is difficult to make the next step. One either needs the precision $\mathcal{O}(E/\Lambda)$, to see the need for non-renormalizable corrections, or energies $E \sim \Lambda$, to directly discover new particles with the masses $M \sim \Lambda$. The neutrino sector provides us the first experimental evidence for non-renormalizable corrections to the Standard Model viewed as an effective theory.

4. More on the origin of neutrino masses

The description of the neutrino masses in the low energy effective theory is provided by the operator (11). It is interesting, however, to discuss how such an operator can potentially originate from renormalizable interactions of a deeper theory, with the characteristic scale M of its heavy particles, and can we explain that way the smallness of the neutrino masses. Imagine, we supplement the SM with another left-handed particle ν^c, a singlet of $SU(2) \times U(1)_Y$, i.e. a field such that

$$CP\nu^c(CP)^{-1} \equiv \nu_R \tag{29}$$

can be interpreted as a right-handed neutrino field. A right-handed Majorana mass term can then be added to the SM

$$(\nu^c)^A M^{AB}_{MAJ}(\nu^c)^B + hc \tag{30}$$

Moreover, we can construct Yukawa interactions

$$\epsilon_{ij} H_i (\nu^c)^B Y_\nu^{BA} l_j^A + hc \tag{31}$$

with a new set of (neutrino) Yukawa couplings Y_ν^{BA}. Both terms are $SU(2) \times U(1)_Y$ invariant and renormalizable. We can consider then the diagram shown in Fig. 5.

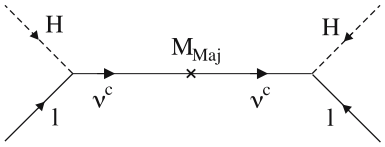

Figure 5. Diagram generating the dimension 5 operator.

At the electroweak scale v, if $M_{MAJ} \gg v$, we obtain the effective interaction shown in Fig. 6

described by the operator

$$\epsilon_{ij} H_i l_j^A Y_\nu^{DA} (M^{-1}_{MAJ})^{DC} Y^{CB} \epsilon_{ij} H_i l_j^B \tag{32}$$

We recognize the previously introduced operator (11) with

$$\frac{\lambda_{AB}}{M} = (Y_\nu^T M^{-1}_{MAJ} Y_\nu)_{AB} \tag{33}$$

This is the so-called see-saw mechanism for neutrino masses.

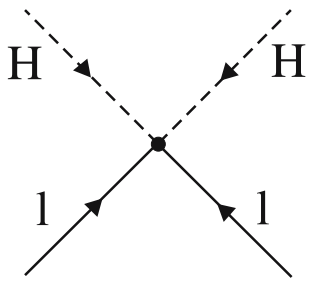

Figure 6. Effective dimension 5 operator.

With the Yukawa couplings (31) present in the Lagrangian, we can also contemplate the possibility that there is no right-handed Majorana neutrino mass term. The masses of the three active neutrinos would then be simply Dirac neutrino masses. In general both terms should be included in the theory defined by the $SU_L(2) \times U_Y(1)$ gauge symmetry and renormalizability. The Majorana mass term (III.23) can be, however, eliminated by imposing the additional global U(1) symmetry ensuring conservation of the lepton number L. The fields ν^c must then have $L = -1$, opposite to L of the leptonic doublets l_A. Neutrinos are then Dirac particles like the other fermions. There are two reasons why Dirac masses are not so attractive. One is the need for very small numerical values of the Yukawa couplings Y_ν. The other is that the lepton number conservation has to be imposed as an additional global symmetry (remember that in the SM it is a consequence of the field content and of the renormalizability and not an additional assumption).

The presence of the Majorana mass term (III.23), which breaks the global U(1) symmetry, inevitably makes neutrinos Majorana particles (in fact what allows to interpret in the SM the two helicity states described by the ν field as a particle and an antiparticle is just the lepton number!)

This discussion gives us the opportunity to distinguish between the Dirac and Majorana neutrinos. It is interesting to ask if those two possibilities can be distinguish experimentally. Unfortunately , the question: are ν and $\bar{\nu}$ two helicity states of a Majorana neutrino or differ also in some other way, by some quantum numbers (L) which would justify to call them particle and antiparticle, cannot be easily settled experimentally. The conservation of the lepton number and the Dirac nature of neutrinos imply that neutrinoless double-β decay is forbidden but the effects predicted with Majorana neutrinos are generically very small. One may wonder if there are other physical effects that can distinguish theories of neutrino masses with L conserved (Dirac neutrinos) from those with $\Delta L \neq 0$ (Majorana neutrinos). Unfortunately, one sees that

the effects that can be explained by assigning the lepton number and conserving it, can also be obtained from helicity structure of the weak lagrangian (one can see that by, for instance, considering the processes $\pi^+ \to \mu^+\nu_\mu$ and $\pi^- \to \mu^-\bar{\nu}_\mu$). The above is true up to corrections proportional to neutrino masses and with the masses $< \mathcal{O}(1eV)$ no observable effects can be expected.

In our discussion of the renormalizable extensions of the Standard Model based on both terms eq. (III.23) and (31) we could have assumed that the Dirac mass (31) and Majorana mass (III.23) are of the same order of magnitude . This is again unattractive as a natural value for the Majorana mass M is at least of the order of the cut-off to the Standard Model (its value is not protected by any known low energy symmetry). Finally, there is also a possibility to couple two lepton $SU_L(2)$ doublets to a $SU_L(2)$ Higgs triplet (singlets would violate the electric charge conservation). This possibility is not particularly attractive for several reasons. The smallness of the neutrino masses would require either $v_{\text{triplet}} \ll v_{\text{doublet}}$ or the triplet couplings orders of magnitude smaller than the other Yukawa couplings. Moreover, the introduction of the triplet would make the parameter ρ a free parameter of the theory (hence not calculable) with its own counterterm allowing to adjust its value at will. Note that the mass term obtained from the coupling of two lepton $SU(2)_L$ doublets with a Higgs triplet would be a Majorana mass term, too. It would violate the lepton number by $\Delta L = 2$.

Our summary on neutrino masses is as follows: Majorana masses for the three active neutrinos can be generated by extended Higgs sector (Higgs triplets). Another possibility is to supplement the SM with right-handed neutrinos ν_R. If only their Yukawa interactions are present, neutrinos are Dirac particles. If ν_R s have also Majorana masses, all neutrinos are Majorana particles. With small $M \sim M_W$, in addition to Majorana $(\nu + \bar{\nu})$ there are also light sterile Majorana $(\nu^c + \bar{\nu}^c)$. None of those possibilities looks particularly attractive, for the reasons discussed above.

If the scale M is high enough, the effects of the Majorana masses for the three active neutrinos can be discussed in the SM or MSSM supplemented by the operator (11). We should also stress that this operator (11) is the only one of dimension 5 contributing to the neutrino Majorana mass matrix. Other possible contributions are of higher dimension. Thus, one may expect that in the effective SM or MSSM the neutrino masses are indeed described by that operator, even if its origin is different form the see-saw mechanism.

We end this section by stressing that the see-saw mechanism with some large scale M in eq.(11) looks like the most plausible explanation of small neutrino masses. Indeed

i) the smallness of m_ν is then related to its zero electric charge

ii) the smallness of m_ν is also related to lepton number violation at the scale M

iii) with $m_\nu \sim Y^2 \frac{v^2}{M}$, v=240 GeV and for $Y \sim 0(1)$ we get $m_\nu \sim (0.01 - 0.1)eV$ for $M \sim (10^{15} - 10^{13})GeV$ So, the scale M fits nicely with the GUT scales and is natural (a cut-off to the SM or MSSM)

iv) ν^c completes the spinor representation of S0(10)

v) heavy ν^c can play important role in bariogenesis via leptogenesis.

In the rest of these lectures we take therefore the point of view that neutrino masses are a manifestation of <u>new</u> physics at <u>high</u> energy scale and can be effectively described by the operator (11). We do not know yet the theory that gives (11) but we can read off the structure of (11) from experimental data and try to look for theories that can explain it.

5. Phenomenology of neutrino masses

We shall discuss now in more detail phenomenology of neutrino masses in the effective theory defined by the SM or MSSM with three light neutrinos and supplemented by the dimension five operator (11). They are described by 9 real parameters: $m_1,\ m_2,\ m_3,\ \theta_{12},\ \theta_{23},\ \theta_{13}$ and the phases $\delta,\ \alpha_1,\ \alpha_2$ (see eq.(III.16,24)). The angles θ_{ij} parametrize the V matrix present in eq.(24):

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \quad (34)$$

Let us order the neutrino masses so, that m_3 is the most splitted mass and $m_2 > m_1$. Also, we define

$$\Delta m^2_{ij} = m^2_j - m^2_i \quad (35)$$

The experimental data then tell us that $\Delta m^2_{23},\ \theta_{23}$ are "atmospheric" parameters, and $\Delta m^2_{12},\ \theta_{12}$ are "solar" parameters. We have reasonably good knowledge of $\Delta m^2_{12},\ |\Delta m^2_{23}|,\ \theta_{23},\ \theta_{12}$ and a limit on θ_{13}. These parameters are accesible in neutrino oscillation experiments.

Unless otherwise explicitly stated, we shall always work in the weak eigenstate basis which is the charged lepton mass eigenstate basis i.e. after the unitary transformation E_L (see(14)) performed on the lepton doublets l_A defined in (2).

According to the present experimental information (and assuming CP conservation) the most plausible approximate form of the mixing matrix is:

$$U = \begin{pmatrix} c_{12} & s_{12} & 0 \\ -\frac{s_{12}}{\sqrt{2}} & \frac{c_{12}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{s_{12}}{\sqrt{2}} & -\frac{c_{12}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{36}$$

The maximal atmospheric mixing and small (1,3) mixing with $\theta_{13} < 0.2$ follow from the Super-Kamiokonde and Chooz experiments, respectively. The most recent SNO data strongly favour also a large solar mixing angle (LA solution). Moreover, all oscillation patterns are consistant with two independent mass squared differences

$$\begin{array}{ll} |\Delta m^2_{23}| \sim 10^{-3} eV^2 & \text{(atmospheric)} \\ \Delta m^2_{12} \sim 10^{-5} eV^2 & \text{(solar)} \end{array} \tag{37}$$

Recent results on the large scale structure formation and anisotropies in the temperature of the cosmic background radiation obtained by the WMAP experiment give upper bound $\Sigma m_a \sim 0.7$ eV, i.e. for degenerate neutrinos we get $m_a \sim 0.2$ eV.

Since the mass eigenvalues are not measured, contrary to the quark and charged lepton case, several mass eigenvalue patterns are possible, consistently with the values (37). These are:

I hierarchical

$$|\Delta m^2_{23}| \sim m^2_3 \gg m^2_2 > m^2_1; \tag{38}$$

here we expect $|m_3| \sim \sqrt{|\Delta m^2_{23}|}$ and

$$|\frac{m_3}{m_{2,1}}| \sim \sqrt{\frac{|\Delta m^2_{23}|}{\Delta m^2_{12}}} \sim 10 \tag{39}$$

II inversly hierarchical

$$|\Delta m^2_{23}| \approx m^2_2, m^2_1 \gg m^2_3 \tag{40}$$

with $m^2_1, m^2_2 \gg \Delta m^2_{12}$

III degenerate

$$m_3^2 \approx m_2^2 \approx m_1^2 \gg |\Delta m_{23}^2| \qquad 0.05 \;\; eV \sim |m_a| \sim 0.2 \;\; eV \quad (41)$$

In summary,we have already got a good deal of experimental information enabling us to reconstruct from eq. (III.16) the low energy mass matrix $\hat{m}$, however up to several important ambiguities which are absent for the charged lepton masses. These are: the three different acceptable sets of absolute values of mass eigenvalues

$$|m_3|, |m_2|, |m_1| \sim \left\{ \begin{array}{ccc} 5 \cdot 10^{-2}, & 5 \cdot 10^{-3}, & 0 \\ 0, & 5 \cdot 10^{-2}, & 5 \cdot 10^{-2} \\ 0.1 & 0.1 & 0.1 \end{array} \right\} eV \qquad (42)$$

where the numbers mean orders of magnitude, and the relative signs of the masses are not fixed. Thus, using eq. (III.16) we get a number of acceptable possibilities for $\hat{m}$. Also, one should stress that the two large mixing angles make the neutrino mixing different from the quark mixing.

In principle, accessible in oscillation experiments also are: $sign(\Delta m_{23}^2)$ and the phase δ. For instance, the earth matter corrections to $\nu_\mu \to \nu_e$ transition depend on the sign of Δm_{23}^2. The CP violating phase δ can be measured by measuring $P(\nu_e \to \nu_\mu) - P(\bar{\nu}_e \to \bar{\nu}_\mu)$. Finally, the Majorana phases are in principle accessible by measuring the rates for the neutrinoless double β decays. The chances for observing experimentally all those, so far, unaccessible effects strongly depend on the value of θ_{13} and on the pattern of mass eigenvalues.

We can illustrate that latter point by discussing in more detail the feasibility of experimental observation of the $2\beta 0\nu$ decay, for various patterns of neutrino masses. The elementary diagram for such decays is shown in Fig. 7:

where the neutrino couplings are $\bar{e}\bar{\sigma}_\mu \nu_e$ (in the charged lepton eigenstate basis) and the neutrino mass insertion is $m_{ee}\nu_e\nu_e$. Hence, the rate for the decay is proportional to

$$|m_{ee}| = |\Sigma U_{ei}^2 m_i| \qquad (43)$$

and is sensitive to the Majorana phases:

$$|m_{ee}| = |\cos^2\theta_{13}(m_1 \cos^2\theta_{12} + m_2 e^{2i\alpha_2} \sin^2\theta_{12}) + m_3 e^{2i\alpha_3} \sin^2\theta_{13}| \quad (44)$$

The inspection of (44) helps to understand that, as it has been shown by detailed calculations, with the hierachical neutrino mass pattern (and

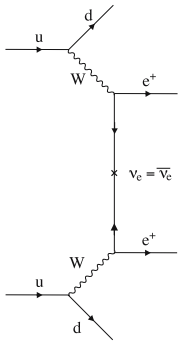

Figure 7. Neutrinoless double β decay.

LA solution) the $2\beta 0\nu$ decays can be accessible experimentally only for the angle θ_{13} close to the CHOOZ bound. Furthermore, for the inverted hierachy we get

$$(\Delta m^2_{32})^{1/2} \cos^2 2\theta_{12} < |m_{ee}| < (\Delta m^2_{32})^{1/2} \approx 5 \times 10^{-2} \quad eV \tag{45}$$

Finally, for the degenerate neutrinos with $|m| < 0.23$ eV and $\tan^2 \theta_{12} = 0.45 \pm 0.06$ we have

$$|m_{ee}| = |m| \cos^2 \theta_{12} (1 + e^{2i\alpha_2} \tan^2 \theta_{12}) < 0.15 eV \tag{46}$$

still below the present experimental limit $(0.3 - 0.5)$eV.

5.1 Quantum corrections to neutrino masses and mixing

We have discussed the low energy effective description of Majorana masses for the three active neutrinos. The present experimental neutrino data is consistent with such a description and the low energy Majorana mass matrix can be reconstructed from eq. (III.16), up to several important ambiguities, which are absent in the charged fermion mass sector. Those ambiguities were discussed at the end of the previous section. Effectively, they mean that, even with high precision data we shall always be left with a number of candidate low energy mass matrices consistent with the data. We have also argued that most likely the origin of the Majorana neutrino masses is some physics at high scale M (eq.(11)). A very important question is how the low energy "measured" (up to the ambiguties) mass matrices are related to the physics at the scale M. More precisely, we "measure" the Majorana mass matrix $\hat{m}(M_{\rm W})$ at the scale, say, $M_{\rm W}$ and the high energy physics will hopefully give us $\hat{m}(M)$. It is well known that in order to relate to each other the values

of the parameters of a theory at different energy scales one has to include quantum corrections. They are unambigous in the SM or MSSM, so from $\hat{m}(M_{\rm W})$, we can indeed infer the shape of $\hat{m}(M)$, provided the correct theory below the scale M is SM or MSSM. Thus, under the latter assumption we can obtain from the set of acceptable matrices $\hat{m}(M_{\rm W})$ the matrices $\hat{m}(M)$, to be understood in a deeper theory. Or vice versa, given a prediction fo $\hat{m}(M)$, to relate it to experimental observations we have to calculate $\hat{m}(M_{\rm W})$ by inclusion of quantum corrections. It has recently been understood that, contrary to the quark sector where such quantum corrections are generically small, the same may not be true for the quantum corrections (particularly in the MSSM) to the neutrino mass matrices. Thus, it is important to check their role in the theory of neutrino masses.

The standard technique to calculate the energy scale dependence of various lagrangian parameters are the Renormalization Group Equations (RGE). This technique can be used also to calculate the energy scale dependence of the neutrino mass matrix $\hat{m}(Q)$. One-loop RGE's sum up the so-called leading logarithmic contribution $\sum_n g^n ln^n(\frac{M}{Q})$ from quantum corrections. Such contributions may be significant when $M \gg Q$.

There are several ways to write down the RGE's for the neutrino mass matrix. The most transparent approach is to use the RGE's written directly for the mass eigenvalues and for the mixing matrix U, i.e. for the physical observables. With $t = \frac{1}{16\pi^2}\frac{M}{Q}$, in the SM and MSSM they read:
for the mass eigenvalues m_a

$$\frac{d}{dt}m_a(t) = -\sum_{i=e,\mu,\tau}(K + 2Y_i U_{ia}^2)m_a(t) \tag{47}$$

where K is some universal (neutrino flavour independent) factor and Y_i are charged lepton Yukawa couplings, and for the mixing matrix U

$$\frac{d}{dt}U_{ia}(t) = \sum_{b\neq a}\frac{m_b + m_a}{m_b - m_a}U_{ib}(U^T Y^2 U)_{ba} \tag{48}$$

where Y is the charged lepton Yukawa coupling matrix.

We see two important qualitative facts. One is that only the charged lepton Yukawa couplings can be relevant for changing the pattern of $\hat{m}(Q)$ by quantum corrections. In fact, the effects of Y_e and Y_μ are negligible and it is only Y_τ which is important. This is why the effects can be stronger for MSSM than for SM and in the MSSM, stronger for large than for low tg β.

The second point is even more crucial. We see that the evolution of the angles depends on the coefficient

$$A_{ba} \equiv \sum_{b \neq a} \frac{m_b + m_a}{m_b - m_a} \tag{49}$$

One can expect that when $|A_{ba}| \gg 1$ i.e. for the degenerate or inversly hierarchical structures, with the same CP parity $m_1 \approx +m_2$, the quantum corrections generated by Y_τ could become important and destabilize the initial pattern $\hat{m}(M)$ (for a review see ref.[2]). It has been shown that in those cases, independently of the initial structure of the $\hat{m}(M)$, the Y_τ generated quantum corrections drive the mixing angles to the so-called quasi-fixed point relation (one relation for the three mixing angles) at the scale $M_{\rm W}$:

$$\sin^2 2\theta_{12} = \frac{\sin^2\theta_{13} \quad \sin^2 2\theta_{23}}{(\sin^2\theta_{23} \quad \cos^2\theta_{13} \quad + \quad \sin^2\theta_{13})^2} \tag{50}$$

where θ_{12}, θ_{13} and θ_{23} are the solar, Chooz and atmospheric mixing angles, respectively. Using the present experimental information about θ_{23} and θ_{13}, the relation (50) predicts small solar mixing and is in contradiction with experiment, which favours the LA solution for the solar angle.

The significance of that result depends, of course, on the range of parameters for which the quasi-fixed point (50) is reached. This is shown in Fig. 8.

We note that the evolution of the angles depends on the paremeters $A_{ij}\epsilon$ where A_{ij} is given by eq.(49) and $\epsilon = \tan^2\beta \times 10^{-5}$, with the most relevant variable being A_{21}. Clearly, there are three regions in $A_{21}\epsilon$. For $A_{21}\epsilon \lesssim 1$ the RG evolution is negligible. For $A_{21}\epsilon \gtrsim 3$ the limit (50) is reached and, therefore, such models are excluded by the data. In the small range $1 \lesssim A_{21}\epsilon \lesssim 3$ there is a strong dependence of the renormalized angles on their boundary values and the low energy mixing pattern is very unstable under small changes of the initial values of the angles at the large scale M. So, the question is how constraining is the bound $A_{21}\epsilon \lesssim 1$ for neutrino mass models. Using experimental value $\Delta m^2_{12} \approx 5 \times 10^{-5}\text{eV}^2$ and for $\tan\beta = 10$ we get the condition $m_1 + m_2 < 0.25$ eV. WMAP gives us $|m| < 0.23$ eV (for degenerate neutrinos). We conclude that quantum corrections can be important for large values of $\tan\beta$ and then they make the neutrino mass models with degenerate neutrinos or with inversly hierarchical neutrinos with $m_1 \approx +m_2$ incompatible with the data.

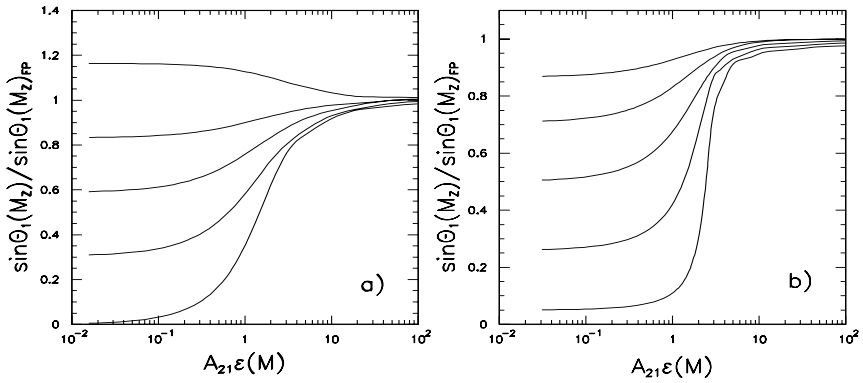

Figure 8. Approach to the IR fixed point for the angle θ_1 ($\theta_1 \equiv \theta_{12}$) as a function of $A_{21}\epsilon(M)$ for two different values of s_{13}: a) $s^2_{13} = 0.2$ and the hierarchy $|m_3| \gg |m_2| \sim |m_1|$, b) $s^2_{13} = (0.025)^2$ and the hierarchy $|m_2| \sim |m_1| \gg |m_3|$.

6. Models of neutrino masses and mixing

We conclude these lectures with a brief discussion of models for neutrino masses and mixing (for a more extensive discussion see e.g. [3]). In the bottom-up approach, one can reconstruct the neutrino mass matrices in the charged lepton diagonal basis and then try to understand how can they be obtained, e.g. by the see-saw mechanism, in a "natural" way. In the top-down approach one postulates some theories of fermion masses, including neutrinos, and compares them with the data. One interesting possibility is to base such theories on Grand Unification and some flavour (horizontal) symmetries. We begin our discussion with the bottom-up approach.

The attractive criterion of stability of neutrino mass textures with respect to quantum corrections selects the hierarchical mass pattern and the inversly hierarchical pattern with the two degenerate neutrinos of different CP parties, $m_1 \approx -m_2$, as the most plausible candidates for such textures. For several reasons, the hierarchical pattern looks particularly natural and we shall focus on it. The question is then whether the hierarchical pattern of the mass eigenvalues is compatible with two large mixing angles and whether both can be explained by the see-saw mechanism.

Under the assumption of a hierarchical structure, the neutrino mass matrix reconstructed from the data is:

$$\hat{m} = U \begin{pmatrix} \epsilon_1 & & \\ & \epsilon_2 & \\ & & m \end{pmatrix} U^T \tag{51}$$

where U is the measured mixing matrix with two large mixing angles, small θ_{13} and $\epsilon_{1,2}/m \sim \frac{1}{10}$. Thus, we get

$$\hat{m} = m \begin{pmatrix} \delta & \epsilon & \epsilon \\ \epsilon & 1+\eta & 1+\eta \\ \epsilon & 1+\eta & 1+\eta \end{pmatrix} \tag{52}$$

where δ , ϵ , η $\ll$1 and they can differ from each other by factors 0(1). The zeroth order approximation to the matrix (52) reads

$$\hat{m}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{53}$$

Its eigenvalues are $\lambda_{1,2,3} = 0,\ 0,\ 2$ and the mixing angles $\theta_{23} = \pi/4$, $\theta_{13} = 0$ and θ_{12} is undefined. The first question to address is whether we can get the matrix (53) by the see-saw mechanism, with its eigenvalues and the mixing angles stable without a fine-turning in the underlying high energy physics. Note that for the hierarchy of the eigenvalues in (53) it is not sufficient that the non-zero entries are of order 1. They must be fine-tuned to be very close to one or must have the structure

$$\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix} \tag{54}$$

which also gives det=0, now with no fine-turning of the parameters a and b. The see-saw mechanism gives

$$m = Y^T \frac{1}{M} Y \tag{55}$$

and taking a diagonal matrix $\frac{1}{M}$ and a Yukawa coupling matrix

$$Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & Y_\mu & Y_\tau \end{pmatrix} \tag{56}$$

we indeed get

$$\hat{m} = \frac{1}{M} \begin{pmatrix} 0 & 0 & 0 \\ 0 & Y_\mu^2 & Y_\mu Y_\tau \\ 0 & Y_\mu Y_\tau & Y_\tau^2 \end{pmatrix} \tag{57}$$

Thus, the stability of the hierarchy of the mass eigenvalues is assured by the see-saw mechanism provided only one right-handed neutrino has large Yukawa couplings to only the two active neutrinos ν_μ and ν_τ.(The same effect can be obtained by making other right-handed neutrinos very heavy.) The structure (57) does not assure a large mixing angle yet. For that we need $Y_\mu \approx Y_\tau$ but no excessive fine-tuning is needed. To reproduce the full matrix (52) we need some corrections to the approximation (53) and (57). We can get some information about the needed structure of such corrections by rotating the matrix (52), first, by an angle θ_{23} close to $\pi/4$ and then by $\theta_{13} \sim \epsilon$. An inspection of the mass matrix in that new basis shows that for a large solar angle one needs $|\eta - \delta| < \epsilon$. To avoid cancellations, let us take $\delta < \epsilon$ and $\eta < \epsilon$. One can consider then several possibilities. Taking $\delta \approx \epsilon$ and $\eta < \epsilon$ one gets $\Delta m^2_{23} \approx 4m^2$, $\Delta m^2_{12} \approx m^2\epsilon^2$, $m_{ee} \approx \sqrt{\Delta m^2_{12}}$, $\theta_{12} \sim O(1)$ but smaller than $\pi/4$, $\theta_{13} \approx \sqrt{\Delta m^2_{12}/\Delta m^2_{23}}$. With $\eta \approx \epsilon$, $\delta \ll \epsilon$ we obtain $\Delta m^2_{12}/\Delta m^2_{23} \approx \epsilon^2$ and $\theta_{13} \approx \sqrt{\Delta m^2_{12}/\Delta m^2_{23}}$ (as before) but $m_{ee} \approx 0$. If $\delta, \eta \ll \epsilon$, e.g. $\eta = 0$, $\delta \approx \epsilon^2$ we have $m_a = \pm m\epsilon, m$, $\sin^2 2\theta_{12} \approx 1 - \epsilon/4$, $\Delta m^2_{12} \approx m^2\epsilon^2$, which are difficult to be reconciled with the LA solution. Those examples show the importance of small parameters.

Consider, for instance,

$$Y = \left|\begin{array}{cccc} \nu_e & \nu_\mu & \nu_\tau & \\ \hline 0 & 0 & 0 & \nu_1^R \\ Y_e^2 & Y_\mu^2 & Y_\tau^2 & \nu_2^R \\ Y_e^3 & Y_\mu^3 & Y_\tau^3 & \nu_3^R \end{array}\right|$$

One can easily check that the matrix (55) has det $m = 0$ and its structure is of the form (52) with $\delta \sim (Y_e^i)^2$, $\epsilon \sim Y_e^i Y_{\mu,\tau}^i$. So, with $Y_e^i \ll Y_\mu^i, Y_\tau^i$ and $Y_\mu^2 \ll Y_\mu^3$, $Y_\tau^2 \ll Y_\tau^3$ we indeed get small parameters $\delta \ll \epsilon$. Moreover, the 2×2 large submatrix now reads:

$$\left(\begin{array}{cc} \frac{(Y_\mu^2)^2}{M_2} + \frac{(Y_\mu^3)^2}{M_3} & \frac{Y_\mu^2 Y_\tau^2}{M_2} + \frac{Y_\mu^3 Y_\tau^3}{M_3} \\ \frac{Y_\tau^2}{M_2} + \frac{Y_\tau^3 Y_\mu^3}{M_3} & \frac{(Y_\tau^2)^2}{M_2} + \frac{(Y_\tau^3)^2}{M_3} \end{array} \right) \qquad (58)$$

Its det $= \frac{1}{M_2 M_3}(Y_\mu^2 Y_\tau^3 - Y_\mu^3 Y_\tau^2)^2 \equiv \eta < \epsilon$. The masses M_2 and M_3 are the Majorana masses of the right-handed neutrinos ν_2^R, ν_3^R respectively. So, with the hierarchical couplings of the active neutrinos to two right-handed neutrinos we can obtain a stable hierarchy of the mass eigenvalues $m_i = 0, \eta, 2$, and two large mixing angles (for more details see ref. ([3])). The see-saw mechanism can explain the hierarchical pattern of masses and the observed large mixings provided there are some hierarchies in the neutrino Yukawa couplings and/or the right-handed neutrino Majorana masses.

We conclude these lectures with a few more remarks on the top-down approach to neutrino masses. Usually, that problem is addressed as part of a more general question about the theory of fermion masses. It seems that the most attractive possibility is some family (horizontal) symmetry combined with the Froggatt-Nielsen mechanism. In supersymmetric models, horizontal symmetries may also correlate the fermion and sfermion masses and could solve the fermion mass and the supersymmetric $FCNC$ problems, simultaneously. The simplest possibility is the horizontal $U(1)$ symmetry, assumed to be a spontaneously broken gauge symmetry. The scalar field that breaks spontaneously the gauge symmetry is often called the Froggatt-Nielsen field.

We shall consider $U(1)$ models with single $U(1)$ horizontal symmetry and one Froggatt-Nielsen field Φ which, we assume carries the charge $q_\Phi = -1$ of the horizontal $U(1)$. Furthermore, we assume that all $U(1)$ matter charges are non-negative, with the third generation fermion doublets and the Higgs doublets (we talk about $MSSM$) having charge zero, $q_3 = h_1 = h_2 = 0$. Acceptable fermion masses and mixings (including the neutrino sector) can be obtained from a superpotential of the form

$$W \sim y^u_{ij}\Theta(q_i + u_j + h_2)(\frac{\Phi}{\Lambda})^{q_i+u_j+h_2} Q^i U^j H_2 \tag{59}$$

plus similar terms for the down quarks, charged leptons, Dirac neutrino terms and RH neutrino Majorana masses. The neutrino masses are obtained from the see-saw mechanism. Here $\Phi/\Lambda \sim \lambda$ (the Cabibbo angle), Λ is some fundamental scale and Φ is the vaccum expectation value of the scalar component of a chiral superfield Φ. The constants y_{ij} are arbitrary $\mathcal{O}(1)$ coefficients, not predicted by the model. The Yukawa couplings then read $Y_{ij} = y_{ij}\lambda^{q_i+u_j+h_2}$ etc. Yukawa interactions and the RH Majorana masses are constrained by the family $U(1)$ charges, once we demand the $U(1)$ invariance of the theory. With a proper $U(1)$ charge assignement one can "explain" the observed pattern of masses and mixing. We get, e.g. $m_{u_i} \sim v\lambda^{q_i+u_j+h_2}$ and $V^{CKM}_{ij} \sim \lambda^{|q_i-q_j|}$.

An interesting and ambitious approach is to consider theories that are $G_{GUT} \times G_{FLAVOUR}$ invariant. For a very nice discussion of neutrino mass models based on those ideas we refer the reader to ([3]).

7. Summary and outlook

The evidence for neutrino masses and mixing is the first experimental indication for physics beyond the Standard Model. There are strong theoretical arguments suggesting that this is an evidence for new physics at high energy scale, close to the Grand Unification scale. The SM or

its supersymmetric extension should then be viewed as the effective low energy theory, with the active neutrino Majorana masses described by additional (non-renormalizable) terms in the effective theory.

The reconstruction of the neutrino mass matrix from experimental data is subject to a number of ambiguities, not only related to the present experimental precision, but following from the fact that only the mass squared differences and mixing angles are accessible experimentally. In consequence, several different textures can be considered in agreement with the data. If the origin of the neutrino mass is at a large energy scale M, the "measured" textures are obtained by evolving the original mass matrices from the scale M to low energies, i.e. by the inclusion of quantum corrections. An attractive requirement of small quantum corrections may be a constraint on the acceptable textures of the neutrino mass matrices.

A very plausible source of small neutrino masses is see-saw machanism, with very heavy right-handed (sterile) neutrinos. It can naturally accommodate a hierarchical pattern of the neutrino masses and the two large mixing angles.

Horizontal (flavour) symmetries are likely to play a role in understanding the fermion masses (including neutrinos). Such models can be predictive (and tested) in other sectors (e.g. leptogenesis, $FCNC$ and CP violation in supersymmetric models).

Acknowledgements

I would like to thank Piotr Chankowski for continuous enjoyable collaboration on the subject of neutrino physics. This work is partially supported by the Polish State Committee for Scientific Research grant 5 P03B 129 24 for years 2003-2005. I thank the group of Hans Peter Nilles for its hospitality during my visit to the University of Bonn. That visit has been possible thanks to the Research Award of the Humboldt Foundation.

Appendix

The two-component Weyl spinor notation is particularily convenient for dealing with Majorana particles. Here we explain this notation shortly. More details can be found in modern textbooks on QFT (see e.g. [1, 4]).

In four dimensions, the Lorentz group or more precisely its covering group $SL(2, C)$ has two non-equivalent complex two-dimensional representations denoted as $(1/2, 0)$ and $(0, 1/2)$. The Grassmann fields (or

fermionic field operators) transforming according to these representations are conventionally written as λ_α and $\bar{\chi}^{\dot{\alpha}}$ and called left- and right-handed spinors, respectively. Since the complex conjugation of a left-handed spinor λ_α transforms as a right-handed one ($(\lambda_\alpha)^* \sim \bar{\lambda}_{\dot{\alpha}}$), the fermion content of any Lagrangian can be specified by listing only the left-handed spinors used for its construction.

If the two left-handed fields λ_α and χ_β transform as representations R and R^*, respectively under some symmetry group (global or local) of the theory, they can be combined to form a Dirac bispinor:

$$\psi_{(\lambda)} = \begin{pmatrix} \lambda_\alpha \\ \bar{\chi}^{\dot{\beta}} \end{pmatrix}, \qquad \bar{\psi}_{(\lambda)} = \left(\chi^\alpha, \bar{\lambda}^{\dot{\beta}}\right), \tag{60}$$

transforming as R and R^*, respectively. The raising and lowering of Weyl spinor indices is done with the help of the antisymmetric tensors $\epsilon^{\alpha\beta}$, $\epsilon_{\alpha\beta}$, $\epsilon_{\dot{\alpha}\dot{\beta}}$ and $\epsilon^{\dot{\alpha}\dot{\beta}}$:

$$\lambda^\alpha = \epsilon^{\alpha\beta}\lambda_\beta, \quad \lambda_\alpha = \lambda^\beta \epsilon_{\beta\alpha}, \quad \bar{\chi}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\bar{\chi}_{\dot{\beta}}, \quad \bar{\chi}^{\dot{\alpha}} = \bar{\chi}_{\dot{\beta}}\epsilon^{\dot{\beta}\dot{\alpha}}. \tag{61}$$

The two kinetic terms for λ and χ can be then rewritten in the familiar form

$$\begin{aligned} \mathcal{L}_{\text{kin}} &= i\bar{\lambda}\bar{\sigma}^\mu\partial_\mu\lambda + i\bar{\chi}\bar{\sigma}^\mu\partial_\mu\chi \\ &= i\bar{\lambda}\bar{\sigma}^\mu\partial_\mu\lambda + i\chi\sigma^\mu\partial_\mu\bar{\chi} + (\text{total der}) = i\bar{\psi}_{(\lambda)}\gamma^\mu\partial_\mu\psi_{(\lambda)} \end{aligned} \tag{62}$$

where the Dirac matrices γ^μ in the Weyl representation are constructed as

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu_{\alpha\dot{\beta}} \\ \bar{\sigma}^{\mu\dot{\alpha}\beta} & 0 \end{pmatrix} \tag{63}$$

with $\sigma^\mu \equiv (I, \boldsymbol{\sigma})$, $\bar{\sigma}^\mu \equiv (I, -\boldsymbol{\sigma})$ ($\boldsymbol{\sigma}$'s are the Pauli matrices). For such a pair of Weyl fields also a Dirac mass term can be constructed

$$\mathcal{L}_{\text{mass}} = -m\left(\lambda^\alpha\chi_\alpha + \bar{\lambda}_{\dot{\alpha}}\bar{\chi}^{\dot{\alpha}}\right) = -m\bar{\psi}_{(\lambda)}\psi_{(\lambda)} \tag{64}$$

If the field λ (χ) has no left-handed partner transforming in the complex conjugate representation R^* (R), it is convenient to introduce chiral Dirac bispinors

$$\psi_{(\lambda)L} = \begin{pmatrix} \lambda_\alpha \\ 0 \end{pmatrix}, \quad \psi_{(\chi)R} = \begin{pmatrix} 0 \\ \bar{\chi}^{\dot{\alpha}} \end{pmatrix}. \tag{65}$$

For chiral Dirac bispinors e.g. $\psi_{(\lambda)L} = P_L\psi_{(\lambda)L}$ etc., where $P_L \equiv (1 - \gamma^5)/2$. Note that (see eq.(60)),

$$\overline{\psi_{(\lambda)L}} = (\,0 \quad \bar{\lambda}_{\dot{\alpha}}\,), \qquad \overline{\psi_{(\chi)R}} = (\,\chi^\alpha \quad 0\,). \tag{66}$$

The typical Yukawa coupling of a scalar field ϕ in the representation R_ϕ and two left-handed Weyl spinors λ and χ transforming as representations R_λ and R_χ, respectively (such that $1 \subset R_\phi \times R_\lambda \times R_\chi$) can be written as (we omit the Clebsch-Gordan coefficients)

$$\begin{aligned}\mathcal{L}_{\rm Yuk} &= -Y\phi\lambda\chi - Y^*\phi^\dagger\bar\lambda\bar\chi \\ &= -Y\phi\overline{\psi_{(\chi)R}}\psi_{(\lambda)L} - Y^*\phi^\dagger\overline{\psi_{(\lambda)L}}\psi_{(\chi)R}\end{aligned} \tag{67}$$

The Yukawa part (12) of the SM Lagrangian and the mass terms (10) are the example of (67) and (64), respectively, with fields u^c, d^c and e^c playing the role of χ, and q and l (or u, d and e) playing the role of λ.

Finally, Weyl spinor fields λ_α which are singlets of all unbroken symmetries of the theory can form 4-component Majorana bispinors

$$\psi_{(\lambda){\rm Maj}} = \begin{pmatrix}\lambda^\alpha \\ \bar\lambda^{\dot\beta}\end{pmatrix}. \tag{68}$$

Of course in this case $\psi_{(\lambda)} = C\bar\psi^T_{(\lambda)} \equiv \psi^c_{(\lambda)}$ which means that the field is self-conjugate. For such a field a Majorana mass term can be formed

$$\begin{aligned}\mathcal{L}_{\rm Maj} &= -\frac{1}{2}m(\lambda\lambda + \bar\lambda\bar\lambda) = -\frac{1}{2}m\left(\overline{\psi_{(\lambda)R}}\psi_{(\lambda)L} + \overline{\psi_{(\lambda)L}}\psi_{(\lambda)R}\right) \\ &\equiv -\frac{1}{2}m\bar\psi_{(\lambda)}\psi_{(\lambda)} \equiv -\frac{1}{2}m\psi^T_{(\lambda)}C\psi_{(\lambda)}\end{aligned} \tag{69}$$

where C is the charge conjugation matrix. The Majorana mass term (11) is precisely of this form.

A quantum field λ_α, say in representation R of a symmetry group reads:

$$\lambda_\alpha \sim \sum\!\!\!\!\!\!\int d^3k[b_L(k)a_\alpha(k)e^{-ikx} + d^\dagger_R(k)b_\alpha(k)e^{ikx}] \tag{70}$$

It annihilates particle (charges R) with helicity $\lambda = -1/2$ and momentum k, and creates antiparticle (charges R^*) with $\lambda = +1/2$ and momentum k. The field $\lambda^\dagger \sim \bar\lambda$ gives opposite effects. The fields λ and $\bar\lambda$ are connected to each other by the CP transformation:

$$CP\lambda(CP)^{-1} = \bar\lambda \tag{71}$$

Parity and charge conjugation can be separately defined on the set $(\lambda) = \lambda_i + \lambda_i^c$ such that the fields $\lambda_i(\lambda_i^c)$ are in representation $R(R^*)$ of the symmetry group, or if λ is neutral under all unbroken internal symmetries.

References

[1] See, for instance, Pokorski S., Gauge Field Theories, Cambridge Monographs on Mathematical Physics, Cambridge University Press, Second Edition 2000

[2] For a review on quantum corrections to neutrino masses and mixing see P.H. Chankowski and S. Pokorski, International Journal of Modern Physics A, Vol. 17, 575-614 (2002)

[3] For a review of the recent work on the neutrino mass theory see G. Altarelli and F. Feruglio, Theoretical Models of Neutrino Masses and Mixing, hep-ph/0206077

[4] Peskin M. E. and Schroeder D. V., An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts, 1997.

NEUTRINOS

Jacques Bouchez
CEA/DSM/DAPNIA
CEA-Saclay F91191 Gif-sur-Yvette Cedex
bouchez@hep.saclay.cea.fr

Abstract We give an overview of the experimental activities on neutrino properties, with an emphasis on the mass of the neutrino. After a historical introduction, we review the present experimental situations and describe the projects foreseen in future years.

Keywords: Neutrinos, mass, oscillations

Introduction

These lectures are focused on the determination of the properties of the neutrinos, with a strong emphasis on their mass, and mainly from an experimental point of view. Neutrinos were postulated in 1930 by W. Pauli, discovered only 26 years later, and have slowly revealed some of their properties, while some still remain unknown. This is of course due to the difficulty of detecting these particles, the only ones in nature which only feel the weak interaction, so that their detection necessarily involves large detectors to get a useful signal, and these detectors have to be efficiently shielded against any backgrounds which would very quickly overwhelm the faint neutrino signals.

This explains for example why the massive character of neutrinos has only been proven in the last few years (more than 40 years after their discovery), although we still ignore the value of their masses, or are unable to tell if neutrinos and antineutrinos are or not the same particle.

The first chapter gives a brief historical survey of neutrino physics.

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 111–170.

The second chapter describes attempts to directly determine the neutrino masses and related properties.

Chapters 3 to 5 are devoted to the flavor oscillations of neutrinos; the observation of these oscillations has definitely proven that neutrinos have mass, but more experiments are needed to completely describe the mass pattern. Chapter 3 focusses on solar neutrinos, chapter 4 on atmospheric neutrinos and chapter 5 on future projects trying to measure a still unobserved flavor oscillation, between ν_e and ν_μ; its observation would open the exciting possibility to study CP symmetry in the leptonic sector, which, if violated, would have fundamental cosmological consequences.

Finally, the chapter 6 gives the prospects to determine the absolute mass scale for the neutrinos.

For those interested in neutrino physics and their latest developments, I recommend to consult the following web page :
http://www.hep.anl.gov/ndk/hypertext/nu_industry.html
which has links to most experiments and neutrino meetings.

Some recent reviews on neutrinos can be found in the litterature [1].

1. Neutrinos : a historical survey

1.1 The continuous beta spectrum

When in 1914, J. Chadwick measured the beta spectrum of radium E (that is ^{210}Bi), its continuous character came as a big surprise (as other types of radioactivity were characterized by monoenergetic lines, interpreted as the energy difference between initial and final states). This led to many speculations which were refuted one after the other by very smart experiments.

- It was first supposed that each radioactive transition emitted several beta rays. But in 1924, K.G. Emeleus [2], using a Geiger counter, was able to count separately beta and alpha emissions of the Bismuth-Polonium cascade and found equality.

- The explanation pushed forward by Lise Meitner was that the beta rays were secondary particles carrying away only part of the total energy. C.D. Ellis [3], in 1927, made a calorimetric experiment to measure the heat deposited by a radium E source, and found that the temperature increase (1 milli degree) corresponded exactly to the mean beta energy (390 keV) and not to the end-point energy (1.05 MeV).

So the mystery remained : A single beta ray was emitted with the full energy detectable in a calorimeter.

1.2 The neutrino hypothesis

Two equally challenging hypotheses were formulated to explain this energy crisis :

- N. Bohr suggested that energy was **not** conserved in beta decays, and that could explain the mysterious (at that time) source of energy in stars.
- W. Pauli, on the contrary, insisted on the necessity of energy conservation and proposed, in a celebrated letter [4], that the missing energy was carried away by a hypothetical, light and very penetrating particle, which he nicknamed neutron (and that Fermi renamed neutrino when the neutral partner of the proton was later discovered by Chadwick).

1.3 Early attempts at detection

The first attempts at neutrino detection relied on the possibility for this particle to ionize atoms owing to a magnetic moment (which Pauli estimated 1/100 or less than that of the electron, due to the high penetrating power of the neutrino). Nahmias [5] in particular made careful measurements and to get rid of the cosmic background, installed its detector in the deepest station of London subway : this was the first **underground** neutrino experiment. Negative results allowed Nahmias to put a limit of $2\ 10^{-4}$ Bohr magneton on the neutrino magnetic moment.

Other approaches consisted in proving that beta decays obeyed standard 3 body kinematics, with energy and momentum conservation :

- Henderson [6], following an idea by Ellis, showed that the 2 cascades $^{212}Bi \rightarrow^{212} Po \rightarrow^{208} Pb$ $(\beta\alpha)$ and $^{212}Bi \rightarrow^{208} Tl \rightarrow^{208} Pb$ $(\alpha\beta)$ gave the same energy balance when taking the **maximal** electron energy, while using the mean energy gave inconsistent results
- Crane and Halpern [7], using ^{38}Cl produced by the newly born cyclotron, showed in 1938 that the decays observed in a cloud chamber were consistent with the 3 body kinematics by building what we now call a Dalitz plot
- In 1951, Sherwin could show the equality, with a 10 % accuracy, of missing energy and momentum in ^{52}P decays [8].

However, these tests could not logically refute the non-conservation of energy. The only real proof of existence was to detect a neutrino far away from its production. Theoreticians (Bethe, Peierls [9]) had however estimated cross-sections of neutrinos on nuclei in the absence of magnetic

moment and found desperately low values around 10^{-44} cm^2, which, according to them, made impossible any neutrino detection. One should however mention the unsuccessful attempt by Crane in 1939 [10] who put a 1 millicurie source of radium inside a bag filled with 3 pounds of salt (NaCl) with the hope to observe the transmutation of some ^{35}Cl atoms into ^{35}S, which has a half-lifetime of 87 days. After 3 months of irradiation followed by a chemical extraction of sulfur, he found no signal of ^{35}S decay and could put an upper limit of 10^{-30} cm^2 on the capture cross-section : this certainly was the very first **radiochemical** experiment for neutrino detection, and the same principle was later used in the first solar neutrino experiment in 1968.

1.4 The neutrino discovery

Things became more favorable for neutrino hunters after the second world war, with the appearance of nuclear reactors. It was soon realized (actually by Fermi) that they were extremely strong sources of neutrinos (a fission dissipates 200 MeV and gives 6 neutrinos through cascade decays of fission products, so that a reactor with a thermal power of 1 GW produces isotropically 2 10^{20} neutrinos per second !)

The most favorable reaction to detect these neutrinos (actually anti-neutrinos) was to use the inverse beta decay reaction on free protons,

$$\bar{\nu} + p \rightarrow e^+ + n$$

since this reaction has a threshold of 1.8 MeV while the neutrino spectrum goes up to 8 MeV (thresholds for other atoms are much higher). A few such interactions could thus be observed every hour in a detector with several tens of kilograms of hydrogen-rich material, if located close enough to the nuclear core.

This is how Cowan and Reines finally discovered neutrinos in 1956 [11]. Their detector, located 11 meters from the core, consisted in layers of cadmium-loaded water used as neutrino target sandwiched with layers of liquid scintillator observed by photomultipliers. The signal consisted in, first 2 synchronous signals in the upper and lower layers of scintillator produced by the two 511 keV gammas emitted by the positron annihilation, followed a few microseconds later by gammas in the 2 same layers produced by the capture of the final neutron on cadmium after it was thermalized in the water.

This very specific signature helped a lot in discriminating against backgrounds, which were also reduced by strong passive shieldings against gammas and neutrons surrounding the detector. The candidate events

increased from 1/hour when the reactor was off to 4/hour when it was on. A true neutrino signal of 3/hour (which did not vary when the shielding was increased, but decreased when normal water was replaced by heavy water) had been seen.

This technique is still basically used to detect neutrinos from reactors: It has been improved over the years so that the signal efficiency is now above 50% rather than the few % in the pioneer experiment.

1.5 Neutrinos and antineutrinos

After Dirac produced his theory of spin 1/2 particles and predicted the existence of antiparticles with opposite charges, it was natural to wonder if neutrinos and antineutrinos were or not different particles. As neutrinos have no electric charge, there is a possibility that they are truly neutral and carry no charge of whatever nature. This possibility has been put forward by E. Majorana, after whom self-conjugate neutrinos are now named. If some internal charge (such as a leptonic charge) is carried by neutrinos, antineutrinos will carry the opposite charge and be different : they will be so-called Dirac neutrinos.

Before answering the question, one has to label what would be a neutrino and what would be an antineutrino : one has decided to call antineutrinos those which are produced together with an electron, while neutrinos are produced by β^+ decays together with a positron (so that β decays produce a lepton-antilepton pair, one electrically charged and the other neutral). In a nucleus, the transition between a neutron and a proton produces an antineutrino. This antineutrino is thus able to turn a proton into a neutron (this is the discovery detection). Now neutrinos will for the same reason certainly be able to transform a neutron into a proton inside a nucleus. But will neutrinos be also able to transform a proton into a neutron, or equivalently antineutrinos be able to transform a neutron into a proton ? If yes, one would tend to admit that neutrinos are their own antiparticle, while they would be different particles if the answer is no. (But as we will see later, this reasoning happens to be too naïve and is wrong). Anyhow, an experimental test was done, where a tank filled with chlorine, a prototype done by R. Davis for his solar neutrino experiment, was brought near the Savannah River reactor (used by Reines) to eventually observe the transmutation of ^{37}Cl into ^{37}Ar by the reactor antineutrinos. No such transmutation was observed, and it was concluded that neutrinos and antineutrinos actually were different particles, carrying a leptonic charge (1 for neutrinos and electrons, -1 for antineutrinos and positrons) which was conserved in interactions. This meant that neutrinos were Dirac particles.

But things became much more complicated when parity conservation was shown to be violated in weak interactions by Miss Wu in her celebrated experiment [12]. This violation was found to be maximal, and this meant that spin was playing an essential role in weak interactions: amplitudes depend upon the helicity of the (anti)neutrino produced in beta decays. As the parity violation was found to be maximal, it meant that only left-handed helicity neutrinos were produced in β^- decays, and conversely β^+ decays were producing neutrinos of right helicity.

This peculiarity had been guessed by Lee and Yang [13] who modified the Fermi theory by adding a factor $(1 - \gamma_5)$ in the current-current Hamiltonian describing beta decays. This factor actually selects a given chirality for the neutrino, and the opposite one for the antineutrino. When these neutrinos are relativistic, chirality and helicity are nearly the same, and this so-called V-A theory could explain Miss Wu's observations.

But this imposes a reconsideration of the difference between neutrinos and antineutrinos : is the observed difference between them when looking at Chlorine transmutation due to an intrinsic difference (leptonic charge) or just a spin effect ? It could well be that neutrinos and antineutrinos are the same particle, but that due to V-A, only left-handed neutrinos are produced with electrons and right-handed neutrinos are produced with positrons. As long as neutrinos stay ultrarelativistic, so that chirality and helicity are nearly equivalent, V-A prevents left-helicity neutrinos from interacting with protons and right-helicity neutrinos from interacting with neutrons.

In the limit where neutrinos are massless, chirality and helicity are intrinsic conserved quantities, and the distinction between the two vanishes : neutrinos are described as Weyl particles (2-component massless spinors) and the Dirac or Majorana descriptions become mathematically equivalent. This is actually the way neutrinos enter the minimal standard model of electroweak interactions : neutrinos are left-handed, antineutrinos are right-handed, and the two other degrees of freedom, which would be anyway perfectly sterile due to V-A, simply do not exist. The apparent lepton number conservation is just a consequence of V-A.

When neutrinos have a mass, the alternative between Majorana and Dirac descriptions could in principle be tested : for example, a neutrino beam impinging on a fixed nucleus target will be described as left helicity particles, while a nucleus beam with a higher speed than the neutrino beam, and going in the same direction, would see them as right helicity particles provided we use the same helicity convention in the center-of-mass frame; then Dirac neutrinos would produce electrons in the first case and be sterile in the second case, while Majorana neutrinos would

produce electrons in the first case and positrons in the second case (all this being true up to correction factors of order $(m_\nu/E_\nu)^2$). Such a gedanken experiment is evidently totally unrealistic and in practice, only neutrinoless beta decays, addressed in the last chapter, could allow us to determine whether neutrinos are Dirac or Majorana particles.

1.6 Three families of neutrinos

W. Pauli, when postulating the neutrino, increased the elementary bricks of the microscopic world from 2 (proton and electron) to 3. But as we all know, the zoology of "elementary" particles showed an exponential increase with the experimental progress. The neutron and the positron (the first antiparticle) were soon discovered. The muon was seen in cosmic rays, and it took some time to realize it was not the hypothetic pion mediating the nuclear force, although it had the expected mass, but rather a heavy electron. Pions were copiously produced with the start of GeV accelerators, and the study of their decays was puzzling. Why was the main decay into a muon and a neutral light particle (a neutrino so that the lepton number was conserved) rather than electron-neutrino ? The V-A theory had the explanation: lepton-neutrino decay is forbidden by V-A for spin 0 particles like pions in the limit of massless leptons. And this interdiction is only violated due to the helicity-chirality mismatch for massive charged leptons. So, although the phase space for electron-neutrino is much higher than for muon-electron, V-A rule dominates and the electron-neutrino branching ratio is only $1.2\ 10^{-4}$ due to the much lower electron mass. But a question was to be answered: was this ν_π neutrino the same as the ν_β emitted in radioactive decays ? If yes, this neutrino, when interacting with nuclei, should produce muons and electrons roughly in equal numbers. If it was a second neutrino variety, specifically related to the muon, it should produce only muons and no electrons. The absence of radiative decays of muons into electrons, but rather in electron and two neutrinos, suggested that there were 2 species of neutrinos associated with two species of charged leptons, with different lepton numbers between the two lepton families. But this had to be tested. The new accelerator at Brookhaven was used to produce a secondary pion beam which after decay, sent neutrinos to a detector placed behind a very thick steel shielding to absorb all charged particles [14]. 34 interactions producing a muon were observed, while only 6 electron or gamma showers were observed. The ν_π was different from the ν_β, and the two neutrino species were labelled ν_μ and ν_e, referring to their associated charged lepton.

When the third charged lepton (the τ) was discovered at SLAC, it became natural to link it with a third variety of neutrino, the ν_τ. The direct proof of the ν_τ existence was brought only in 2000, when the DONUT experiment [15], using the beam dump technique with the high energy proton beam at Fermilab, could produce a so-called prompt neutrino beam. Among all secondary particles produced by proton interactions in a thick shielding, only those containing a charm quark, having a very short lifetime, decayed quickly enough to emit high energy decay particles, and among them tau-ν_τ pairs, while pions and kaons, being much more longlived, lost their energy before decaying and emitted only very low energy ν_e's and ν_μ's. A detector consisting of photographic emulsions could observe the interactions of these high energy neutrinos producing τ leptons, recognized by their short path of hundreds of microns inside the emulsion before their decay.

The existence of 3 families of neutrinos had previously been proven indirectly by LEP experiments, which deduced from the width of the Z^0 gauge boson that it has to decay into 3 different varieties of ν-$\bar{\nu}$ pairs, each contributing 110 MeV to the total width.

1.7 Checking the standard model with neutrinos

The latest result relies on the so-called standard model of electroweak interactions, which was slowly built from experimental observations and theoretical progress during the 60's and the 70's. The success of this theory culminated with the discovery at CERN of the 2 gauge bosons W and Z. The study of neutrino interactions played an important role in the conception of this standard model, which unifies weak and electromagnetic interactions in a single theoretical frame.

1.8 The discovery of neutral currents

As we have seen, neutrino species are linked to the charged lepton with which they are associated. This association can be in a decay, a hadron decaying leptonically or semi-leptonically (that is together with other hadrons) into a charged lepton l and an antineutrino $\bar{\nu}_l$ (or an antilepton $\bar{l}$ and a neutrino ν_l); or it can be in so-called charged current interactions (mediated by W) where a ν_l interacts with a hadron to give a lepton l together with hadrons. This association defines the neutrino flavor as the flavor of the associated charged lepton.

An important discovery was done in 1973 when an experiment with the bubble chamber Gargamelle submitted to a ν_μ beam observed first one [16], then several events [17] interpreted as the elastic diffusion of a neutrino upon an electron. Kinematically, these knocked electrons keep

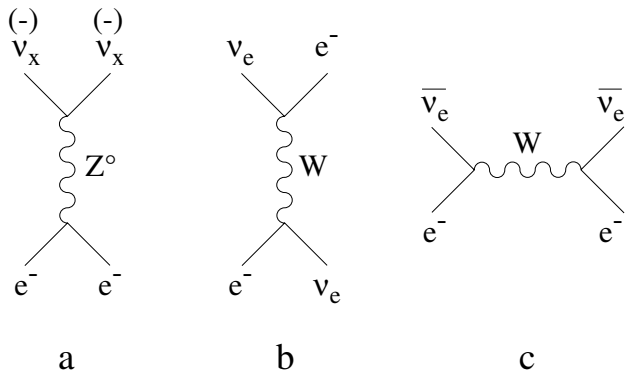

Figure 1. Exchange diagrams describing the diffusion of neutrinos on electrons : a) Z^0 neutral current diagram, valid for all neutrino flavors; b and c) W charge current exchange diagrams, only present for ν_e (resp. $\bar{\nu_e}$)

the direction of the beam when the neutrino energy is much higher than the electron mass. But such a diffusion is impossible if the neutrino interactions are charged-current interactions, since necessarily a muon should appear in the final state (as for example $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$). This new way of interacting can only be explained if there exists neutral currents, mediated by the boson Z^0, as shown on fig. 1, where the initial neutrino appears also in the final state.

This observation was an important step towards electroweak unification, since at the time several scenarios were possible, with or without neutral currents. The existence of these neutral currents was later confirmed in neutrino interactions on nuclei in which no final charged lepton was observed.

A more precise measurement of ν_μ and $\bar{\nu_\mu}$ cross-sections on electrons was performed by experiments like CHARM at CERN and gave a quantitative estimate of the so-called Weinberg angle, θ_W, the free parameter in the Weinberg-Salam scenario of unification, corresponding to the "direction" of the spontaneous breaking of SU(2)xU(1). These reactions being purely leptonic do not suffer from hadronic corrections and are readily computable [18]. Cross-sections can be expressed as functions of two vector and axial coupling constants, g_V and g_A, whose values are $g_V^{SM} = -0.5 + 2\sin^2\theta_W$ and $g_A = -0.5$ in the standard electroweak model. The interpretation of these measurements is shown on figure 2.

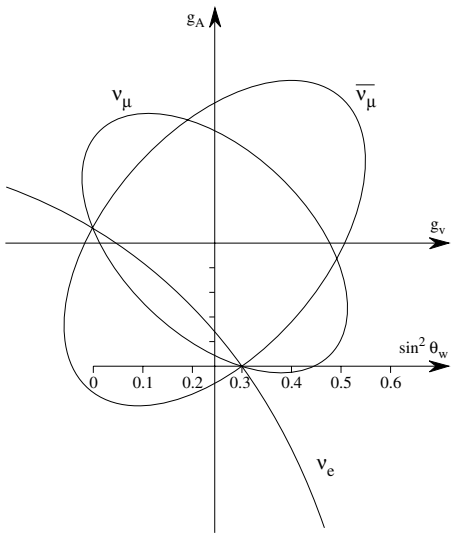

Figure 2. Determination of vector and axial coupling constants g_V and g_A from neutrino cross-sections on electrons. The cross sections for ν_μ, $\bar{\nu_\mu}$ and ν_e define 3 different ellipses which should intersect on the standard model line and allow to determine the Weinberg angle θ_W (the curves are drawn for $\sin^2\theta_W = 0.3$) For clarity, experimental errors are omitted, their effect is to give some thickness to the ellipses

1.9 The W-Z interference

Once neutral currents were established, it was important to check if they were correctly described by the Weinberg-Salam model. One very clean way to test the model was to measure the diffusion of ν_e on electrons, since no strong interaction corrections are necessary. In this case, both charged-current and neutral current interactions are possible (see figure 1), and since they can contribute to the same helicity amplitudes, one should see the interference term between the 2 processes in the total cross-section.

The standard model actually predicts σ_W, σ_Z and σ_I corresponding to W exchange, Z exchange and the interference term, to be in the ratio $4 : 0.35 : -2.1$. An experiment was performed in Los Alamos in 1990 [19]. It sent the high intensity proton beam on a heavy target in which produced pions were stopped and either decayed into μ^+ and ν_μ for π^+'s or were absorbed by a nucleus for π^-'s, the subsequent μ^+ decaying into $\bar{\nu_\mu}$ and ν_e. The ν_e diffusion on electrons dominates, due to the possibility of charged-currents, while ν_μ's and $\bar{\nu_\mu}$'s give a small computable background due to the neutral currents. The extracted signal (234 ± 35 events) proved the significant presence of a negative

interference term for ν_e diffusion with the expected intensity (within the 20% precision). As shown on fig. 2, this result conforted the Weinberg-Salam model.

2. Are neutrinos massless or massive ?

It was realized from the beginning that neutrinos had to be light particles. ¿From the difference observed in beta decays between the electron and the missing (that is neutrino) mean energies, F. Perrin suggested that mean momenta were probably equal, and that implied a neutrino mass much lighter than the electron mass. In a 1936 review, Bethe and Bacher wrote that "the neutrino mass ... was probably zero".

The standard model was built with the explicit hypothesis of massless neutrinos; the grand unified theories based on SU(5) made the same hypothesis, since there was no room for a ν_R in the 15-plets, while SO(10) unified theories could accomodate massive neutrinos in the 16-plets.

But the solar neutrino problem (in 1968, see below), was the main reason for reconsidering the massless feature of the neutrino. Flavor oscillations, by analogy to the $K - \bar{K}$ oscillations, was a possible explanation for the solar ν_e deficit, but this implied massive neutrinos. These oscillations were actively searched for, specially near nuclear reactors, after Reines claimed in 1980 a positive signal from a CC/NC anomaly in neutrino interactions on deuterium [20], which was later refuted.

A non-zero mass for the neutrinos has a strong impact for cosmology, as these particles could then explain the dark matter in the Universe. For some time, the best models for the apparition of large scale structures preferred a mixture of cold dark matter (weakly interacting heavy particles, or WIMPs) and hot dark matter (for which neutrinos with few eV masses were perfect candidates). This is no longer true after a non zero cosmological constant has been introduced in these models. But this triggered experiments in the 90's searching for ν_μ - ν_τ oscillations in the few eV range, such as NOMAD [21] and CHORUS [22] at CERN.

The see-saw mechanism, proposed in 1979 [23], brought a natural explanation for light neutrinos in grand unified theories. More recently, the CP violation induced by a complex neutrino mixing matrix is considered as the best candidate to explain matter-antimatter asymmetry in our Universe, through leptogenesis. All these theoretical arguments are developped in other lectures in this school. They explain how neutrino properties can shed light on physics at the grand unification energy scale, and this is why the determination of neutrino properties is considered nowadays of fundamental importance.

2.1 Direct determinations of the mass

Soon after the neutrino was postulated by Pauli, the neutron was discovered and in 1934, Fermi proposed his theory of beta decay where a neutron decays into a proton, an electron and a neutrino. This effective theory was successful in explaining the electron spectrum, rates were determined by the Fermi constant and a phase space factor. In his original publication [24], Fermi noticed that the shape of the electron spectrum near the end point depends crucially upon the neutrino mass, and offers a possibility of determining it, or at least of putting an upper limit. The experimental challenge is however enormous, since one needs a very strong source to study the tiny part at the end of the spectrum with sufficient statistics. Moreover, the detector energy response (measured electron energy versus true energy at emission) has to be perfectly understood, since it creates spectral distorsions which, if not properly accounted for, would translate into a fake m_ν^2, either positive or negative, when fitting the spectrum. As a matter of fact, many experiments each published in the 80's and 90's neutrino masses compatible with zero, but adding up all results gave a m_ν^2 negative at 2 sigmas ! This was due to improper accounting of some experimental effects. To list but a few, electron energy losses by ionization in the source, backscattering of electrons on walls, energy resolution, and so on.

The most precise results come from two experiments, Troitsk [25] and Mainz [26], which both use a tritium source (long-lived isotope with a very low energy end point around 18 keV) for which molecular and atomic effects are easy to compute with precision, and a so-called solenoid retarding spectrometer (see figure 3a) which uses both an adiabatic guiding magnetic field for bigger acceptance of electrons and an electrostatic barrier to let only the highest energy emitted electrons to get to the detecting device. This detector measures the rate of electrons above the barrier energy, and the spectrum is deduced by varying the electrostatic barrier potential (so called integral measurements, as opposed to differential measurements using sophisticated topologies of magnetic fields aimed at measuring the electron momentum, and which were used for some time with less success).

Both experiments encountered many problems before being able to reach a good sensitivity on the neutrino mass. For example, Mainz got for some time neutrino mass values depending upon the energy range below end point used for fitting, with a tendancy to get negative square masses: it was discovered that the effect was due to a roughening transition of the tritium source, made of solid tritium deposited on a substrate; the thickness of the source was originally well defined, but later became

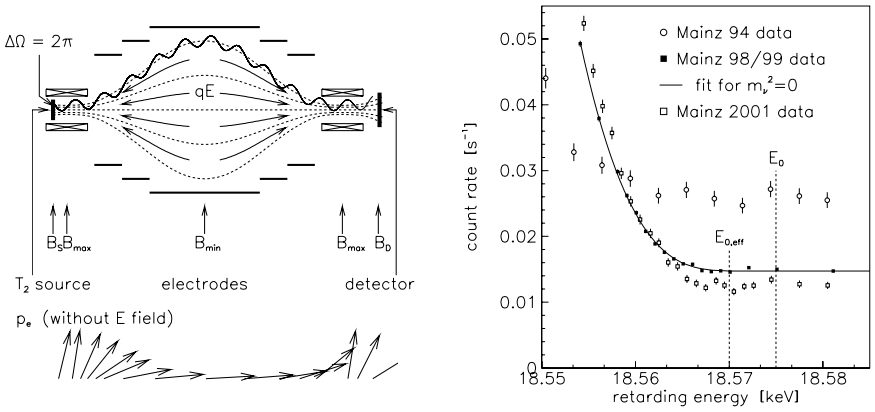

Figure 3. a) left : schematics of the Mainz solenoid retarding spectrometer, which adiabatically suppresses the transverse electron momentum for maximal acceptance b) right : Mainz beta spectrum of tritium, before and after detector improvements (lower temperature of the source, spectrometer modifications). Notice the improvement in background level, and the good fit to $m_\nu = 0$

chaotic, leading to an energy loss of the electron in the source varying with time and location; this was finally corrected by lowering the temperature of the source below 2 K, freezing out this unwanted transition. It took also several years to lower the background of electrons to a sufficiently low level (see figure 3b).

Troitsk using a gaseous molecular source of Tritium did not encounter this problem, but unfortunately found in its data a monochromatic spurious energy line just below the end point, with intensity and location fluctuating with time (with some sinusoïdal pattern which gave birth to lot of speculations on a possible cloud of relic neutrinos around our Sun). This line was not seen by Mainz when both experiments took data at the same time, and the consensus is now that the Troitsk line is an instrumental artefact of unknown origin. When Troitsk removes this line before fitting its spectrum, they get a result totally compatible with Mainz: both experiments announce a neutrino mass below 2.2 eV at 95 % CL. The sensitivity of both experiments is limited by the stiffness of the acceptance for electrons, varying typically from 0 to 1 over a few eV around the value of the barrier potential. To improve this stiffness, directly related to the energy resolution, the only solution is to build a bigger spectrometer : this has led to the KATRIN project [27], proposed by a collaboration containing Troitsk and Mainz experts, and which aims at a neutrino mass sensitivity of 0.3 - 0.4 eV and is due to start in 2007. The proposed spectrometer is a scaling-up of the present

ones, with a 20 m length and 7 m maximal diameter, and is designed to use both gaseous and frozen tritium sources. It will run at Karlsruhe near a facility producing the needed tritium.

2.2 Electromagnetic properties

Another way of proving that neutrinos are massive is to exhibit their electromagnetic properties. For example, a magnetic moment would prove their massive character. In the minimal standard model with a neutrino mass put by hand, the predicted magnetic moment is computable and found desperately low, but higher magnetic moments could be possible in case of new physics induced for example by grand unified scenarios. Since the early days, the limit on this magnetic moment has improved, and now reaches $10^{-10}\,\mu_B$.

In order to determine the neutrino magnetic moment, one studies the diffusion of neutrinos on electrons at low energy. If the neutrino has a magnetic moment, the cross section contains 2 terms:

$$\sigma = \sigma_W + \sigma_{em}$$

the weak cross section σ_W is the usual one.
the electromagnetic cross section σ_{em}, which does not interfere with σ_W since it contributes to different helicity amplitudes, shows a pole in $1/E_e$ where E_e is the energy of the final electron:

$$d\sigma/dy = (\mu_\nu/\mu_B)^2 \pi\alpha^2/m_e^2[1/y - 1]$$

where $y = E_e/E_\nu$.
Thus, if one integrates on the electron energy from a threshold energy E_{thr} up to the kinematic limit, σ_W grows like E_ν while σ_{em} grows like $\log(E_\nu/E_{thr})$. This means that one should use as low a threshold as possible to get high sensitivity on the electromagnetic part, and stay at low energy so that the weak cross section does not overwhelm the electromagnetic cross-section. In order to keep the background low, one should avoid hydrogenated targets so that the CC cross-section on proton is absent. Best experiments as of today have been run at nuclear reactors (as neutrino energies are only a few MeV).

The best result comes from the MUNU experiment at Bugey reactor [28], where the target was the freon gas inside a 1 m^3 TPC surrounded by an active shielding of liquid scintillator. Recoil electrons give sizeable tracks and are easily recognized compared to heavy particles (recoil nuclei induced by fast neutrons or alphas from radioactivity) giving highly ionizing and very short tracks. A difference of electron spectra, starting at 900 keV, between reactor ON and reactor OFF periods gives the

neutrino signal, which is found compatible with σ_W and translates into a limit of $10^{-10}\mu_B$ on the electronic neutrino magnetic moment [1].

A new idea, recently proposed, would use a spherical TPC filled with a noble gas (for example Helium) [29]. The drift field would be radial, and the reading would be done at the center of the sphere using a Micromegas detector. Sub-keV thresholds are technically possible, and if the background is manageable at such low energies (where it was never measured), then the sensitivity on the neutrino magnetic moment could be improved by 1 to 2 orders of magnitude near a nuclear reactor. R & D is presently performed on such a detector.

3. The phenomenon of flavor oscillations

Attempts at a direct proof of neutrino masses have been up to now unsuccessful. But there is an indirect way to prove their massive character, which consists in looking for flavor oscillations. This phenomenom is predicted by standard quantum mechanics, and based on the fact that if neutrinos are massive, the 3 flavor eigenstates (ν_e, ν_μ, ν_τ) need not coïncide with the 3 mass eigenstates (ν_1, ν_2, ν_3). We have then 2 distinct bases connected through a unitary 3x3 matrix. The Schrödinger equation describing the free propagation of a neutrino predicts the appearance of different flavors with time. We will first see the 2 flavor case, technically simpler, and postpone the 3 flavor case, which is more intricate.

3.1 2-flavor formalism

Let us restrict to a world where only 2 flavors (ν_e, ν_μ) are present so that we have 2 mass eigenstates (ν_1, ν_2) with masses m_1 and m_2. The unitary matrix linking the 2 bases is just a rotation by an angle θ.

$$|\nu_e\rangle = \cos\theta\ |\nu_1\rangle + \sin\theta\ |\nu_2\rangle$$

$$|\nu_\mu\rangle = -\sin\theta\ |\nu_1\rangle + \cos\theta\ |\nu_2\rangle$$

Let us consider a ν_e produced with momentum p at t=0.
After a time t, it will be :

$$|\nu(t)\rangle = \cos\theta\ e^{-iE_1 t}\ |\nu_1\rangle + \sin\theta\ e^{-iE_2 t}\ |\nu_2\rangle$$

with

[1] only Dirac neutrinos may have intrinsic magnetic moments. But both Dirac and Majorana neutrinos may have transition magnetic moments, when the initial and final neutrinos have different flavors

$$E_i = \sqrt{p^2 + m_i^2}$$

The probability to interact as a ν_μ at time t is given by:

$$|\langle \nu_\mu | \nu(t) \rangle|^2 = 4 \sin^2\theta \cos^2\theta \sin^2 \frac{(E_1 - E_2)\, t}{2}$$

If the neutrino is relativistic (which is always the case) then $E_1 - E_2 = \frac{m_1^2 - m_2^2}{2p}$ and we can write:

$$P(\nu_e \to \nu_\mu,\ t) = \sin^2 2\theta \sin^2 \frac{\Delta m^2}{4p}\ t$$

As can be seen on figure 4a, the ν_μ component and the ν_e component (which add up to 1) oscillate sinusoïdally with time with a period $T = \frac{4\pi\ p}{\Delta m^2}$, corresponding to an oscillation length

$$L_{\rm osc} = cT = 2.5 \text{ meters} \times \frac{E_\nu(\text{MeV})}{\Delta m^2(\text{eV}^2)}$$

the maximal amplitude of the oscillation being given by $\sin^2 2\theta$.

Actually, the correct way to derive this formula implies to describe localized neutrinos with wave packets, but the result is the same [30]. As a bonus, one finds however that the oscillation pattern fades away after p/σ_p oscillations, where σ_p is the width in momentum of the wave packet. In practice, the pattern is not experimentally observable after p/σ_{exp} oscillations, where σ_{exp} is the experimental resolution on the neutrino momentum, or the natural width of the source if this momentum is not measured. After this damping has occured, the transition probability becomes constant at $0.5 \sin^2 2\theta$.

3.2 Oscillation experiments, exclusion plots

As we have seen, there are two possible ways to look for an oscillation. Flavors are observable through charge currents on nuclei, the produced charge lepton identifying the flavor of the interacting neutrino. Either one looks for a deficit in the initial flavor (disappearance experiment) or for the appearance of a flavor initially absent (appearance experiment). For small oscillation amplitudes, appearance is certainly better since in case of no background and a pure flavor source (for example ν_e), a single interaction producing a muon will prove the oscillation, while for disappearance the sensitivity to the oscillation amplitude, $\sin^2 2\theta$, is limited by statistical fluctuations on the number of ν_e interactions.

But appearance is not always possible. In the previous example, if the neutrino energy is below the muon production threshold, it will not

be possible to sign ν_μ appearance and we have to limit ourselves to disappearance experiments : this will be the case for low energy ν_e like those from the Sun or from nuclear reactors.

The principle of oscillation searches is to use a detector far from the source. But this distance is to be compared to the oscillation length, which goes like the inverse of Δm^2. Any experiment will then be sensitive to Δm^2 values above a lower limit defined by the distance of the detector and the energy of the neutrinos. An experimental complication arises when the flux, or the flavor composition of the source (in neutrino beams, mainly ν_μ from π decays, there is always a small ν_e component from μ or K decays) is not perfectly known. In this case, a remedy consists in using two detectors, one near and the other far from the source, and compare observations at the 2 locations. Any difference in flavor content will prove the presence of flavor oscillations. This will however work only if the oscillation length is higher than the near location distance from the source; this means that such comparisons will be blind to oscillations of too high frequency, that is too high Δm^2. The sensitivity on Δm^2 will then be limited from below and from above.

In the design of an experiment looking for neutrino oscillations, one always has to think beforehand to all the possible backgrounds which could mimick the signal. To decrease the background, shieldings have to be used. The best shielding against cosmic rays is to go deep underground. Furthermore, local backgrounds due to radioactivity, and specially gammas and neutrons, impose in the case of low energy neutrinos to design passive and/or active shieldings surrounding the detector. In some cases, the remaining background can be measured when the neutrino source (beam, reactor) is off. But it is not always possible (think of the Sun !).

Finally, when the result is obtained, it is translated into an acceptance (in case of a positive result) or into an exclusion (in case of negative result) domain in the plane of the 2 physical parameters, $\sin^2 2\theta$ and Δm^2 (see figure 4b).

One should however keep in mind that there are 3 neutrino flavors in nature, so that actual oscillations are governed by more than 2 parameters. Fortunately, nature has been kind enough so that these 2-flavor parameters are easily reinterpreted in the 3-flavor case.

3.3 An example : the Chooz experiment

As an example of oscillation searches, I recommend to read the publication of the Chooz experiment [31], which has looked for the disappearance, 1 km away from the source, of nuclear reactor antineutrinos.

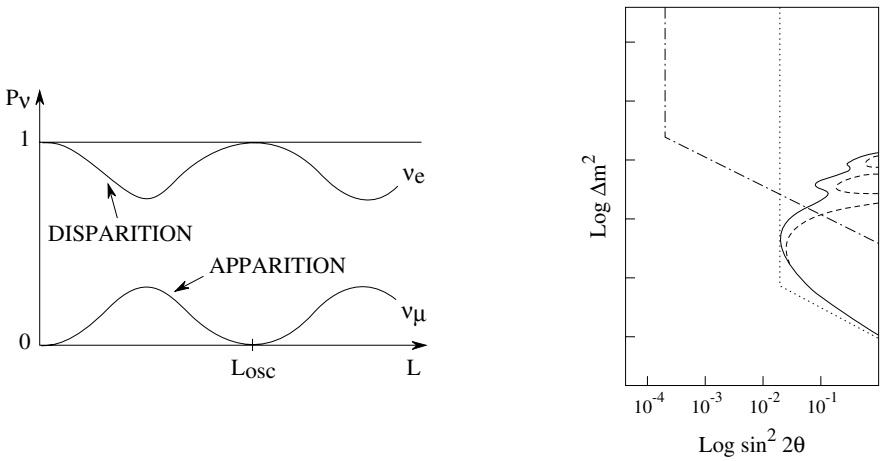

Figure 4. a) Left : oscillation pattern between two neutrino flavors for a given neutrino energy
b) Right : Typical result of an oscillation experiment with a negative result. The dash-dotted curve corresponds to an appearance experiment, the dotted curve to a disappearance experiment, showing the better sensitivity to small oscillations for appearance experiments. The vertical line at high Δm^2 corresponds to oscillation lengths smaller than the distance between source and detector, while the lower line with a slope corresponds to oscillation lengths larger than this distance. The continuous curve shows a typical exclusion domain obtained from the comparison of a near and a far detector, and shows a loss in sensitivity for high Δm^2 when the oscillation length becomes much smaller than the distance between source and near detector

Previous experiments have measured the initial neutrino energy spectrum with precision, so that a single far detector (see figure 5a) was judged sufficient. It was placed 100 meters below ground in order to reduce the cosmic ray induced backgrounds. The active target was 5 tons of gadolinium-loaded liquid scintillator, shielded with unloaded scintillator between the target and the photomultipliers.

As in Reines experiment, thermalized neutrons are identified by their capture on Gadolinium (the capture cross-section is 160 000 barns !) which emits several gammas of 8 MeV total energy (above natural radioactivity energies). A first scintillation signal of at least 1 MeV, due to the positron and its annihilation in the target, followed within 100 μs by a neutron capture, will identify a neutrino interaction. The rate of accidental coincidences can be computed from single rates, and is about 2 candidates per day; true coincidences due to backgrounds (so-called correlated events), coming mainly from fast neutrons entering the detection volume, was kept at about 1 candidate per day. These accidental and correlated backgrounds could actually be measured while the

2 nuclear cores were off. The true neutrino signal was then obtained by substracting this background from measurements done during reactor on periods. It was found around 25 events per day, consistent with the expectation in case of no oscillation. Figure 5b shows data during ON and OFF periods, and figure 5c shows the deduced neutrino spectrum together with the Monte Carlo expectation. The agreement is perfect both for rate and for energy shape. The observed flux compared to the expectation with no oscillation gives a ratio :

$$R = 1.01 \pm 2.8\%(stat) \pm 2.7\%(syst)$$

The systematic error is dominated by the uncertainty on the neutrino flux and the target fiducial mass. An exclusion domain for oscillation parameters was deduced from this absence of signal (see figure 5d), which is relevant for the interpretation of the atmospheric neutrino anomaly (see below).

4. The solar neutrinos

4.1 Neutrinos from the Sun

The Sun gets its energy by burning its hydrogen into helium. In this process, 2 protons are transformed into 2 neutrons and 27 MeV are released. This energy is radiated away as photons, typically that of a black body radiation at 6000 K, its surface temperature. In each fusion process, 2 electronic neutrinos are emitted. The majority of these neutrinos come from the fusion of 2 protons into deuterium, and their maximal energy is 420 keV. But there are many possible branches leading to helium, some of which emit neutrinos of higher energy (see figure 6a). The neutrino spectrum is actually dominated by 3 contributions

- the pp neutrinos, ν_{pp} mentionned above
- the Be neutrinos, ν_{Be}, which are monoenergetic, corresponding to the electron capture by ^{7}Be (the dominant line is at 860 keV)
- the boron neutrinos, ν_B, corresponding to the beta decay of ^{8}B into ^{8}Be, with a maximal energy of 14 MeV.

The flux of pp neutrinos can be deduced from the Sun luminosity with few percent precision. The other components are more sensitive to the details of the Sun modelization, and are known with a precision around 10% for ν_{Be} and 20 to 30 % for ν_B. Other cycles (such as CNO) are believed to play a minor role in neutrino production.

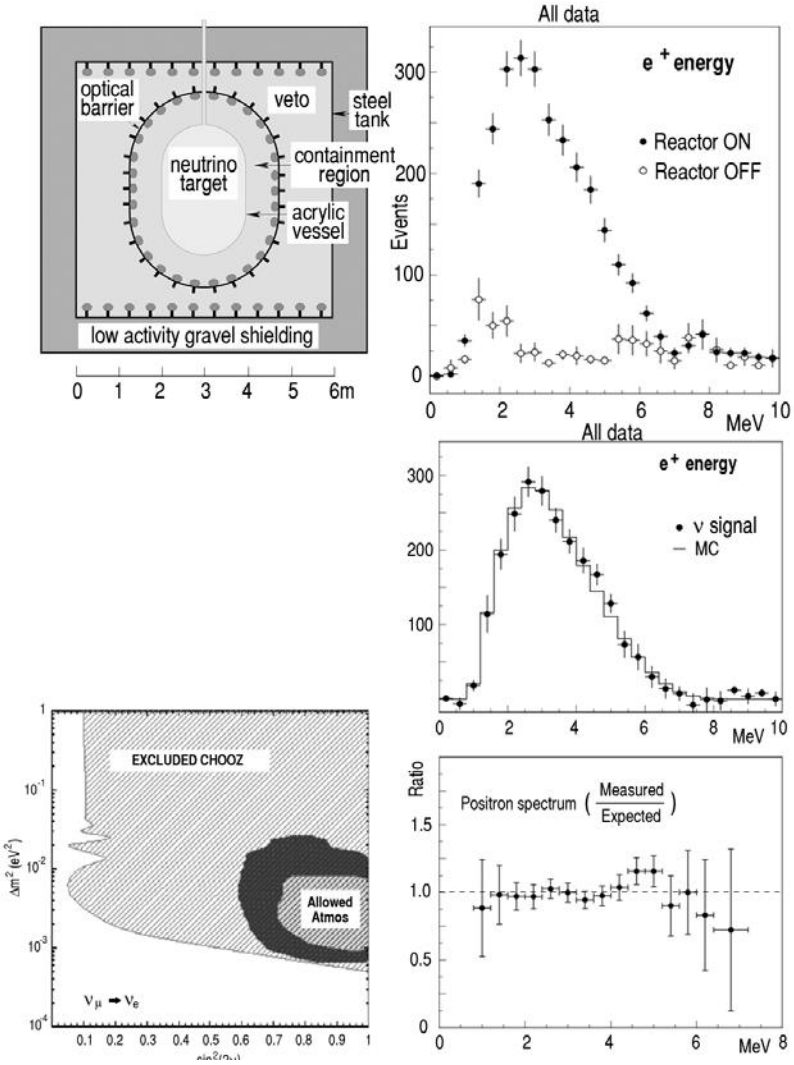

Figure 5. a) upper left : schematics of the Chooz detector
b) upper right : positron spectra measured during reactor ON and OFF periods
c) lower right : net neutrino signal and its ratio to Monte Carlo prediction in the absence of oscillations
d) lower left : Chooz exclusion contour, containing the domain obtained from a fit [33] of Superkamioka atmospheric data in the $\nu_\mu \leftrightarrow \nu_e$ hypothesis

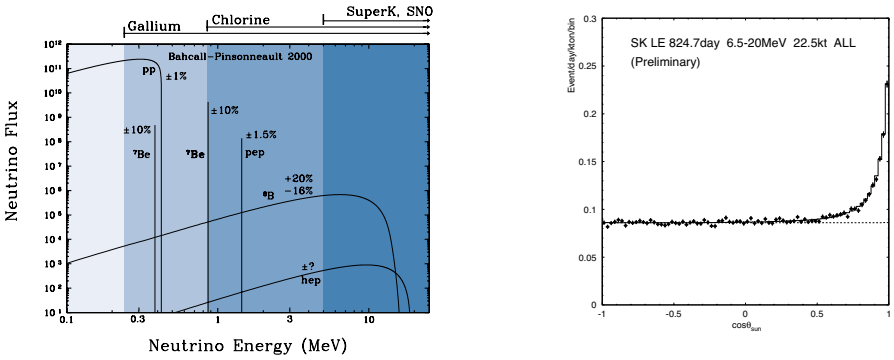

Figure 6. a) Left : Energy spectrum of neutrinos produced in the Sun (the kinematic thresholds for CC interactions on Ga and Cl, and the experimental threshold of Superkamioka and SNO are shown)
b) Right : Angular distribution with respect to the Sun direction of single electron ring events of energy between 5 and 20 MeV. Notice the forward peak proving the detection of solar ν_e's above an isotropic background

4.2 The pioneer experiment at Homestake

R. Davis started in 1968 the first experiment [32] to detect solar neutrinos. It consisted in a big tank located deep underground in the Homestake mine and filled with 390 m^3 of $C_2\,Cl_4$ (liquid used for cleaning clothes). Solar neutrinos of energy above 814 keV should transform ^{37}Cl into ^{37}Ar, a radioactive isotope of argon with a half-lifetime of 35 days. It was a radiochemical experiment, since the target is totally passive, and every 45 days the tank was flushed during 1 day with helium to extract the few radioactive atoms of Argon, while less than 1 cm^3 of ^{36}Ar had been added beforehand into the tank in order to extract a macroscopic volume of argon (it also provides a check of the extraction efficiency). The extracted argon is separated from other gases by using charcoal traps and gas chromatography, and then stored inside a small proportional counter; 2.8 keV Auger electrons coming from the decays of ^{37}Ar by electronic capture are searched for for several months.

After less than one year, it became evident that a signal corresponding to the production of 0.5 ^{37}Ar per day was seen, but the prediction was three times larger !

After more than 20 years of running, the final result gives an observed neutrino flux of 2.5 ± 0.25 SNU [2] while modern solar models predict 7.5 ± 1 SNU. This gave birth to the so-called solar neutrino problem.

4.3 Gallium experiments

The threshold for production of ^{37}Ar is 814 keV, so that Davis' experiment is not sensitive to the most abundant ν_{pp}. This is why, at the beginning of the 90's, two similar experiments, GALLEX [34] and SAGE [35], were launched respectively at Gran Sasso underground laboratory in Italy and in Baksan underground laboratory in the USSR. Both are radiochemical experiments and use Gallium as neutrino target, because the threshold for conversion of ^{71}Ga into ^{71}Ge is only 233 keV, well below the ν_{pp} maximal energy of 420 keV, so that the observed signal should be dominated by these neutrinos. The principle of operation is exactly the same as in Davis' experiment. SAGE uses 55 tons of liquid metal of Gallium in several containers, while GALLEX uses a $GaCl_3$ solution with 30 tons of gallium in a single tank. After about 1 month, GALLEX flushes its tank with nitrogen, taking out the $GeCl_4$ which is gazeous, and after chemical treatment, fills small proportional counters with GeH_4, a suitable gas for detection of X-rays and/or Auger electrons emitted by the electron capture of ^{71}Ge, which has a half-lifetime of 11 days. After 6 years (corresponding to 65 extractions), GALLEX published an observed neutrino flux of 71 ± 6 SNU, while solar models predict between 125 and 130 SNU. The deficit was confirmed, even when the main ν_{pp} component is detected. SAGE, which uses different extraction procedures, found after 10 years the same result (70.9 ± 6.5 SNU). Two very important checks were performed in Gallex :

- An artificial source of ^{51}Cr, which emits neutrinos of 870 keV, was manufactured using several kg of ^{50}Cr irradiated with a high flux of neutrons in a research reactor, and the achieved activity was 60 PBq. Brought to Gran Sasso (its half-lifetime is 25 days), this source, playing the role of an artificial sun with a "luminosity" 10 times higher than our Sun, was placed in the middle of the tank and several extractions were done within a few weeks : the source induced signal was found (within a 10% uncertainty), totally consistent with expectations. This exercise was repeated twice during the experiment [36].

[2] The solar neutrino unit (SNU) corresponds to 1 interaction per second on 10^{36} atoms, and was devised for this experiment so that numbers get close to unity.

- At the end of the experiment, ^{71}As was introduced into the tank. It decays into ^{71}Ge, and the extraction of this germanium was found consistent with expectations at the 1% level [37].

These two tests have proven that ions of ^{71}Ge produced in microscopical quantities inside the tank are extracted as efficiently as macroscopic quantities of neutral atoms added in the liquid: the hypothesis of exotic hot chemistry for these ions which would prevent their efficient extraction, and put forward to explain Davis' result, was definitely ruled out.

4.4 Real-time detection: the Superkamioka experiment

A third result on the solar neutrinos came from an experiment in Japan, Superkamiokande, using a large cylindrical tank (40 m diameter, 40 m high) filled with 50 kilotons of water (preliminary results had been obtained with a smaller detector, Kamiokande, containing 1 kton of water). The walls of this cylinder are covered (coverage of 40%) with large photomultipliers detecting the Cerenkov light induced by charged particles propagating in water. At energies below 100 MeV, only electrons (or converted gammas) can give such a signal. The Cerenkov light cone gives both the energy and the direction of the particle. After several years, the purification of water was brought to such a high level that an experimental threshold at 5 MeV could be achieved. It was then possible to observe the electrons diffused by the higher energy solar neutrinos, that is the ν_B. By kinematics, such electrons point away from the Sun within a few degrees, and can be discriminated against isotropic backgrounds (see figure 6b). The observed neutrino flux [38] (the latest result is
2.35 ± 0.02 (stat) $\pm$ 0.08 (syst) 10^6 cm^{-2} s^{-1}) shows a strong deficit compared to the solar model prediction of 5 ± 1 in the same units. Thus, the first real-time detection of solar neutrinos confirmed the deficit observed by radiochemical experiments. This new type of detection gives more than a total flux: the energy spectrum, its possible distorsion and/or its potential variations with time, can equally be studied.

4.5 Are solar models reliable ?

Starting with the pioneer work of J. Bahcall [39], solar models have been developped for more than 30 years [40]. They consist in describing the Sun evolution from its formation to its present state, based on some simplifying hypotheses (spherical symmetry, no rotation, no magnetic field), and evolution equations (energy transport by radiation in the in-

ner part, by convection in the outer part, thermal equilibrium between produced and emitted energy, hydrostatic equilibrium between radiative pressure and gravity). The actual evolution depends upon physical inputs such as nuclear cross-sections and their corrections (screening), and opacities depending upon the chemical composition and the temperature for energy transport. An iterative procedure has to converge towards a present Sun compatible with observations (mass, radius, surface temperature and luminosity). A recent very important input concerns heliosismology data, as measured by GOLF [41] on the satellite SOHO, which give access to the sound speed along the radius of the Sun (down to 0.1 solar radius). Model builders have converged towards a somewhat standard solar model, and discrepancies are now very small. In particular, such models reproduce the sound speed at the 1 per thousand level (except in the transition zone between radiative and convective transport near 0.6 solar radius, where some turbulence is supposed to occur). With such a maturity, the precision on expected neutrino fluxes has reached 1 % on ν_{pp}, 10 % on ν_{Be} and 20 % on ν_B. More recently, there are some attempts at determining possible effects of magnetic fields inside the Sun on neutrino transport, in case they have a magnetic moment near the experimental limit [42].

After the results of Homestake, Gallex and Sage, and Superkamiokande, some attempts were done to reconcile observations with solar models, by changing drastically some parameters (such as opacities). These ad-hoc models were not very successful, and we now know that they fail totally to reproduce (by several %) the heliosismological constraints. It was actually shown by Hata and Langacker [76] that the 3 results of Chlorine, Gallium, and of diffusion on electrons, together with the Sun luminosity, were enough to extract in a "solar model independant way" the 3 dominant fluxes of neutrinos, ν_{pp}, ν_{Be} and ν_B, using only the known shape of their energy spectrum. The surprising output is that one then obtains a significantly **negative** ν_{Be} flux, and this stays true even if one discards any one of the 3 measurements. There was no hope to reconcile solar neutrino measurements on Earth with solar models, unless one accepts new neutrino properties.

4.6 Oscillations and matter effects

The straightforward explanation, as already suggested by several authors [44], was to suppose that neutrinos are massive and subjected to oscillations [3], so that they arrive on Earth with different flavors:

[3] the possibility of decay for an unstable massive neutrino was also proposed [45].

Chlorine and Gallium experiments measure only the ν_e flux, while Superkamiokande measures a pseudo ν_e flux equal to $\nu_e + 0.15\ (\nu_\mu + \nu_\tau)$, due to the smaller cross-section on electron of the 2 latter flavors.

But solar neutrinos begin their travel to the Earth inside very dense matter, and it was realized by Wolfenstein in 1978 [46] that the presence of electrons would modify the oscillation pattern of ν_e compared to what happens in vacuum. Although neutrinos have negligible interactions with matter, these interactions will however generate an index of refraction, linked to the elastic amplitude in the forward direction. All flavors have the same amplitude on nuclei, but not on electrons (see figure 1). ν_μ and ν_τ will be subjected to the same refractive index, but this index will be different for ν_e. The effect of this refractive index, which acts as a potential to be added to the vacuum Hamiltonian , has to enter the Schrödinger equation. This potential is diagonal in flavor basis, while the free Hamiltonian is diagonal in the mass basis. Adding both will define propagation eigenstates which are different from any one of these 2 bases, and are labelled ν_{1m}, ν_{2m} and ν_{3m}. This matter basis will be constant in matter of constant electronic density, but will vary with time when the electronic density varies along the path of the neutrino (as it is the case inside the Sun from its center to its surface). We will study in the following the 2 cases of constant and varying electronic densities, restricted to the 2 flavor case.

Constant density. In vacuum, the Hamiltonian is diagonal is the mass basis :

$$\begin{aligned} H_V|\ \nu_1\rangle &= E_1|\ \nu_1\rangle \\ H_V|\ \nu_2\rangle &= E_2|\ \nu_2\rangle \end{aligned}$$

When neutrinos go through matter, a potential has to be added to the vacuum Hamiltonian. This potential is diagonal in the flavor basis :

$$\begin{aligned} V|\ \nu_e\rangle &= \left(C + \sqrt{2}\ G\ \rho_e\right)\ |\nu_e\rangle \\ V|\ \nu_\mu\rangle &= C|\ \nu_\mu\rangle \end{aligned}$$

The term C describes the neutral current interactions on nuclei (or nucleons) and electrons, it is common to all flavors (if only C was present, oscillations would not be modified). The extra term for ν_e, proportional to the Fermi constant G and the electron number density ρ_e corresponds to charge currents of ν_e on electrons. Please note that this extra term changes its sign when going from neutrinos to antineutrinos.

The total Hamiltonian is diagonal in a new basis ν_{1m} and ν_{2m} deduced from the flavor basis by a rotation θ_m given by [4]

$$\tan(2\theta_m) = \frac{(E_2 - E_1)\sin(2\theta)}{(E_2 - E_1)\cos(2\theta) - \sqrt{2}\ G\ \rho_e}$$

When the electron density is constant, the oscillation formula has the same structure as in vacuum, but the mixing angle θ is replaced by θ_m and the oscillation length is multiplied by $\sin(2\theta_m)/\sin(2\theta)$.

One sees immediately that oscillation amplitudes will be enhanced with respect to vacuum for neutrinos and damped for antineutrinos when $E_2 > E_1$, that is $m_2 > m_1$. If $m_2 < m_1$, oscillations will be enhanced for antineutrinos and damped for neutrinos. Notice also that the oscillation length increases with respect to vacuum for the enhanced oscillation and decreases for the damped oscillation. Thus matter effects create an asymmetry between neutrinos and antineutrinos which should not be confused with CP violation; it is just due to the fact that this matter is not CP symmetric. This effect, if detected, gives access to the mass hierarchy between m_1 and m_2.

One can also compute the density of electrons for which the enhanced oscillation becomes maximal; it is

$$\rho_R = \Delta m^2 \cos(2\theta)/2\sqrt{2}GE_\nu$$

Varying density. When neutrinos travel through matter with a varying electron density, the propagation equations cannot be solved analytically in the general case, so that one has to resort to numerical simulations. However, there is a special case, called the **adiabatic** case, where the solution is simple. It happens when variations of density are small over one oscillation length, so that evolution equations can be rewritten in the **variable** basis of instantaneous propagation eigenstates while neglecting terms induced by its varying character. This is legitimate when the rotation speed of these eigenstates in the fixed basis of mass or flavor is negligible compared to the oscillation frequency. We actually are familiar with this simplification, when studying how the spin of a particle at rest evolves in space when the particle is submitted to a **slowly varying** (in direction and in strength) magnetic field; one knows that the spin will precess around the magnetic field (Larmor precession) and the axis of this precession will stay aligned with the magnetic field

[4] we need a convention to label m_2 and m_1; here we decide that the m_1 component is the dominant mass component in ν_e, or equivalently that θ is between 0 and $\pi/4$

direction : this is how one rotates the polarisation direction of a polarized target. The analogy is perfect when instead of using the standard orthogonal bases ν_1, ν_2 or ν_e, ν_μ, one uses the so-called Poincaré representation where a neutrino state $\mid \nu >= \cos\theta \mid \nu_e > + e^{-i\phi} \sin\theta \mid \nu_\mu >$ is ascribed a point on a sphere of unit radius with a polar angle 2θ and an azimuth ϕ. One notices that orthogonal states (like ν_e, ν_μ or ν_1, ν_2) will be represented by 2 points opposite on the sphere, so that any orthogonal basis corresponds to a given direction on the sphere. (Furthermore, the probability for a state P to be observed in the state M is just $(1 + \vec{OP}.\vec{OM})/2$, O being the center of the sphere). The equations for neutrino evolution become the same as the evolution of a spin in a magnetic field, with the following correspondances : the field direction corresponds to the direction of the instantaneous neutrino propagation eigenstates, and the strength of the magnetic field (mutliplied by the particle magnetic moment) is replaced by the difference in energy eigenvalues of the two neutrino propagation eigenstates. To summarize, in the adiabatic approximation, a neutrino propagating through matter with slowly varying density will precess (on the sphere) around the axis of instantaneous eigenstates and follow it (see figure 7).

Interpretation of solar data. The matter in the Sun corresponds to the case of varying electron density. The adiabatic approximation will hold when the "Larmor frequency" on the Poincaré sphere is higher than the rotation speed of the direction of the propagation eigenstates. Taking into account the known exponential decrease of electron density with solar radius, this condition will hold when

$$\Delta m^2(eV^2) \sin 2\theta \tan 2\theta \gg 5\,10^{-9} E_\nu(MeV)$$

We produce ν_e near the center of the Sun, and the matter effect will dominate over Δm^2 in the energy splitting if

$$\Delta m^2(eV^2) \cos 2\theta \ll 1.5\,10^{-5} E_\nu(MeV)$$

If this second condition holds, ν_e and ν_μ are the propagation eigenstates at the production point.
When both conditions are fulfilled, the ν_e born as a propagation eigenstate will stay at all times a propagation eigenstate up to its exit from the Sun, so that it will leave the Sun either as a ν_1 (if $m_1 > m_2$) or a ν_2 (if $m_1 < m_2$). The second case is the most interesting (remember that the main mass component in ν_e is ν_1) and is called the MSW effect, after the name of the 2 russian physicists (Mykheyev and Smirnov) who first noticed this effect [47], the W standing for Wolfenstein who had exhibited the importance of matter effects. It corresponds to an adiabatic

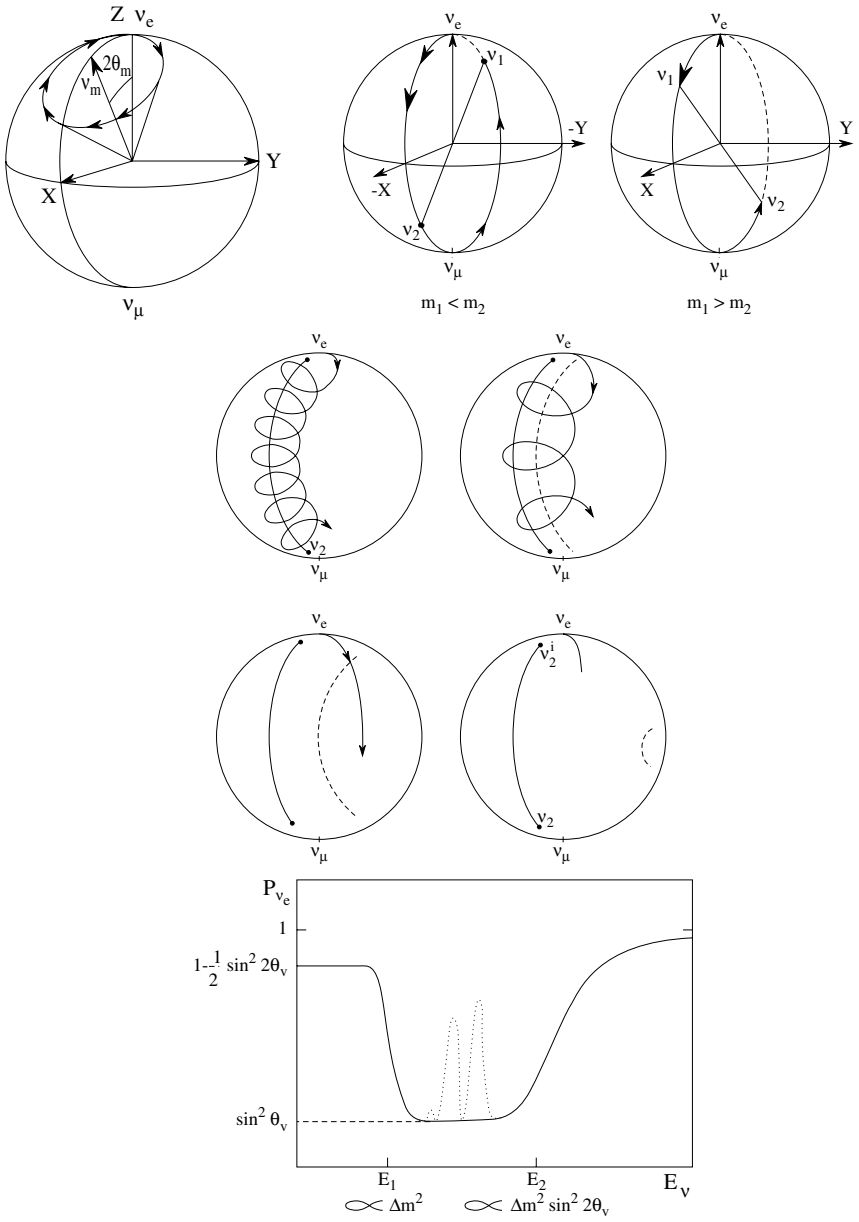

Figure 7. upper left : matter with constant density. The neutrino state precesses around the fixed axis of propagation eigenstates in matter
Upper right : evolution of propagation eigenstates, starting from ν_e at infinite density and ending for null density either as ν_2 if $m_1 < m_2$ or ν_1 if $m_1 > m_2$
Middle : matter with decreasing density (as in Sun) : evolution of an initial ν_e state between Sun center and surface for $m_1 < m_2$ when the adiabaticity condition is less and less satisfied, from top left to bottom right; the driving towards ν_2 becomes less and less efficient.
Bottom : neutrino spectral distorsion when the MSW effect is fully active between E_1 and E_2; below E_1, the central solar density is too small for matter effects to be sizeable, and above E_2, we have a slow loss of the adiabatic condition; the dotted line is for detection during the night when ν_e's are partially regenerated in the Earth

driving in the Sun of a ν_e into a ν_2 (actually no oscillation takes place), which will be predominantly a ν_μ (the smaller θ is, the bigger the ν_μ component will be). The neutrino exiting the Sun as a ν_2 will arrive at Earth as a ν_2, since it is a propagation eigenstate in vacuum. This MSW effect will be effective for a span in neutrino energy given by the 2 above conditions (see figure 7), and can explain naturally as big a ν_e deficit as requested, while vacuum oscillations would be at pain to explain large deficits (that is factors bigger than 3).

When fitting the above-mentionned experiments (Homestake, Gallex and Sage, Superkamioka) [48], several scenarios could explain these data (see figure 8) and were labelled SMA (small vacuum mixing angle, the real archetype of an MSW effect), LMA (large vacuum mixing angle, with a sizeable MSW effect), LOW (with large vacuum mixing angle but with smaller Δm^2 than LMA, and requiring also some MSW effect) and finally the VAC solution where Δm^2 is much smaller, no sizeable matter effects occur and large vacuum oscillations develop between Sun and Earth.

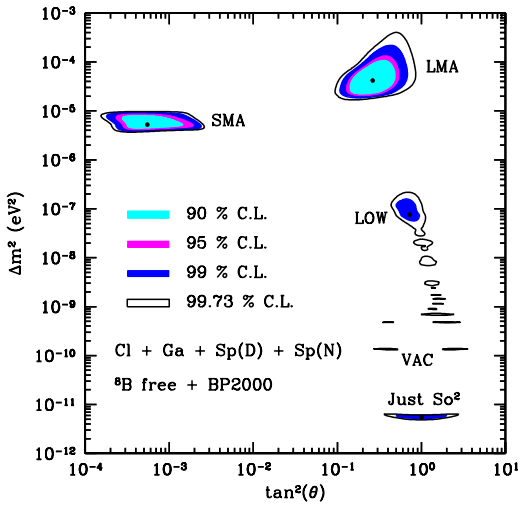

Figure 8. Solar oscillation fits to Homestake, Gallex, Sage, Superkamioka (with day and night energy spectra, to be sensitive to potential regeneration in the Earth). SMA, LMA, LOW are different solutions where the MSW effect is active on part of the spectrum, while VAC and Just-So are vacuum oscillations between Sun and Earth, with no sizeable matter effects. See footnote 5 for an explanation on the abcissa.

In order to discriminate between these scenarios, some observables can help, like diurnal ν_e flux variations (at night, neutrinos travel through the Earth and matter effects - with nearly constant density - take place inside our planet, and might produce some ν_e regeneration), seasonal variations for VAC due to the eccentricity of the Earth orbit, and spectral distorsions. But most of these effects could only be seen in Superkamioka, the only real-time experiment; unfortunately, these effects are bigger at sub-MeV energies and much smaller above the 5 MeV threshold of Superkamiokande, and this prevented any definite discrimination (although with accumulation of data, the LMA solution seemed to be slightly preferred).

4.7 The direct proof: the SNO experiment

The direct proof of flavor change for the solar neutrinos was obtained in 2001 by the SNO experiment [49]. Like Superkamiokande, the detector is a water Cerenkov, but filled with **heavy** water. The detector is spherical and its mass is 1 kiloton. Replacing hydrogen by deuterium has fundamental consequences: apart from their diffusion on electrons, it offers two new ways of detecting solar neutrinos, by their charged current and their neutral current on deuterium. In the first case, deuterium nucleus is transformed into 2 protons, and the emitted electron has a characteristic angular distribution proportional to 1- 1/3 $\cos(\theta)$, where θ is the angle of the electron with respect to the opposite of the Sun direction. This reaction will be triggered only by the electronic flavor. In the second case, the deuterium is simply broken into a proton and a neutron; this is invisible, since nucleons are far below their Cerenkov threshold. The signature comes from the subsequent capture of the neutron by deuterium, giving a tritium and a gamma of 6.25 MeV. This monoenergetic (but energy resolution is poor at this energy) and isotropic gamma is the signature for this reaction, which measures the **total** flux of neutrinos (above the threshold of 2.2 MeV), independant of their flavor.

The first 2 processes are measured using the angular distribution of the electron with respect to the direction of the Sun (more precisely its opposite), described by 3 components: Forward peak for electron scattering, 1- 1/3 $\cos(\theta)$ for charge current, flat for backgrounds.

The third process is obtained from the neutron capture gamma rate.

And the result was magnificent: the 3 measurements gave a total flux 3 times larger than the ν_e flux: for each ν_e arriving on earth, there are also 2 ν_μ or ν_τ. The flavor conversion was proven (see figure 9).

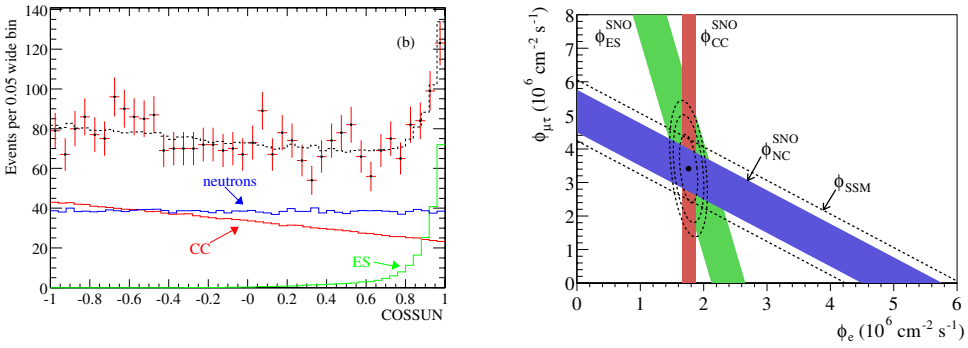

Figure 9. a) Left : electron and gamma angular distribution with respect to the Sun measured by SNO in its second phase (salt added) with the fit of the different components.
b) Right : Flavor composition of the solar neutrino flux as deduced from SNO measurements : ν_e in abcissa, $\nu_\mu + \nu_\tau$ in ordinate.

In a second period, salt was added to the heavy water, in order to obtain a better signature of neutral currents on deuterium, since the capture of neutrons on Chlorine releases 8.6 MeV in gammas, much easier to observe above backgrounds. The first result was confirmed and made more precise [50]. When adding this result to the other experiments, the LMA solution becomes strongly favoured, although the LOW solution survives with very low probability.

4.8 The ultimate proof : the Kamland experiment

But was the oscillation mechanism really proven ? some alternative scenarios, like resonant spin-flavor precession in the Sun in presence of strong magnetic fields, could equally describe all the solar data [51]. One had to observe the oscillation implied by the LMA solution with neutrinos travelling in vacuum, to get rid of all the matter effects complicating the interpretation. This was achieved by another experiment in Japan, Kamland [52]. This experiment did not try to detect the solar neutrinos (they hope to be able to do it in the future), but rather detected the electronic antineutrinos emitted by the japanese nuclear reactors. At the location chosen for Kamland (the Kamioka site), the nearest reactor is at 80 km, and more than 30 reactors are within a distance of less than 200 km. If LMA is right, these neutrinos have time to oscillate on their way to Kamland, and a significant deficit should be observed. The detection technique is standard : it uses 1 kiloton of liquid scintillator

and triggers on the delayed coincidence between a positron and a neutron capture (on protons, since no doping is added to the scintillator). The result, published in december 2002 and shown on figure 10, speaks for itself. 54 events have been seen while 87 were expected. Furthermore, although statistics are low, the observed spectrum is compatible with the expected distorsion due to oscillations. Oscillation fits select an area

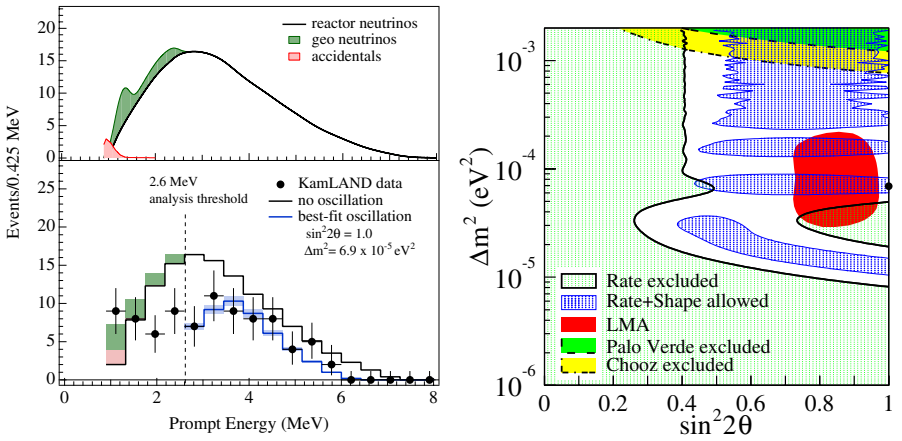

Figure 10. a) Left : the expected $\bar{\nu_e}$ spectrum expected from japanese nuclear reactors in absence of oscillations and below the observed spectrum together with the oscillation fit. The green (or shaded) area corresponds to $\bar{\nu_e}$ emitted by the Earth and creating a poorly known background below 2.6 MeV, so that this value is taken as analysis threshold.
b) Right : Acceptance contour for oscillations from Kamland data (dotted area) together with the LMA domain (in red or shaded) from solar experiments (LMA). There are two allowed bands for the intersection, but since then, the salt data from SNO have killed the higher Δm^2 alternative.

of parameters compatible with (and only with) the LMA solution from previous experiments. When fitting all data, one obtains [53]
$\Delta m^2 = 7.3^{+0.4}_{-0.6}\, 10^{-5} eV^2$ and $\tan^2\theta = 0.41 \pm 0.04$

A maximal oscillation in vacuum, corresponding to $\tan^2\theta = 1$, is excluded at more than 2 sigmas [5].

[5]Notice that here $\tan^2\theta$ is used instead of $\sin^2 2\theta$; this allows to cover in the same plot the two cases $m_1 < m_2$ and $m_1 > m_2$ by redefining θ as the angle between ν_e and the **lighter** of ν_1 and ν_2, so that θ will be smaller than $\pi/4$ when $m_1 < m_2$ and between $\pi/4$ and $\pi/2$ when $m_1 > m_2$. The MSW effect will then occur in the Sun when $\tan^2\theta$ is smaller than 1.

The interpretation of the solar neutrino deficit becomes unambiguous : it is due to oscillations driven by these parameters, with the MSW effect acting on this oscillation. The solar problem is finally solved.

4.9 Future projects

Before SNO and Kamland results, many projects were proposed to measure in real-time the low energy part (ν_{pp} and/or ν_{Be}) of the neutrino spectrum, as the different possible scenarios (SMA, LMA, LOW, VAC) differ most in this low energy region. As of today, only one such experiment, Borexino [54], is due to start soon at the Gran Sasso underground laboratory. It aims at measuring the diffusion on electrons of ν_{Be}, in a spherical tank filled with 300 tons of ultrapure liquid scintillator. The purity requirements are formidable, since the signal is a single electron below 500 keV that any gamma from radioactivity can mimmick. It was proven in a prototype that the 10^{-16} g/g of Uranium, and 10^{-18} g/g of $^{14}C\,/\,^{12}C$ could be achieved.

Other more ambitious projects have become less urgent, and in view of their difficulty, it is not clear if they will ever be done, as their initial aim, selecting **the** solution, has been fulfilled by SNO and Kamland.

5. The atmospheric neutrinos

Large underground detectors were initially built to search for proton decay, as predicted by grand unified theories. Going underground was a must for such a rare process. But after some depth, an irreducible background will dominate, generated by interactions of GeV neutrinos produced by cosmic rays in the atmosphere. It was thus very important to study the flux of such neutrinos. While the proton decay signal was never seen, the study of this background turned out to be extremely interesting.

5.1 The Superkamioka observations

Neutrinos are produced in hadronic showers induced in the high atmosphere by pion decays, followed (if the energy is not too high) by the subsequent decay of the muon. Thus one should observe a ratio of ν_μ over ν_e equal or higher to 2 (here, we do no distinction between neutrinos and antineutrinos). In Superkamiokande, ν_e and ν_μ interacting on nuclei are identified by the topology of the Cerenkov ring of the final lepton. Muons suffer little multiple scattering, so that the edge of the Cerenkov ring is sharply defined. On the contrary, electrons will give an electron shower inside water, with many secondary electrons, and the resulting Cerenkov ring will be fuzzy. Beam tests have shown that is is possible to

separate almost perfectly single ring events produced by electrons and by muons. It is also possible to determine whether or not the muon leaves the detector : in case it leaves, all photomultipliers inside the Cerenkov cone (drawn from the muon starting point) will receive light; for stopping muons, photomultipliers inside the smaller Cerenkov cone drawn in the same direction from the muon stopping point will receive no light; of course, the energy of muons leaving the detector cannot be determined.

Superkamiokande found the surprizing result [55] that the observed ratio ν_μ over ν_e was nearly 1, instead of 2. This ratio depends little on the exact modelization of neutrino production in the atmosphere. Furthermore, owing to the high statistics accumulated over the years by Superkamioka, it was possible to study this ratio as a function of the zenithal angle of the final lepton (which is nearly that of the neutrino), and it was found that the anomaly increased when going from vertical downgoing neutrinos to vertical upgoing neutrinos (see figure 11).

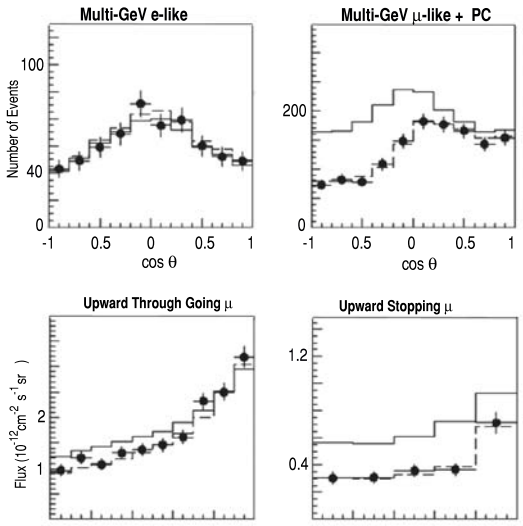

Figure 11. Zenithal distributions of Superkamioka atmospheric data ($\cos\theta = 1$ for vertical downgoing, $\cos\theta = -1$ for vertical upgoing); the left column corresponds to electron-like events, the right column to muon-like events. The first line is for low energy events (E $<$ 1.4 GeV), while the second is for high energy events (E $>$ 1.4 GeV)

As the zenithal angle is directly related to the path length of the neutrino, this suggested strongly a possible flavor oscillation of the neutrinos. Actually, such an hypothesis gives a perfect fit to the data if one supposes an oscillation between ν_μ and ν_τ with a maximal amplitude

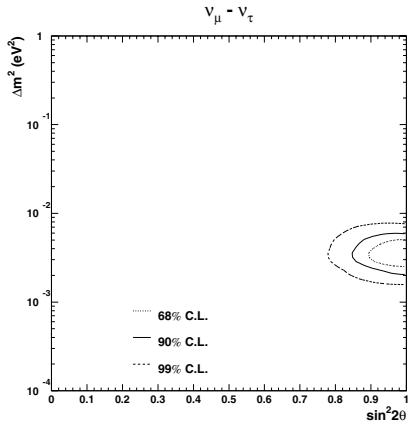

Figure 12. Acceptance domain in $\nu_\mu \leftrightarrow \nu_\tau$ oscillation parameters from Superkamiokande atmospheric data.

($\sin^2 2\theta = 1$) and a frequency given by Δm^2 between 2 and 3 10^{-3} eV^2 (see figure 12). Why is the oscillation supposed to occur between ν_μ and ν_τ, and not between ν_μ and ν_e ? There are several reasons :

- in case of an oscillation between ν_μ and ν_e, the anomaly is decreased due to the partial compensation of flavor conversions (if ν_μ and ν_e fluxes were equal in flux and energy spectrum, there would be perfect compensation and no anomaly would be seen), and the observed anomaly is too important.

- the observed ν_e flux is consistent with Monte Carlo simulations of cosmic ray hadronic showers in the atmosphere, while there is a deficit of observed ν_μ.

- The CHOOZ experiment excludes any sizeable ν_e oscillation at the frequency needed to explain the atmospheric anomaly.

Superkamioka has tried to prove that ν_τ are actually produced by the oscillation mechanism. At typical energies around 1 GeV, ν_τ cannot produce charged current interactions, as the τ is too heavy. But they can produce neutral currents. The idea was then to select events with 2 electron rings compatible with the decay of a π^0 produced by neutral current. With this method, a significant excess in the NC rate compared to the CC rate has actually been observed. This conforts the ν_μ - ν_τ hypothesis and also puts strong constraints on an alternative hypothesis, in which ν_μ would oscillate into a sterile neutrino which has no interactions with matter.

It would seem rather strange to suppose the existence of a fourth family of sterile neutrinos for which there is no experimental evidence. But this hypothesis has been proposed to explain the results of an experiment at Los Alamos, LSND [56], which claimed evidence for $\bar{\nu_\mu}$ - $\bar{\nu_e}$ oscillations. This result, although marginally consistent with a similar experiment, KARMEN [57], which found no signal, cannot presently be disproved. The need for a fourth neutrino comes from the fact that the required oscillation frequency for LSND corresponds to a Δm^2 above 0.1 eV^2, inconsistent in the 3-neutrino case with the 2 values of Δm^2 obtained for solar and atmospheric neutrinos, since the biggest Δm^2 value should be the sum of the 2 others. It should be added that models with 4 neutrinos give poor fits to the data on solar neutrinos, atmospheric neutrinos and LSND. An experiment called MiniBooNE [58] is presently running at FermiLab to check the LSND result, and its result is (impatiently) awaited for 2005.

5.2 Checking oscillation with accelerators

Although the evidence for atmospheric neutrinos is now very strong, it was not so when first anomalies were reported by experiments preceeding Superkamiokande (like Kamiokande, IMB, Soudan, Frejus) and the community, in order to clarify the situation (there were debates over the way cosmic showers were simulated, experiments like Frejus saw no effect,...) decided to check this result using ν_μ beams from accelerators. At the time, the preferred value of Δm^2 was rather around 10^{-2} eV^2, and this is with this value that accelerator projects were designed. But, as time went on, and with the much higher statistics of Superkamiokande, the value of Δm^2 shifted down to 2.5 10^{-3}, so that the projects had to adjust to this much lower value. Lower Δm^2 value means longer oscillation length for a given energy, or a lower neutrino energy if the distance between source and detector cannot be modified.

There have been 3 such projects around the world :

- In Japan, the K2K experiment started in 1999 [59]. It uses a 1.3 GeV neutrino beam made with the 12 GeV protons at KEK and shot towards the Superkamiokande detector, 250 km away. Near detectors, among them a 1 kiloton water Cerenkov, help in determining the characteristics of the beam and predict the expected number of ν_μ CC interactions in Superkamiokande. ¿From data taken between 1999 and 2001, a confirmation of the ν_μ disappearance suggested by the atmospheric anomaly could be obtained, since only 56 events were observed while 80 $\pm$ 6 were expected. The energy spectrum suggests the distorsion implied by the os-

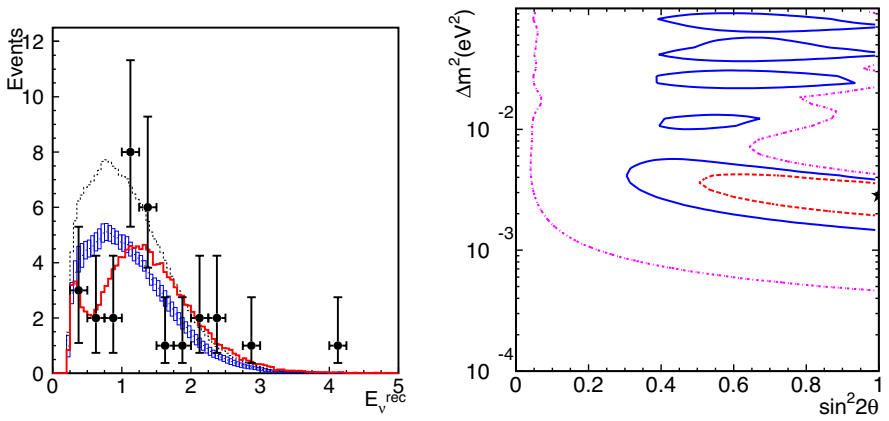

Figure 13. a) Left : The ν_μ energy spectrum measured by K2K (points with error bars); the dotted line is the predicted spectrum in the absence of oscillations, while the solid (red) line is the oscillation best fit
b) Right : Accepted domain for ν_μ disappearance deduced from K2K data (68, 90 and 99 % confidence levels). The star corresponds to the best fit.

cillation (but contains only 29 muon-like single-ring events), see figure 13. This experiment will take data until march 2005, and expects to prove the deficit at the 3 σ level. This experiment is mainly limited by its very low statistics, and illustrates the need for much higher intensity beams.

- In the US, the MINOS experiment [60] is due to start in 2005. It is installed in the Soudan mine, 730 km away from FNAL. The detector has been completed in 2003, and waits for the neutrino beam. It consists of 5 kilotons of magnetized iron equipped with trackers to reconstruct neutrino interactions. The beam was tuned to a lower energy than previously planned, and the neutrino energy will be below 5 GeV. This experiment should clearly see the neutrino deficit and measure the oscillation frequency (that is Δm^2) much better than Superkamiokande.

- In Europe, a high energy neutrino beam produced by the SPS is being built and should be ready in 2006. The choice was made to be complementary to MINOS, namely to explicitely sign the appearance of ν_τ (At MINOS energies, ν_τ cannot give CC interactions). Two detectors are being installed in Gran Sasso underground laboratory, 730 km away from CERN. As the beam mean energy is around 17 GeV, the oscillation is still far from its maximum at

Gran Sasso, but nevertheless each detector expects to see around 10 ν_τ events in 5 years running.

OPERA [61] is a detector using lead and emulsions, and will look for $\tau \to \mu$ decays. If a ν_τ interacts in the lead, the final τ lepton will be either seen in emulsion foils before and after its decay, so that a distinctive kink in the track will sign the τ; or the τ will decay inside the lead before the first emulsion foil and only its decay track (a muon) will be reconstructed : in this case, the signature will be based on the mismatch between this decay track and the vertex determined from the hadronic shower produced by the neutrino interaction. In order to avoid the background from decays of charmed particles, the detector has dipole magnets measuring the sign of the candidate muon : it will be negative for τ decay and positive for charm decay.

ICARUS [62] uses a new technology, issued from many years of R & D, using a liquid argon TPC. In a TPC, the ionization left by charged particles in the medium is drifted by a transverse electric field towards a reading plane, made of wires where electrons shower and give a strong signal (as in proportional chambers). Pads placed behind these wires will observe a signal by influence and locate the 2 coordinates of the shower. The third coordinate is deduced from the drift time (the time it took to the electron to reach the wire). Usual TPC's are filled with gas. Here the medium is liquid, and needs to be extremely pure to avoid attachment (that is capture of the drifting electron by electronegative impurities). Drifts over more than 1 meter without sizeable loss has been achieved in liquid argon. A 600 ton module has been built and ICARUS will use 5 such detectors, reaching a mass of 3 kilotons. Such TPC's give a very precise visualization of the interactions (comparable to the old bubble chambers), but spatial resolution will not be sufficient to explicitely see the short τ track before its decay. The ν_τ CC signal is thus selected by topological and kinematic cuts mainly based on the missing transverse momentum in the final state.

To conclude this part, let us give the results obtained on atmospheric oscillation parameters from 3-family global fits to solar data, Kamland, Superkamiokande atmospheric data and Chooz [63] :

$$\Delta m^2_{atm} = (2.6 \pm 0.4)\, 10^{-3} eV^2 \text{ and } \sin^2(2\theta_{atm}) = 1 \pm 0.05$$

6. Towards a full determination of the mixing matrix

If we stick to 3 flavors of neutrinos, we already know the possible oscillation frequencies. In the following, we will develop the formalism for 3 neutrino mixing, and see how it is possible to determine completely the mixing matrix.

6.1 The mixing for 3 families

The unitary matrix linking mass and flavor eigenstates can be written :

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \qquad (1)$$

This unitary matrix U can be decomposed as the product of 3 rotations, complemented with extra phases responsible for CP violation :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & c23 & s23 \\ 0 & -s23 & c23 \end{pmatrix} \begin{pmatrix} c13 & 0 & e^{i\delta}s13 \\ 0 & 1 & 0 \\ -e^{-i\delta}s13 & 0 & c13 \end{pmatrix} \begin{pmatrix} c12 & s12 & 0 \\ -s12 & c12 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \cdots$$

$$\cdots \times \begin{pmatrix} e^{i\phi_1} & 0 & 0 \\ 0 & e^{i\phi_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $c$ij and $s$ij stand for cosine and sine of θ_{ij}.

The rightmost matrix is only present if neutrinos are Majorana particles, but these phases ϕ_1 and ϕ_2 do not enter oscillation formulae, so they are irrelevant for oscillation experiments. They are however important for other processes such as neutrinoless double beta decays. The 3 other matrices are the quasi standard representation of a rotation in 3-D space with 3 Euler angles, corresponding to successive rotations (from right to left) around ν_3 axis by θ_{12}, then around the transform of ν_2 by θ_{13}, and finally around the transform of ν_1, (that is ν_e) by an angle θ_{23} (see figure 14). One sees however that the θ_{13} rotation matrix is modified by a phase δ which will enter oscillation formulae and induce, if non zero, a CP violation in oscillations.
When switching from neutrinos to antineutrinos, it is enough to change the sign of the CP phases.

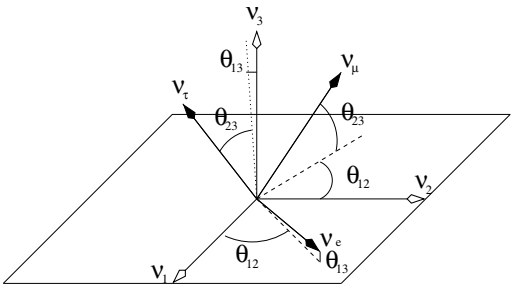

Figure 14. Definition of the 3 angles used in the neutrino mixing matrix

By convention, mass indices 1 and 2 will be used for solar oscillations, while the mass index 3 is used for atmospheric oscillations. This does not imply that m_3 is the highest mass.

From this matrix, one can derive oscillation formulae between an initial flavor l and a final flavor l'; we note Δ_{ij} the oscillating term $(m_i^2 - m_j^2)\, t \,/\, (4\, E_\nu)$:

$$P_{ll'} = \delta_{ll'} - 4Re \sum_{i<j} U^*_{l'i} U_{l'j} U_{li} U^*_{lj} \sin^2 \Delta_{ij} + 2Im \sum_{i<j} U^*_{l'i} U_{l'j} U_{li} U^*_{lj} \sin 2\Delta_{ij}$$

We deduce the formulae for the already observed oscillations, taking into account the fact that the solar frequency is much smaller than the atmospheric one and the smallness of θ_{13} as deduced from CHOOZ:

- For atmospheric oscillations, neglecting Δ_{12} terms and equating Δ_{13} and Δ_{23}:
$$P_{\mu\tau} = 4c_{13}^4 s_{23}^2 c_{23}^2 \sin^2 \Delta_{23}$$
which is exactly the 2-family formula, except for the c_{13}^4 factor which is bigger than 0.92 according to CHOOZ.
The ν_e disappearance relevant for CHOOZ can also be written with the same approximations:
$$P_{ee} = 1 - 4s_{13}^2 c_{13}^2 \sin^2 \Delta_{23}$$
that is exactly the 2-family formula (taking into account the very beginning of the solar oscillation would modify this probability by less than half a percent)

- For solar oscillations in vacuum (relevant for KamLand), we must first take into account the damping of the fast (atmospheric) oscillations and replace $\sin^2 \Delta_{23}$ and $\sin^2 \Delta_{13}$ by 0.5:
$$P_{ee} = (1 - 2s_{13}^2 c_{13}^2) - 4c_{13}^4 s_{12}^2 c_{12}^2 \sin^2 \Delta_{12}$$

which, taking into account the fact that θ_{13} is small can be rewritten :

$$P_{ee} = (1 - 2\theta_{13}^2)(1 - \sin^2 2\theta_{12} \sin^2 \Delta_{12})$$

that is , apart from an overall factor between 0.92 and 1, the same formula as in the 2-family case

These formulae explain why the 2-family fits for the observed oscillations are relevant and have allowed to determine the splittings in squared masses, $\Delta m^2_{12} = 7\,10^{-5} eV^2$ and $\Delta m^2_{23} = 2.5\,10^{-3} eV^2$, and the angles $\theta_{12} \simeq 35^o$ and $\theta_{23} \simeq 45^o$.

θ_{13} is the only unknown mixing angle, while the CP phase δ creates no sizeable effect in the above-mentionned oscillations. The only way to determine these 2 parameters is to look for oscillations involving ν_e at the atmospheric oscillation frequency. The formula for ν_μ to ν_e transition, which in practice will be the searched for oscillation, at distances where the atmospheric oscillation is fully developped while the solar oscillation is just at its very beginning will be more intricate, as on the one hand the terms for atmospheric and solar frequencies will be both small and may well be of the same order of magnitude, and on the other hand the CP violating phase δ will explicitely appear and may create big asymmetries. An additional complication is due to matter effects which will be important for neutrino beams above 1 GeV, since the detectors will then be located at typical distances of 1000 km or more.

General formulae can be found in the litterature [64]. We will here restrict to the vacuum case, and write the formula developped in the 2 small parameters $\epsilon = \theta_{13}$ and $\epsilon' = \Delta_{12}$, while Δ will be a shorthand for Δ_{13} and Δ_{23} (the derivation needs some care as the difference between Δ_{13} and Δ_{23} is ϵ')

$$P_{\mu e} = (2\epsilon s_{23} \sin\Delta)^2 + (2\epsilon' c_{12} s_{12} c_{23})^2 + 2(2\epsilon s_{23} \sin\Delta)(2\epsilon' c_{12} s_{12} c_{23}) \cos(\Delta + \delta)$$

this formula exhibits the positivity of $P_{\mu e}$ which is of the form
$X^2 + Y^2 + 2XY \cos(\phi)$

To best determine θ_{13}, a detector should be placed in a neutrino beam so that the distance roughly corresponds to the first atmospheric oscillation maximum; this optimal distance is approximately 500 km $\times E_\nu$ (GeV). It is then interesting to rewrite the formula when the atmospheric oscillation phase Δ is exactly $\pi/2$, and use the fact that $\theta_{23} = \pi/4$:

$$P_{\mu e}^{max} = A^2 + S^2 + 2AS \sin(\delta)$$

where $A = \sqrt{2}\theta_{13}$ is the "atmospheric" term,
and $S = \Delta_{12} \sin(2\theta_{12})/\sqrt{2}$ is the "solar" term.

One clearly sees how the CP violating term comes from an interference between the 2 oscillation amplitudes (solar and atmospheric). When these two terms are equal (which happens for $\theta_{13} \simeq 1^o$ as $S \simeq 0.03$), a maximal CP violation will totally cancel one of the oscillations (either ν or $\bar{\nu}$), while the other is twice the expected value without CP violation. When A and S are different, the maximal asymmetry becomes smaller. From this it follows that the sensitivity on δ is roughly constant as soon as A is bigger than S (the bigger is A, the higher the statistics but the smaller the asymmetry, and both variations compensate each other).

In order to extract the 2 unknown parameters θ_{13} and δ, it is necessary to do several measurements, for example with ν_μ and $\bar{\nu_\mu}$. A single measurement will be sensitive to a given area in the 2-dimensional plot of these 2 parameters, as shown in fig. 15. By convention, it is often

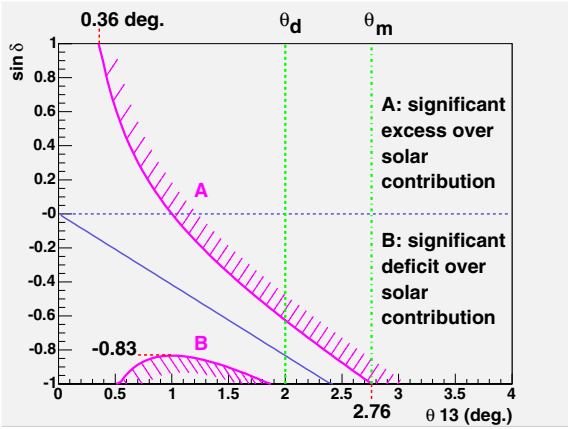

Figure 15. Sensitivity domain in the θ_{13} $\sin\delta_{CP}$ plane of an appearance experiment. A significant effect will be present whatever the value of δ could be for $\theta_{13} > \theta_m$, but one sees that the actual sensitivity (as defined by a significant excess of events, to the right of curve A, or deficit, below curve B), goes well below this value. In contrast, a disappearance experiment (not sensitive to δ) having a sensitivity on θ_{13} above θ_d, even though θ_d is smaller than θ_m, will not cover areas of the plane where the appearance experiment has sensitivity.

quoted a sensitivity on θ_{13} for a single measurement : in this case, δ is fixed at zero by convention (to compare easily different projects); we implicitely use this convention in the following. Some authors [65] prefer to quote the value of θ_{13} above which an oscillation signal will be seen whatever the value of δ; this more conservative value might be labelled as pessimistic, since for some values of δ, a significant oscillation signal might still be seen for θ_{13} values smaller than this limit.

6.2 The ultimate tool : the neutrino factory

In order to be sensitive to the small ν_μ - ν_e oscillation, one needs very intense neutrino beams. For 5 years now, physicists have studied the so-called neutrino factory. This facility was derived from a much more ambitious project, the muon collider, which is an a priori surprising alternative to the electron colliders which would follow LHC. In case LHC finds Higgs particle(s), one would like to study them with high statistics in a clean environment. In the same way as LEP was the perfect machine to study the gauge bosons W and Z, e^+ e^- machines would copiously produce Higgs accompanied only by Z^0. At high energies however, it will be impossible to make electron rings, as the energy loss in the arcs would be tremendous (by synchrotron radiation). One has to resort to linear accelerators of electrons,but in order to avoid lengths of reaching hundreds of kilometers, new acceleration schemes have to be designed, and a lot of R & D is presently going on. For muons however, circular rings are still possible [6]. The challenge is also formidable, as one has to produce a lot of muons, collect and accelerate them before they decay, and also cool them so that beam densities, whose square enters the collider luminosity, is sufficiently high. The R & D has started on such machines, and it was quickly realized that with more modest demands, it would be possible, in a first stage, to produce very intense neutrino beams from the decay of muons stored at a few tens of GeV in a single storage ring. The concept of neutrino factories was born, and the european design for such a machine is shown on fig. 16a.

With neutrino factories, the typical baselines for θ_{13} and δ_{CP} studies would be 1000 to 7000 kilometers. As muon decays produce both ν_μ and $\bar{\nu_e}$, the best strategy is then to use magnetized detectors to observe $\bar{\nu_\mu}$ CC interactions, producing wrong sign muons, that is opposite in charge to those produced by CC interactions of the initial ν_μ; these wrong sign muon events would then prove the oscillation of the initial $\bar{\nu_e}$ into $\bar{\nu_\mu}$. To study CP, it is enough to switch from muons to antimuons in the storage ring. But at so big distances and large neutrino energies (10 GeV or more), matter effects will be very strong and dominate over genuine CP violation induced by δ, as can be see on fig. 16b. This is why more than one distance between source and detector is necessary, and these distances must be carefully chosen to improve the sensitivity to δ. Lots of groups work on this problem, and a somewhat preferred

[6]Furthermore, with muon colliders, the **formation** of Higgs, by tuning the collider energy to its mass, becomes possible, contrary to electron machines due to the tiny coupling of the Higgs to electrons (Higgs couplings to $x\bar{x}$ fermion pairs are proportional to m_x)

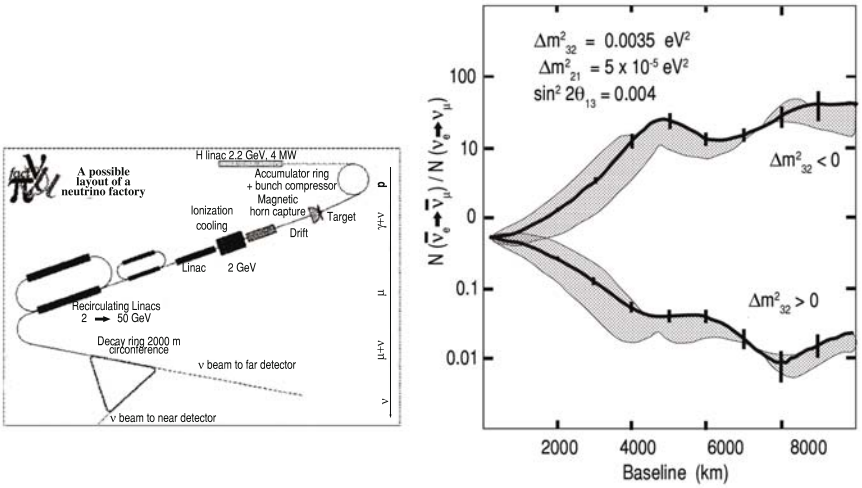

Figure 16. a) left : schematics of the european neutrino factory as envisionned at CERN. The SPL produces 4 MW of protons impinging on a target, muons ring sending neutrino beams towards a far detector.
b) right : Matter and CP violation effects as a function of the distance to the neutrino sources. The two branches correspond to matter effects which allow to determine the mass hierarchy between m_2 and m_3. The thickness of these bands is due to genuine CP violations created by the phase δ_{CP}; the lines in each band correspond to $\delta_{CP} = 0$.

scenario would use 2 detectors located at roughly 3000 and 7000 km from the neutrino source. For details on neutrino factories, I refer to the proceedings of the annual NuFact conference [66].

The fake CP asymmetry due to the matter effects complicates the study of the real CP asymmetry (induced by δ), but it has a very strong advantage, as it will solve the question of mass hierarchy. We have seen that matter effects either enhance or decrease the oscillation vacuum amplitude and length. This will be also true in ν_μ to ν_e oscillation at the atmospheric frequency, and the enhancement will happen for neutrinos or for antineutrinos depending upon what is called the mass hierarchy of the neutrinos :

- If m_3 is higher than m_1 and m_2 (normal hierarchy), matter effects will enhance the ν_e to ν_μ oscillation.

- If on the contrary, m_3 is the lowest mass, while m_1 and m_2 are nearly degenerate (the splitting between m_1^2 and m_2^2 is given by the solar oscillation frequency, and the ordering $m_1 < m_2$ is imposed by the necessity of an MSW effect for solar neutrinos), the

hierarchy is called inverted and ν_e to ν_μ oscillation amplitude will be decreased

Although the sensitivity reach on θ_{13} of neutrino factories is impressive (about 1/10 of a degree), the price of such a machine is such that it will not be financed unless one is sure that a sizeable oscillation signal will be seen so that useful physics is ascertained. This imposes to improve beforehand our sensitivity on θ_{13} with other less expensive facilities; projects using very intense, but conventional, neutrino beams, nicknamed superbeams, will allow to explore θ_{13} values down to 1 degree in the next 10 to 15 years. The construction of a neutrino factory will then depend on the fact that the $\nu_\mu \leftrightarrow \nu_e$ oscillation is discovered or not.

6.3 Present beams and detectors

Although MINOS , OPERA and ICARUS have not been designed to look at ν_e appearance, these detectors should have some sensitivity to θ_{13}. Monte Carlo studies show that the 90% CL sensitivity on θ_{13} will not go below 5 degrees (which however leaves room for discovery !). The CNGS beam is produced by 170 kW of protons, so that the sensitivity is limited mainly by statistics (too few events), keeping in mind that the distance to the source is too short by roughly a factor of 10 to match the oscillation maximum. The FNAL beam uses 400 kW of protons and the distance of MINOS better matches the required distance, but here the main problem will be the π^0 produced by neutral currents and mimicking an electron in the detector.

6.4 A better reactor experiment ?

The present best limit on θ_{13} is given by CHOOZ (see above). Is it possible to improve this result ? The answer is yes : a higher detector mass and a longer running time would easily increase statistics so that the statistical error can be made much smaller. It is less straightforward to decrease at the same time the systematic error, which was 2.7% in CHOOZ. This systematic error comes mainly from uncertainties on the neutrino flux (known with 1% accuracy), the number of effective protons in the target (mainly due to the fiducial cut) and the detector efficiency on the neutrino signal. The only way to get rid of these uncertainties is to build two identical detectors, one near the reactor(s) and the other at the far distance. The disappearance of ν_e will be looked for by comparing the ν_e rates and spectra at the 2 locations. The near detector will in first order measure the same flux of neutrinos (in case of no oscillations) with the same target mass and the same efficiency as

the far detector, so that the remaining systematic error is the precision with which one can ascertain this identity (knowing that counting rates for signal and for background will be different at the 2 locations, the dead times generated by cosmics crossing the detector will be different, energy calibrations will be done at each detector independantly with a finite precision,...). Present studies conclude that a 0.5% systematic error should be reachable.

There are presently several projects worldwide, but the most advanced seems to be the european project double-CHOOZ [67]: it would run near Chooz reactors, using the existing underground cavity for the far detector while a near location has to be prepared 100 m to 200 m from the reactor cores. This location will be much shallower, but an artificial mound on top of it will reduce the cosmic rate to an acceptable level. The detector design has been improved to get a higher target mass, a higher efficiency on neutron capture together with a better definition of the fiducial volume, by adding, between the central target filled with scintillator loaded with Gadolinium and the photomultiplier wall, two layers (first a layer of unloaded liquid scintillator, called γ-catcher, to detect some of the γ emitted by the neutron capture on Gd, and then of a non scintillating liquid which absorbs without giving any signal the majority of the radioactivity induced by the photomultipliers, mainly the γ's from potassium). As in CHOOZ, a final scintillator layer outside the photomultipliers, optically separated from the target, serves as a veto against cosmic activity. Simulations show that a sensitivity on θ_{13} of 5 degrees could be reached after 3 years of data taking.

6.5 The first generation of superbeams

A first series of projects is due to start within a few years. They will use neutrino superbeams produced by proton beams with a power below 1 MW.

At FNAL the NOνA project [68] will use the same beam as MINOS, but with a different detector. There are two reasons for this

1 The MINOS detector is not optimized to detect electrons, as it is a calorimeter with rather low granularity, so that it is difficult to distinguish electrons and π^0 decays. The new detector will on the contrary be a calorimeter with high granularity

2 The new detector will be installed off-axis, that is at an angle with respect to the neutrino beam axis. This trick takes advantage of the fact that if a π decays by emitting a neutrino at angle θ in the decay tunnel, this neutrino will have a maximal energy given by 30 MeV/θ (θ in radians); furthermore, the neutrino spectrum

will show a strong peak at this energy, and very few neutrinos above this maximal energy (this small tail at high energy being due to imperfect focusing of the pions along the beam axis or due to neutrinos coming from μ or K decays). By tuning the angle θ, it is then possible to choose the peak energy and optimize it to have the maximum of the oscillation at the detector. This has also the advantage of reducing the ν_e background under the energy peak of the ν_μ, as ν_e from μ or K decays will keep a wide energy spectrum.

This new detector, 50 kton in mass, should be ready for physics around 2010. The expected sensitivity on θ_{13} is 2.5 degrees at 90% CL.

In Japan a new proton accelerator, with an energy of 50 GeV and a power of 800 kW, is being built (mainly for nuclear physics) at Tokai, north of Tokyo. It has been proposed to produce with this accelerator a very intense neutrino beam aimed at Superkamiokande, 300 km away. This is the T2K (for Tokai to Kamioka) project [69]. The off-axis trick mentionned above will be used to concentrate the neutrino energy around its optimal value of 700 MeV or so. The beam should be ready for 2009, and after 5 years of data taking, a sensitivity on θ_{13} of 2.5 degrees should be achieved.

The two above-mentionned experiments will not have sensitivity on δ, mainly by lack of statistics. But the situation will improve with the next generation of superbeams.

6.6 The second generation of superbeams

The natural way to improve on the first generation superbeam experiments is to increase both the beam intensity and the target mass.

In Japan, the strategy is the following : after 2012, increase to power of the proton accelerator to 4 MW and then, replace Superkamiokande by a bigger detector, 1 megaton in mass, Hyperkamiokande. This new detector is just a scaling-up of Superkamiokande, it will be a huge horizontal cylinder, 50 meters in diameter, and several hundred meters long. A site for this detector has already been selected, not far from Superkamioka, and the off-axis neutrino beam at Tokai will be built so that it makes the same angle with Superkamiokande and Hyperkamiokande. But in the japanese strategy, Hyperkamiokande will not be proposed unless T2K sees an appearance signal. It means that Hyperkamiokande will not be ready before 2020-2025.

In Europe, the Superconducting Linac (SPL), accelerating 4 MW of protons to 2.2 GeV, will be the first accelerating device of the neutrino factory; by hitting a target, pions will be produced and focused by a magnetic system; their subsequent decay will produce the muons for

the neutrino factory, but also a very intense ν_μ beam, with a mean energy of 270 MeV. The optimal distance for ν_e appearance is thus around 130 km, that is precisely the distance between CERN and the existing underground Frejus laboratory, on the franco-italian boundary in the road tunnel between Modane and Bardonecchia. The idea is to excavate a cavity able to house a water Cerenkov detector of the megaton class (or alternatively a liquid argon detector of 100 kilotons). A second gallery will be dug parallel to the existing tunnel in the next few years (for safety reasons), and one could use this opportunity to make the cavity less expensive, due to the presence of all the required machinery. This cavity could be ready around 2012, so that ideally, a big detector could start doing physics as soon as 2015. It has to be emphasized that such a detector, apart from receiving a neutrino superbeam from CERN, will address equally fundamental questions, such as proton decay (with a strong potential for its discovery) and supernova explosion searches. Monte-Carlo studies show a sensitivity on θ_{13} of 1 degree (90% CL) after 5 years (see fig. 18a). Running both in neutrinos and antineutrinos would give sensitivity to CP violation at 3 σ down to θ_{13} values of 4^0 if CP violation is maximal (see fig. 18d). But this imposes to run longer in antineutrinos than in neutrinos, due to several factors : less intensity in $\bar{\nu_\mu}$ as less π^-'s are produced on target, and smaller $\bar{\nu}$ cross-section (by nearly a factor 5 at these low energies). This is why, for 10 years of running time, 8 have to be spent in $\bar{\nu_\mu}$ and only 2 in ν_μ.

For this second generation of experiments, statistics will be much higher (by a factor 250 for Hyperkamioka), which imposes a very good control and understanding of the backgrounds. These backgrounds are dominated first by the intrinsic contamination of ν_e (or $\bar{\nu_e}$) in the beam, and second by the confusion between π^0's produced by neutral current interactions in the detector and genuine electrons. The lower the energy, the lower this last background will be, as less π^0's will be produced, while their decay will more often give two distinctive Cerenkov rings rather than a single one. Near detectors will be necessary in all projects to measure this background before the oscillation develops, so that it is known at the required percent level. Such near detectors are actually already foreseen for the first generation experiments NOνA and T2K, and will allow to experimentally assess if the required precision on backgrounds is reachable (in these experiments, a 10% precision on the background is sufficient).

6.7 A new idea : the beta beams

As we have seen, the sensitivity on CP violation in superbeam projects of second generation is limited by the strong asymmetry in running time between neutrinos and antineutrinos. However, a new idea has been recently proposed [70] which avoids such a limitation, based on the concept of beta beams.

Beta beams start from the same idea as neutrino factories, but replace muons by radioactive nuclei. Producing radioactive ions is routinely done by nuclear physicists, in facilities like Ganil in France, Isolde at CERN, at GSI, etc..The ISOLDE technique consists in sending protons on a suitable target to produce radioactive ions, with lifetimes around 1 second. These ions are collected and accelerated for further studies of nuclear structure far from stability, nuclear processes in stars, and so on. If one were able to accelerate these ions up to relativistic γ factors of 100, and store them in a decay ring with long straight sections, one would obtain neutrino beams which are :

- purely $\bar{\nu_e}$ (in case of β^- decays) or ν_e (for β^+ decays), with impurities in ν_μ in the 10^{-4} range.
- strongly collimated due to the very limited transverse momentum of produced neutrinos (so-called Q value).
- of perfectly known energy spectrum as it is just a boosted beta spectrum.

The best candidate ions are ^{18}Ne for ν_e and ^{6}He for $\bar{\nu_e}$; these rare gases are most easily extracted from the target since they are chemically inert.

Recently, nuclear physicists have proposed a new facility called EURISOL which would use a SPL-like proton driver to increase ion production by 2 orders of magnitude compared to the present state of the art. It has been found that one could then produce neutrino beta beams of at least equal if not superior quality than the SPL superbeam. The general scheme for such beta beams is shown on fig. 17. It makes use of the EURISOL design at low energy, and then takes advantage of the existing accelerator CERN complex (PS and SPS) to accelerate ions to the required energy and finally store them into a decay ring which has to be built. Monte Carlo studies show that running in $\bar{\nu_e}$ would give results similar to the ν_μ SPL superbeam. But a big advantage of beta beams is the possibility to store at the same time ^{18}Ne and ^{6}He in the decay ring (in different bunches), so that the detector receives (alternately) the ν_e and $\bar{\nu_e}$ beams, which is excellent to avoid systematics linked to any slow evolution of the detector, contrary to superbeams which would switch

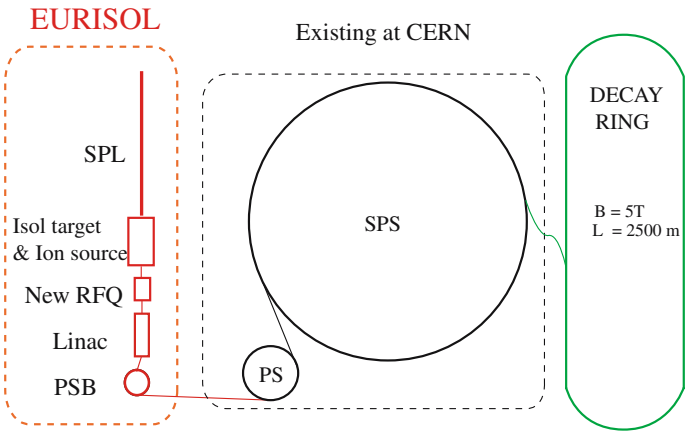

Figure 17. Conceptual design of a beta beam at CERN. At left, the low energy part where radioactive ions are produced using protons from SPL. Middle, acceleration by existing CERN accelerators PS and SPS. Right, the storage and decay ring producing ν_e and $\bar{\nu_e}$ beams sent to Frejus.

polarity at most a few times per year. Studies show that after running the beta beams during 10 years, a 90% CL sensitivity on θ_{13} of 0.5^0 can be achieved (for δ =0), while the 90% CL sensitivity to maximal CP violation extends down to $\theta_{13} = 1^0$ (see fig. 18c). If we can run both the superbeam (2 years of ν_μ, 8 years of $\bar{\nu_\mu}$) and the beta beam with both types of ions during the same 10 years, the 3 σ potential for discovery of CP violation extends down to $\theta_{13} = 0.8^0$ for maximal violation, but is fulfilled for δ above 20^0 when θ_{13} is above 2.5^0 (see fig. 18d). This would be by far the best project in the timescale of these second generation projects (see fig. 18b).

6.8 A european strategy

From the preceeding, Europe has a good opportunity to make a leading contribution to the determination of θ_{13} and δ_{CP}, provided that SPL can be built at CERN for both Neutrino and EURISOL communities. Fruitful contacts have been taken between these two communities, and beta beams are now part of the design studies of the EURISOL project. Concerning the Frejus site, italian and french institutions have recently agreed to jointly push forward the new cavity, which could become together with Gran Sasso a european (multisite) facility open to the world community. Contacts have also been taken with our american colleagues, who propose a Cerenkov detector (called UNO [71]) of half a megaton which could be located in a Colorado mine. But in their case, it is

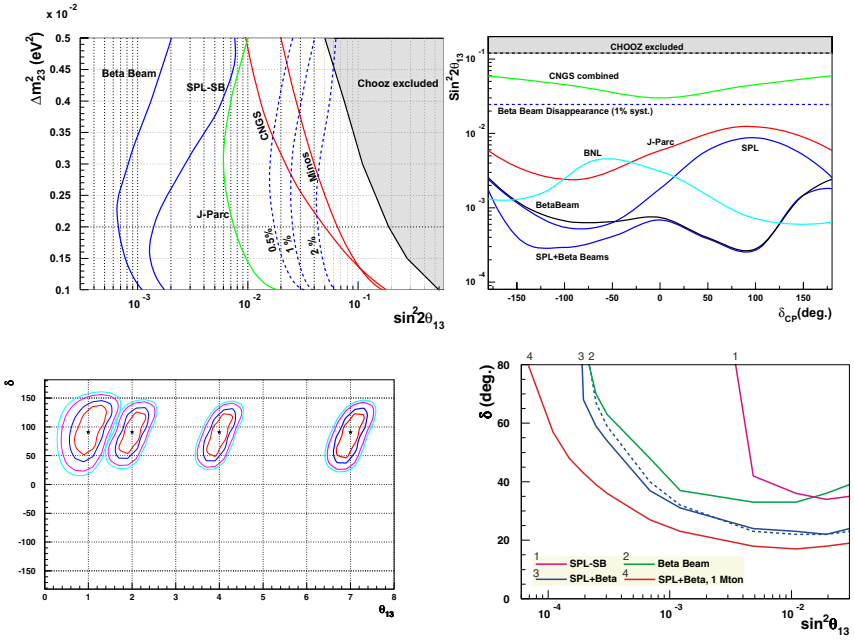

Figure 18. All these figures are based on a detector with 440 kton fiducial mass.
a) upper left : expected 90% CL sensitivity on θ_{13} (for $\delta = 0$) after 5 years of running with ν_μ from the SPL superbeam (SPL-SB), or with the beta-beam running 5 years with both ν_e and $\bar{\nu_e}$, compared to the present limit from CHOOZ and the expected sensitivities after 5 years of ν_μ runs for MINOS, OPERA and ICARUS (curve CNGS) and T2K phase 1 (curve J-Parc)
b) upper right : expected 90% CL sensitivity on θ_{13} as a function of δ for the different projects, in the same conditions (the lower curve corresponds to the combined analysis of beta-beam and SPL-superbeam results). Notice the strong dependance on the sign of δ for superbeams running with only ν beams, while the effect is much smaller for the betabeam running with both ν and $\bar{\nu}$
c) lower left : Fits to θ_{13} and δ after a 10 year betabeam run, in the case of maximal CP violation. Stars represent the true parameters, and contours are fit results for 68%, 90%, 99% and 99.7% CL.
d) lower right : 3 σ discovery potential domain in the $\sin^2\theta_{13}\,\delta$ plane after 10 years of running. The discovery potential is above and to the right of the curves; curve 1 : SPL superbeam (2 year ν_μ and 8 year $\bar{\nu_\mu}$); curve 2 : beta-beam; curve 3 : combined superbeam and betabeam data; curve 4 : as curve 3, but for a detector with a 1 Mton fiducial mass

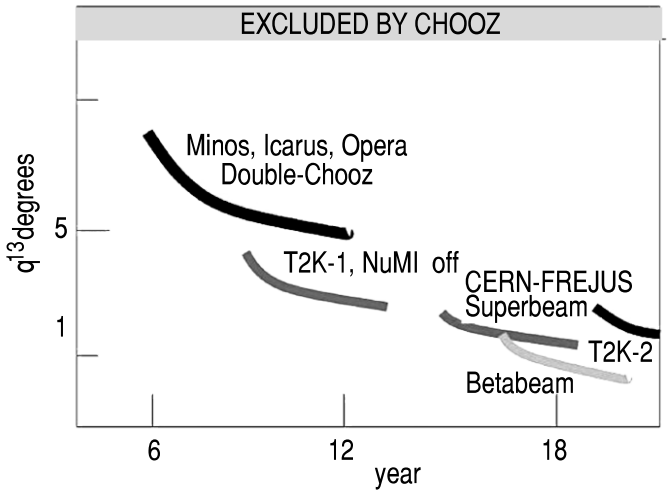

Figure 19. This figure shows how our sensitivity to θ_{13} will improve in the coming years, first with current long baseline and reactor experiments (down to 5^0), then with superbeams of first generation (down to 2.5^0) [NuMI-off stands for NOνA] and finally with superbeams of second generation and beta beams (1^0 and lower).

not clear when and where a neutrino superbeam could be built, and whatever the site (FNAL or BNL), the large distances involved (much bigger than 1000 km) impose multiGeV neutrinos, for which the water Cerenkov technique has problems (π^0 background).

If the cavity in Frejus can be financed and a worldwide community agrees to run there a detector of the megaton class, such a detector could start doing physics on proton decay (this will have to run at least for 20 years) and expect a supernova explosion (which would give 100 000 neutrino events within a few seconds) around 2015. And as soon as CERN has built the SPL, a superbeam could be sent to Frejus. If furthermore, the EURISOL complex is located at CERN, it would boost the possibility to send beta beams to Frejus, hopefully before 2020.

Figure 19 shows how our search for θ_{13} will progress in the coming years. If one wants to look further in the future, precision measurements on neutrino mixings could be performed using a neutrino factory or alternatively a new generation of very high energy beta beams (with ions accelerated with relativistic γ of 600 or higher). Neutrino mixing studies will then reach the degree of precision which is now achieved in the quark sector.

7. The absolute mass scale

Of course, oscillations will not answer all the open questions on neutrinos. Two main questions will have to be answered :

- what is the absolute mass scale, and the exact ordering of mass eigenstates ?
- are neutrinos Dirac or Majorana particles ?

Concerning the mass hierarchy, only neutrino superbeams above 1 GeV or even better the neutrino factory would bring the answer, as we have already seen. But the absolute mass scale will stay unknown. To determine this mass scale, one could hope for a possible answer from KATRIN, if neutrino masses are bigger than 0.4 eV. Due to the small splittings already measured, that would mean that the 3 masses are nearly equal (degenerate case). But from these splittings, we can only say that the highest mass has to be bigger than roughly 50 meV (the square root of Δm_{23} given by the atmospheric oscillation). This leaves room for a negative result from KATRIN. The only hope would then reside in two other experimental approaches, the neutrinoless beta decay and the cosmological observations.

7.1 Double beta experiments

Double beta decays occur for some nuclei for which the single beta decay is energetically forbidden, while the simultaneous transformation of 2 neutrons into 2 protons, with the emission of 2 electrons and 2 antineutrinos, is possible. This is a second order process for weak interactions, which is allowed by standard theory but very rare (very long lifetimes). It was however directly observed within the 10 last years for some isotopes. A more exotic process would be the neutrinoless double beta decay, with only 2 electrons being emitted and no neutrinos. In this process, the 2 neutrinos normally emitted stay virtual and annihilate each other. For this to occur, the neutrino should bear no charge at all, that is be a Majorana neutrino (which can annihilate with itself since it is its own antiparticle). But we must take also into account the spin structure of the neutrinos. An antineutrino emitted with right chirality at one vertex has to be absorbed at the other vertex as a neutrino with left chirality, according to V-A. Chirality has to be violated, which means that neutrino has to be massive, and actually the amplitude for this diagram will vanish for a massless neutrino. For massive neutrinos, one should add 3 such diagrams for the 3 mass eigenstates, and the couplings at each vertex will be proportional to U_{ei} while the propagator

will have a factor m_i at the numerator. The total amplitude is thus proportional to an effective neutrino mass $m_{\rm eff}$ given by:

$$m_{\rm eff} = \sum_{i=1}^{3} U_{ei}^2 m_i$$

One has to remember that the mixing matrix elements U_{ei} will in general be complex, so that cancellations may occur and $\mid m_{\rm eff} \mid$ be smaller than the highest eigenmass. Actually, if neutrinoless double beta decay is observed, $\mid m_{\rm eff} \mid$ can be deduced from the observed decay rate and a direct consequence is that at least one value m_i will be bigger than $m_{\rm eff}$. In order to deduce the effective mass from the decay rate, one has to know the nuclear amplitude describing the overlap between the nucleus initial and final states. This cannot be done exactly and one has to rely on models of nuclear structure. One can say that our present understanding of nuclear physics allows to compute $m_{\rm eff}$ with an uncertainty not bigger than a factor 2.

The signal of neutrinoless double beta decay is a monoenergetic line at the end point location for the energy sum of the 2 electrons. The challenge of such experiments is to use ultrapure material to avoid as much as possible radioactive activity at the energy of interest. Moreover, as the allowed normal double beta decay will certainly occur, its tail will produce an irreducible background which will increase with the energy resolution σ_E on the energy sum. Excellent resolution is a plus to decrease this background. There are two main types of experiments:

- the calorimetric type, where only the total energy deposit is measured. The best example is the Germanium crystal, used as a semiconductor, which measures with high precision the energy sum of double beta decays of ^{76}Ge. Actually, this type of experiment has given the best present limit on $m_{\rm eff}$ at 0.3-0.5 eV obtained by the Heidelberg-Moscow collaboration [72], which used crystals of enriched ^{76}Ge for a total exposure of 40 kg years. Recently, it was even claimed that a positive signal was seen in this experiment [73], but this announcement created a lot of questions and criticisms [74].

- tracking devices, where both electron tracks are measured (for example by a TPC, or a system of Geiger or proportional tubes). Two subcategories may be defined in this second type of devices, depending on whether the source is or not the detecting medium. Xenon TPC's are an example of the first category [75] (the xenon gas being also the emitter, as ^{136}Xe is a double beta isotope) while

the NEMO3 experiment [76] tracks electrons emitted by a foil on which the emitter isotopes are deposited with a system of Geiger tubes and scintillator calorimeters.

The NEMO3 experiment, which is presently running in the underground Frejus laboratory, is expected to improve the $m_{\rm eff}$ sensitivity down to 0.2 eV by using, among other isotopes, 7 kg of enriched ^{100}Mo. Visualizing and measuring each electron helps in rejecting some backgrounds (such as single electrons) and also allows to compare electron energies and angles with models of double beta decays. NEMO3 has already measured normal double beta decays for ^{100}Mo with very high statistics, allowing for tests of nuclear matrix element computations, and also for ^{82}Se, ^{116}Cd and ^{150}Nd with lower statistics. The only drawback of this experiment is the relatively poor energy resolution compared to Germanium devices.

More ambitious projects, with masses of emitting isotopes reaching 1 ton, have been proposed and are at the prototyping level. Let us cite the GENIUS project [77], where 1 ton of naked ^{76}Ge crystals are immersed in a large tank of liquid nitrogen acting as a veto against outside radioactivity, or the CUORE [78] project which plans to use bolometers made of 760 kg of $\mathrm{Te\,O_2}$. If the background counting rate can be kept as low as in previous experiments despite the fact that masses have been increased by more than one order of magnitude, a sensitivity of few tens of meV on $m_{\rm eff}$ could ultimately be achieved by these experiments.

The fig. 20 shows, as a function of the lightest neutrino mass, what could be the value of $m_{\rm eff}$. Bands in this plot are due to the effect of the unknown Majorana phases; it is interesting to notice that a positive signal should be seen above 10 meV if the mass hierarchy is inverted, while prospects are less favorable in case of a normal hierarchy. For a comprehensive survey of double beta projects, see [80].

7.2 Constraints from cosmology

The formation of large structures in the early Universe depends on its matter composition, and neutrinos play a role in this evolution [81]. Relativistic neutrinos tend to stream freely in the gravitational potential and erase density fluctuations. For neutrino masses between 10^{-3} and 0.3 eV, the transition to a non-relativistic regime for neutrinos will take place during structure formation, so that matter density fluctuations will be affected by neutrinos, in a way which depends upon their mass. Briefly speaking, smaller scale structures will be damped while larger will not, and the typical transition scale is roughly proportional to the sum of neutrino masses Σm_i. The study of large scale structures in the Universe (galaxies, galaxy clusters), as obtained by 2dFGRS (2 degree

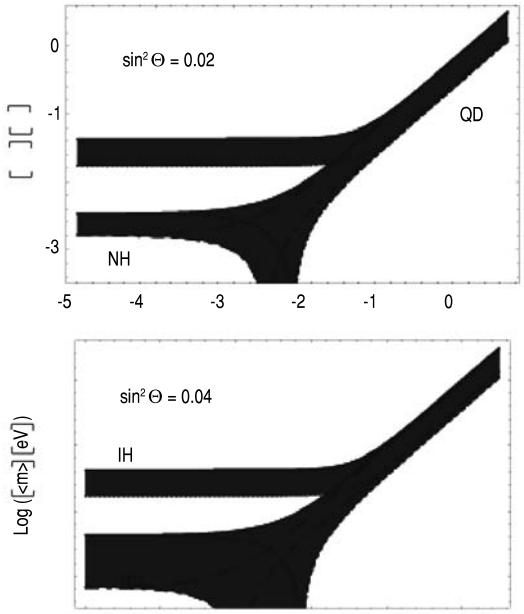

Figure 20. Values of the effective mass for neutrinoless beta decays authorized by present neutrino data as a function of the lightest mass, for different values of $\sin\theta_{13}$. The labels NH, IH and QD correspond to normal hierarchy, inverted hierarchy and quasi-degenerate cases for the masses

field galaxy redshift survey) and SDSS (Sloan digital sky survey) experiments, gives the matter power spectrum of our Universe, sensitive to neutrino masses in a range which covers our expectations. But to derive a value for the neutrino mass, one needs as inputs the cosmological parameters (Ω_m, Λ,...) which are determined from the study of the cosmological microwave background (CMB). After the latest measurements of WMAP [9], it was possible to deduce an upper limit on Σm_i which ranges between 0.6 and 1 eV, depending on the choice of data entering the analysis and some "theoretical" priors [83]. With more precise data on CMB expected from next generation experiments like PLANCK, better constraints on neutrino masses will be reached; the estimated uncertainty on Σm_i will be 0.1 eV using PLANCK and SDSS, and it could go down to 0.06 eV after using data on CMB polarization [84].

Observational cosmology might therefore very well contribute to the determination of the absolute mass scale of the neutrinos.

8. Outlook

We have learnt a lot on neutrinos since they were discovered in 1956, but many of their properties are still unknown.

In the near future, we will get the result from MiniBooNE (first results in 2005) which will hopefully clarify the LSND result. If confirmed, that would be a revolution in neutrino world : either we have a fourth family of neutrinos, or we falsely interpreted one of the oscillation evidences, or neutrinos are very exotic particles.

The result of NEMO3 is also due to come in the next few years, and it will be interesting to see if it confirms with ^{100}Mo the recent claim of a signal seen with ^{76}Ge.

MINOS, OPERA, ICARUS for atmospheric oscillations, and BOREXINO for solar oscillations are expected to confirm and precise these oscillations.

The real challenge will be first to discover the oscillation powered by θ_{13}, and if is actually observed, determine if CP is violated or not in the neutrino sector. This program will take at least 10 to 20 years, which leaves time for other experiments to bring new results or surprises, concerning for example the absolute mass scale.

Neutrino physics will soon enter the era of precision measurements, to the level reached today in the quark sector. This delay is quite understandable, due to the intrinsic difficulties met in the study of this fantomatic particle, but experimentalists have found ways to circumvent them, and the neutrino conferences will continue to bring exciting results in the coming decades.

References

[1] M. C. Gonzalez-Garcia and Y .Nir, Rev. Mod. Phys. 75 (2003) 345; V. Barger, D. Marfatia and K. Whisnant, hep-ph/0308123; W. M. Alberico and S.M.Bilenky, hep-ph/0306239

[2] K.G.Emeléus, Proc.Camb.Phil.Soc.22 (1924) 400

[3] C.D.Ellis, Proc.Roy.Soc.A117 (1927) 109

[4] W.Pauli,"Physik und Erkenntnistheorie" (ed. Friedr, Vieweg und Sohn, Braunschweig 1984) page 156

[5] M.E.Nahmias, Proc.Camb.Phil.Soc.31 (1935) 99

[6] W.J.Henderson, Proc.Roy.Soc.A147 (1934) 572

[7] H.R.Crane and J.Halpern, Phys.Rev. 53 (1938) 789 ; 56 (1939) 232

[8] Ch.W.Sherwin, Phys.Rev.82 (1951) 52

[9] H.A.Bethe and R.Peierls, Nature 133 (1934) 532

[10] H.R.Crane, Phys.Rev. 55 (1939) 501

[11] C.L.Cowan Jr, F.Reines, F.B.Harrison, H.W.Kruse, A.D.McGuire Science, vol.124 (1956) 103; F.Reines and C.L.Cowan Jr, Nature, vol.178 (1956) 446

[12] C.S.Wu *et al.*, Phys.Rev.105 (1957) 1413

[13] T.D.Lee and C.N.Yang, Phys.Rev.104 (1956) 254 and Phys.Rev.105 (1957) 1671

[14] G.Danby *et al.*, Phys.Rev.Letters 9 (1962) 36

[15] K.Kodama *et al.*, NIM A493 (2002) 45

[16] F.J. Hasert *et al.*, Phys.Lett. 46B (1973) 121

[17] J. Blietschau *et al.*, Nucl.Phys. B114 (1976) 189

[18] B. Kayser *et al.*, Phys.Rev. D20 (1979) 87

[19] R.C. Allen *et al.*, Phys.Rev.Lett. 64 (1990) 1330

[20] E.Pasierb *et al.*,Phys.Rev.Letters 43 (1979) 96; F.Reines *et al.*, Phys.Rev.Letters 45 (1980) 1307

[21] P. Astier *et al.*, Nucl.Phys. B611 (2001) 3 ; J. Altegoer *et al.*, NIM A404 (1998) 96

[22] E. Eskut *et al.*, Phys. Lett. B497 (2001) 8 ; E. Eskut *et al.*, NIM A401 (1997) 7

[23] M. Gell-Mann, P. Ramond and R. Slansky, in "Supergravity", ed. by F. van Nieuwenhuizen and D. Freedman, North-Holland, Amsterdam, 1979, p. 315; T. Yanagida, in "Proc. of the Workshop on Unified Theory and the Baryon Number in the Universe", KEK, Japan, 1979 ; R.N. Mohapatra and G. Senjanovic, Phys.Rev.Lett. 44 (1980) 912

[24] E.Fermi, La Ricerca Scientifica 2 (1933) 491 ; Il Nuovo Cimento 11 (1934) 1 ; Zeitschrift für Physik 88 (1934) 161

[25] V. A. Lobashev *et al.*, Phys. Lett. B460 (1999) 227 ; Nucl. Phys. B (Proc. Suppl.) 91 (2001) 280

[26] C. Weinheimer, hep-ex/0210050 ; C. Weinheimer, hep-ex/0306057

[27] A. Osipowicz *et al.*, hep-ex/0109033

[28] Z. Daraktchieva *et al.*, Phys.Lett. B564 (2003) 190 [hep-ex/0304011]

[29] Y. Giomataris and J.D. Vergados, hep-ex/0303045

[30] B. Kayser, Phys. Rev. D24 (1981) 110

[31] M. Apollonio *et al.*, Eur.Phys.J. C27 (2003) 331-374 and Phys.Lett. B466 (1999) 415

[32] Raymond Davis Jr *et al.*, Phys.Rev.Lett. 20 (1968) 1205 ; B.T.Cleveland *et al.*, Ap. J. 496 (1998) 505 ; K.Lande and P.Wildenhain, Nucl. Phys. B (Proc. Suppl.) 118 (2003) 49

[33] M. C. Gonzalez-Garcia and Y .Nir, Rev. Mod. Phys. 75 (2003) 345

[34] M. Altmann *et al.*, Phys. Lett. B490 (2000) 16 ; T. Kirsten, Nucl. Phys. B (Proc. Suppl.) 118 (2003) 33

[35] J. N. Abdurashitov *et al.*, Phys. Rev. C60 (1999) 055801 ; J. N. Abdurashitov *et al.*, Phys. Rev. Lett. 83 (1999) 4686 ; V. N. Gavrin, Nucl. Phys. B (Proc. Suppl.) 118 (2003) 39

[36] W. Hampel *et al.*, Phys. Lett. B420 (1998) 114

[37] W. Hampel *et al.*, Phys. Lett. B436 (1998) 158

[38] M. B. Smy, Nucl. Phys. B (Proc. Suppl.) 118 (2003) 25 (hep-ex/0208004) ; see also S. Fukuda *et al.*, Phys. Rev. Lett. 86 (2001) 5651 (hep-ex/0103032)

[39] see J. N. Bahcall, "Neutrino Astrophysics", Cambridge University Press, 1989

[40] J. N. Bahcall, M.H.Pinsonneault and S.Basu, Ap. J. 555 (2001) 990 ; S.Turck-Chièze *et al.*, Ap. J. Lett. 555 (2001) L69

[41] S. Turck-Chièze *et al.*, Ap. J. Lett. 555 (2001) L69 ; A.S. Brun, S. Turck-Chièze and P. Morel, Ap. J. 506 (1998) 913

[42] S .Couvidat, S. Turck-Chièze and A.G. Kosovichev, Ap. J. 599 (2003) 1434

[43] S. Bludman, N. Hata, P. Langacker, Phys.Rev. D49 (1994) 3622 [hep-ph/9306212]

[44] Z. Maki, M. Nakagawa, S. Sakata, Prog. Theor.Phys. 28 (1962) 870; V. N. Gribov and B. Pontecorvo,Phys. Lett. B28 (1969) 493;

[45] J.N. Bahcall and N. Cabibbo, Phys.Rev.Lett. 28 (1972) 316

[46] L. Wolfenstein, Phys. Rev. D17 (1978) 2369

[47] S. P. Mikheyev and A. Yu. Smirnov, Nuovo Cimento 9C (1986) 17

[48] J. N. Bahcall, P. I. Krastev and A. Yu. Smirnov, hep-ph0103179; G.L. Fogli *et al.*, Phys. Rev. D64 (2001) 093007 [hep-ph/0106247]

[49] Q. R. Ahmad *et al.*, Phys. Rev. Lett. 87 (2001) 071301 [nucl-ex/0106015] and Phys. Rev. Lett. 89 (2002) 011301 J. Boger *et al.*, Nucl. Instr. and Meth. A449 (2000) 172 [nucl-ex/9910016]

[50] S. N. Ahmed *et al.*, submitted à Phys. Rev. Lett., nucl-ex/0309004;

[51] see D.O.Caldwell and P.A.Sturrock, hep-ph/0309191

[52] K. Eguchi *et al.*, Phys. Rev. Lett. 90 (2003) 021802 [hep-ex/0212021]

[53] J. N. Bahcall and C. Peña-Garay, JHEP 0311 (2003) 004 [hep-ph/0305159]

[54] G.Alimonti *et al.*, Astroparticle Physics 16 (2001) 205 [hep-ex/0012030]

[55] for the most recent results, see T. Toshito, hep-ex/0105023

[56] A. Aguilar *et al.*, Phys. Rev. D64 (2001) 112007

[57] B. Armbruster *et al.*, Phys. Rev. D65 (2002) 112001

[58] E. Church *et al.*, FERMILAB-P-0898 (1997) [http://library.fnal.gov/archive/test-proposal/0000/fermilab-proposal-0898.shtml]; see http://www-boone.fnal.gov/publicpages/news.html for latest news

[59] M. H. Ahn *et al.*, Phys. Rev. Lett. 90 (2003) 041801

[60] D. Michael, Nucl. Phys. B (Proc. Suppl.) 118 (2003) 189

[61] P. Strolin, http://operaweb.web.cern.ch/operaweb/documents/archive/operaspsc.pdf

[62] F. Arneodo *et al.*, hep-ex/0103008

[63] G.L. Fogli *et al.*, hep-ph/0308055, hep-ph/0310012

[64] V. Barger, D. Marfatia and K. Whisnant, Int.J.Mod.Phys. E12 (2003) 569 [hep-ph/0308123]

[65] P. Huber *et al.*, Nucl.Phys. B665 (2003) 487; hep-ph/040306

[66] http://www-kuno.phys.sci.osaka-u.ac.jp/~nufact04/index.html

[67] Letter of Intent for Double-CHOOZ: a Search for the Mixing Angle Theta13, hep-ex/0405032

[68] http://www-off-axis.fnal.gov/NOvA proposal March 15 2004.pdf

[69] Y. Itow *et al.*, hep-ex/0106019

[70] P.Zucchelli, Phys.Lett. B532 (2002) 166; J. Bouchez, M. Lindroos and M. Mezzetto, the proceedings of Nufact 03, New York, 2003 (see http://beta-beam.web.cern.ch/beta-beam/references.htm)

[71] http://superk.physics.sunysb.edu/nngroup/uno/whitepaper/index.html

[72] L. Baudis *et al.*, Phys. Rev. Lett. 83 (1999) 41 [hep-ex/9902014]

[73] H.V. Klapdor-Kleingrothaus *et al.*, Mod. Phys. Lett. A16 (2001) 2409; H.V. Klapdor-Kleingrothaus *et al.*, NIM A522 (2004) 371

[74] C.E. Aalseth *et al.*, Mod. Phys. Lett. A17 (2002) 1475; F. Feruglio, A. Strumia and F. Visani, Nucl.Phys. B637 (2002) 345 and addendum ibid B659 (2003) 359; H.V. Klapdor-Kleingrothaus *et al.*, Found. Phys. 32 (2002) 1181 and erratum ibid 33 (2003) 679

[75] R. Luescher *et al.*, Phys.Lett. B434 (1998) 407

[76] NEMO collaboration, hep-ex/0006031 ; A. Barabash, TAUP03 conference, Seattle, sept.2003 (see http://mocha.phys.washington.edu/int-talk/WorkShops/TAUP03/)

[77] H.V. Klapdor-Kleingrothaus, hep-ph/0103074

[78] A. Alessandrello *et al.*, Phys. Atom. Nucl. 66 (2003) 452

[79] S. Pascoli and S. T. Petcov, hep-ph/0310003

[80] S.R. Elliott, TAUP03 conference, nucl-ex/0312013

[81] A.D. Dolgov, Phys.Rept 370 (2002) 333 [hep-ph/0202122]

[82] D.N. Spergel *et al.*, Astrophys.J.Suppl. 148 (2003) 97 [astro-ph/0302208]

[83] V. Barger, D. Marfatia and A. Tregre, hep-ph/0312065 ; P. Crotty and J. Lesgourgues, astro-ph/0402049

[84] J. Lesgourgues, S. Pastor and L. Perotto, astro-ph/0403296

BARYO- AND LEPTOGENESIS (OUTLINE)*

Wilfried Buchmüller
Deutsches Elektronen-Synchrotron DESY,
22603 Hamburg, Germany
buchmuwi@mail.desy.de

Abstract The cosmological matter-antimatter asymmetry can be understood as consequence of *CP* violating quark and lepton interactions in the early universe. We review the theoretical foundations and the main baryogenesis mechanisms: electroweak baryogenesis, leptogenesis, Affleck-Dine baryogenesis.

Keywords: Baryogenesis, CP, electroweak phase transition, flat directions, leptogenesis, neutrino mass, Sphaleron

1. Introduction

The explanation of the cosmological baryon asymmetry is a challenge for particle physics and cosmology. In an expanding universe, which leads to departures from thermal equilibrium, C, CP and baryon number violating interactions of quarks and leptons can generate dynamically a baryon asymmetry [1].

Possible realisations of these conditions have been studied during the past 25 years, starting with detailed investigations in the context of grand unified theories [2]. The classical GUT baryogenesis scenario is strongly restricted by the connection between baryon and lepton number in the high-temperature, symmetric phase of the standard model [3]. In many respects, decays of heavy Majorana neutrinos in the thermal phase of the early universe are an ideal source of the cosmological baryon asymmetry. This leptogenesis idea [4] leads to a beautiful connection

*Lectures given at the Cargese School on Particle Physics and Cosmology: The Interface, August 2003, Cargese, France

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 171–179.

between the matter-antimatter asymmetry of the universe and neutrino properties.

In the following sections the main mechanisms of baryogenesis as well as some recent developments are reviewed. For each section we list a few basic papers and review articles which contain further details and more complete lists of references. This complements the lecture notes handed out during the school. The goal is to help the interested reader penetrate the by now extensive literature on the subject.

2. Baryon asymmetry: evidence and conditions for baryogenesis

One of the main successes of the standard early-universe cosmology is the prediction of the abundances of the light elements, D, ^{3}He, ^{4}He and ^{7}Li. Agreement between theory and observation is obtained for a certain range of the parameter η_B, the ratio of baryon density and photon density [5],

$$\eta_B^{BBN} = \frac{n_B}{n_\gamma} = (2.6 - 6.2) \times 10^{-10} \,, \tag{1}$$

where the present number density of photons is $n_\gamma \sim 400/\mathrm{cm}^3$. Since no significant amount of antimatter is observed in the universe, the baryon density yields directly the cosmological baryon asymmetry, $\eta_B = (n_B - n_{\bar{B}})/n_\gamma$.

On the experimental side, the precision of measurements of the baryon asymmetry has significantly improved with the observation of the acoustic peaks in the cosmic microwave background radiation (CMB). The most recent measurement of the WMAP Collaboration is consistent with the BBN result,

$$\eta_B^{CMB} = (6.3 \pm 0.3) \times 10^{-10} \,, \tag{2}$$

with an error of only 5% [6], which is a most remarkable test of the standard cosmological model.

A matter-antimatter asymmetry can be dynamically generated in an expanding universe if the particle interactions and the cosmological evolution satisfy Sakharov's conditions,

- baryon number violation,
- C and CP violation,
- deviation from thermal equilibrium.

Although the baryon asymmetry is just a single number, it provides an important relationship between the standard model of cosmology, i.e.

the expanding universe with Robertson-Walker metric, and the standard model of particle physics as well as its extensions.

At present there exist a number of viable scenarios for baryogenesis. They can be classified according to the different ways in which Sakharov's conditions are realised. In grand unified theories B and L are broken by the interactions of quarks and leptons with gauge bosons and leptoquarks. This is the basis of classical GUT baryogenesis. Analogously, the lepton number violating decays of heavy Majorana neutrinos lead to leptogenesis.

The crucial deviation from thermal equilibrium can also be realized in several ways. One possibility is a sufficiently strong first-order electroweak phase transition which is required for electroweak baryogenesis. For GUT baryogenesis and for leptogenesis the departure from thermal equilibrium is caused by the deviation of the number density of the decaying heavy particles from the equilibrium number density. How strong this departure from equilibrium is depends on the lifetime of the decaying heavy particles and the cosmological evolution.

3. Equilibrium thermodynamics in the Friedmann Universe

The description of a thermodynamic system in an expanding universe is discussed in detail in the books by Bernstein [7] and by Kolb and Turner [2]. This includes the Robertson-Walker metric, thermal distribution functions of boson and fermions, energy and entropy densities with the corresponding effective number of degrees of freedom, g_* and g_S, the relation between asymmetries $(n_i - \bar{n}_i)$ and chemical potentials μ_i, the conditions for chemical and thermal equilibrium etc.

The basis of baryogenesis via leptogenesis is the connection between a baryon asymmetry B and a lepton asymmetry L in the high-temperature, symmetric phase of the standard model [8],

$$B = \frac{8N + 4}{22N + 13}(B - L) \, , \tag{3}$$

where N is the number of generations.

In Sect. 2 of Ref. [9] a brief derivation is given, and some implications for extensions of the standard model, consistent with baryogenesis, are discussed.

4. Sphaleron processes and electroweak transition

Due to the chiral nature of the weak interactions B and L are not conserved. At zero temperature this has no observable effect due to the smallness of the weak coupling. However, as the temperature approaches the critical temperature T_c of the electroweak phase transition, B and L violating processes come into thermal equilibrium [3, 10].

The sphaleron transition rate in the high-temperature, symmetric phase has been evaluated by combining an analytical resummation with numerical lattice techniques [11]. The result is, in accord with previous estimates, that B and L violating processes are in thermal equilibrium for temperatures in the range

$$T_{EW} \sim 100\text{GeV} < T < T_{SPH} \sim 10^{12}\text{GeV} \tag{4}$$

An important ingredient in the theory of baryogenesis is also the nature of the electroweak transition from the high-temperature symmetric phase to the low-temperature Higgs phase [10, 12, 9]. A first-order phase transition yields a departure from thermal equilibrium. Since in the standard model baryon number, C and CP are not conserved, it is conceivable that the cosmological baryon asymmetry has been generated at the electroweak phase transition. This possibility has stimulated a large theoretical activity during the past decade to determine the phase diagram of the electroweak theory.

For large Higgs masses the nature of the electroweak transition is dominated by non-perturbative effects of the $SU(2)$ gauge theory at high temperatures. At a critical Higgs mass $m_H^c = \mathcal{O}(m_W)$ an intriguing phenomenon occurs: the first-order phase transition turns into a smooth crossover, as expected on general grounds. Numerical lattice simulations have determined the precise value $m_H^c = 72.1 \pm 1.4$ GeV. One can also estimate the critical Higgs mass for supersymmetric extensions of the standard model, for which one obtains $m_h^c \simeq 130 \ldots 150$ GeV (cf. Sect. 1 of Ref. [9]).

5. Electroweak baryogenesis

As already mentioned, the standard model contains all the necessary ingredients for baryogenesis, i.e., baryon number violation, C and CP violation, and the required departure from thermal equilibrium could be generated during the electroweak phase transition [10, 12]. The attractive feature of electroweak baryogenesis is that the baryon asymmetry can be calculated in terms of parameters which are measurable in collider experiments. However, due to the lower bound on the Higgs boson mass

from LEP, $m_H > 114$ GeV, electroweak baryogenesis is now excluded in the standard model. In supersymmetric extensions of the standard model it is still viable for some regions of parameter space [13].

6. Kinetic equations and the standard GUT scenario

The simplest possibility for a departure from thermal equilibrium is the decay of heavy, weakly interacting particles in a thermal bath. The most important processes are decays and inverse decays, which have to be combined with the appropriate zero-width limit of the related $2 \to 2$ scattering processes.

A thorough discussion of the basic picture has been given in Ref. [14], with applications to GUT scenarios. The case of leptogenesis is discussed in Ref. [15].

7. Leptogenesis and neutrino properties

In its simplest version leptogenesis is dominated by the *CP* violating interactions of the lightest of the heavy Majorana neutrinos, the seesaw partners of the ordinary neutrinos. During the past years this case has been studied by many groups in considerable detail.

For a given mass M_1 of the lightest heavy Majorana neutrino N_1, the *CP* asymmetry in N_1 decays satisfies an upper bound [16, 17]. As a consequence, the maximal baryon asymmetry, which can be generated in the leptogenesis process, depends on only four parameters: the CP asymmetry ve_1, the heavy neutrino mass M_1, the effective light neutrino mass $\widetilde{m}_1 = (m_D^\dagger m_D)_{11}/M_1$ and the quadratic mean $\overline{m} = \sqrt{m_1^2 + m_2^2 + m_3^2}$ of the light neutrino masses [18].

One can then study the range of parameters for which leptogenesis can generate the observed baryon asymmetry. It turns out that this yields the stringent upper bound on all light neutrino masses of 0.1 eV. Further, if neutrino masses are larger than 10^{-3} eV, a pre-existing asymmetry is efficiently erased so that the final baryon asymmetry is almost independent of initial conditions. Hence, in the neutrino mass window [19],

$$10^{-3} \text{ eV} < m_i < 0.1 \text{ eV} , \tag{5}$$

leptogenesis emerges as the unique source of the cosmological matter-antimatter asymmetry.

Recently, various corrections affecting the final baryon asymmetry have been studied [20]-[23]. The corresponding upper bound on the neutrino masses, which is quoted in the literature, varies between 0.12 eV

and 0.15 eV. For nearly degenerate heavy Majorana neutrinos this bound can be considerably relaxed [24]. A still open, challenging problem is a complete quantum mechanical description of the leptogenesis process.

Since leptogenesis is directly related to the pattern of neutrino masses and mixings, there exists important connections with *CP* violation in the lepton sector [25] and lepton flavour changing processes [26].

The large temperature required for thermal leptogenesis has important implications for dark matter. In particular, gravitinos rather than WIMPs are attractive candidates for cold dark matter. Recent discussions are given in Ref. [27].

8. Affleck-Dine baryogenesis

In supersymmetric theories the scalar potential generically has flat directions. In connection with inflation allows to relate the cosmological baryon asymmetry to the dynamics of scalar fields associated with these flat directions [28]. The final baryon asymmetry is then determined by the value of the scalar field during inflation, the size of *CP* violating, supersymmetry breaking terms in the effective lagrangian, higher dimensional operators which lift the flat directions etc. This and interesting mechanism has been reviewed in Refs. [29, 30].

In particular the LH flat direction has been studied in detail. In this case the dimension-5 operator which lifts the flat direction is related to neutrino masses. It has been found that the observed baryon asymmetry requires an 'ultralight' neutrino [31],

$$m_1 \simeq (0.1 - 1) \times 10^{-9} \text{ eV} . \tag{6}$$

An important qualitative signature for the Affleck-Dine baryogenesis is the existence of non-thermal dark matter [32, 29].

9. Alternative mechanisms

Electroweak baryogenesis, leptogenesis and the Affleck-Dine scenario are all interesting mechanisms for baryogenesis. The simplest, and quantitatively best studied mechanism is thermal leptogenesis for which the consistency with neutrino masses is impressive. However, there are also other intriguing scenarios of baryogenesis [33, 30]. Here we can just list a few alternatives with some references:

- Inflaton decays
  which can induce non-thermal leptogenesis [34, 35],
- Baryogenesis at low reheating temperatures
  which is important in connection with large extra dimensions [36],

- Quintessence and leptogenesis
  which relates baryogenesis to the slow time evolution of a scalar field [37],
- The lepton leaking mechanism
  where the baryon asymmetry is due to the small coupling of the 'visible' world to a mirror world [38],
- Baryogenesis by brane collisions
  which produce the heavy particles whose decays generate the baryon asymmetry [39],
- Sneutrino dominated universe
  where the scalar partners of the heavy Majorana neutrinos dominate the energy density of the universe before the onset of leptogenesis [16],
- Coherent baryogenesis
  where a nonadiabatically varying fermion mass matrix during reheating leads to a baryon asymmetry.

The various baryogenesis mechanisms lead to significantly different expectations for the nature of dark matter, which may eventually allow us to identify the true origin of the matter-antimatter asymmetry.

Acknowledgment
It is a pleasure to thank Dimitri Kazakov and Gerard Smadja for organizing a lively and very enjoyable school.

References

[1] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, JETP Lett. **5** (1967) 24

[2] E. W. Kolb, M. S. Turner, *The Early Universe*, Addison-Wesley, New York, 1990

[3] V. A. Kuzmin, V. A. Rubakov, M. A. Shaposhnikov, *On anomalous electroweak baryon-number non-conservaton in the early universe*, Phys. Lett. **B 155** (1985) 36

[4] M. Fukugita, T. Yanagida, *Baryogenesis without grand unification*, Phys. Lett. **B 174** (1986) 45

[5] B. D. Fields, S. Sarkar, *Big-bang nucleosynthesis*, in *Review of Particle Physics*, Phys. Rev. **D 66** (2002) 010001

[6] WMAP Collaboration, D. N. Spergel et al., *First year Wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters*, Astrophys. J. Suppl. **148** (2003) 175

[7] J. Bernstein, *Kinetic Theory in the Expanding Universe*, Cambridge University Press, Cambridge, 1988

[8] S. Yu. Khlebnikov, M. E. Shaposhnikov, *The statistical theory of anomalous fermion number non-conservation* Nucl. Phys. **B 308** (1988) 885;
J. A. Harvey, M. S. Turner, *Cosmological baryon and lepton number in the presence of electroweak fermion-number violation*, Phys. Rev. **D 42** (1990) 3344

[9] W. Buchmüller, M. Plümacher, *Neutrino masses and the baryon asymmetry*, Int. J. Mod. Phys. **A 15** (2000) 5047

[10] V. A. Rubakov, M. E. Shaposhnikov, *Electroweak baryon number non-conservation in the early universe and in high energy collisions*, Phys. Usp. **39** (1996) 461

[11] G. D. Moore, *Do we understand the sphaleron rate?*, hep-ph/0009161;
D. Bödeker, *Non-equilibrium field theory*, hep-h/0011077

[12] W. Bernreuther, *CP violation and baryogenesis*, hep-ph/0205279

[13] M. Quiros, *Electroweak baryogenesis and the Higgs and stop masses*, Nucl. Phys. Proc. Suppl. **101** (2001) 401

[14] E. W. Kolb, S. Wolfram, *Baryon number generation in the early universe*, Nucl. Phys. **B 172** (1980) 224; Nucl. Phys. **B 195** (1982) 542 (E)

[15] W. Buchmüller, S. Fredenhagen, *Elements of baryogenesis*, hep-ph/0001098

[16] K. Hamaguchi, H. Murayama, T. Yanagida, *Leptogenesis from an $\tilde{N}$-dominated universe*, Phys. Rev. **D 65** (2002) 043512

[17] S. Davidson, A. Ibarra, *A lower bound on the right-handed neutrino mass from leptogenesis*, Phys. Lett. **B 535** (2002) 25

[18] W. Buchmüller, P. Di Bari, M. Plümacher, *Cosmic microwave background, matter-antimatter asymmetry and neutrino masses*, Nucl. Phys. **B 665** (2003) 445

[19] W. Buchmüller, P. Di Bari, M. Plümacher, *The neutrino mass window for baryogenesis*, Nucl. Phys. **B 665** (2003) 445

[20] S. Antusch, J. Kersten, M. Lindner, M. Ratz, *Running neutrino masses, mixings and CP phases: analytical results and phenomenological consequences*, Nucl. Phys. **B 674** (2003) 401

[21] G. F. Giudice, A. Notari, M. Raidal, A. Riotto, A. Strumia, *Towards a complete theory of thermal leptogenesis in the SM and MSSM*, hep-ph/0310123

[22] T. Hambye, Y. Lin, A. Notari, M. Papucci, A. Strumia, *Constraints on neutrino masses from leptogenesis models*, hep-ph/0312203

[23] W. Buchmüller, P. Di Bari, M. Plümacher, *Leptogenesis for pedestrians*, hep-ph/0401240

[24] A. Pilaftsis, T. E. J. Underwood, *Resonant leptogenesis*, hep-ph/0309342

[25] Z. Z. Xing, *Flavour mixing and CP violation of massive neutrinos*, hep-ph/0307350;
G. C. Branco, *Cosmology and CP violation*, hep-ph/0309215

[26] I. Masina, *Lepton flavour violation*, hep-ph/0210125

[27] W. Buchmüller, K. Hamaguchi, M. Ratz, *Gauge couplings at high temperature and the relic gravitino abundance*, hep-ph/0307181;
M. Fujii, M. Ibe, T. Yanagida, *Upper bound on gluino mass from thermal leptogenesis*, hep-ph/0310142

[28] I. Affleck, M. Dine, *A new mechanism for baryogenesis*, Nucl. Phys. **B 249** (1985) 361

[29] M. Dine, A. Kusenko, *The origin of the matter-antimatter asymmetry*, hep-ph/0303065

[30] K. Hamaguchi, *Cosmological baryon asymmetry and neutrinos: baryogenesis via leptogenesis in supersymmetric theories*, hep-ph/0212305

[31] M. Fujii, K. Hamaguchi, T. Yanagida, *Reheating temperature independence of cosmological baryon asymmetry in Affleck-Dine leptogenesis*, Phys. Rev. **D 63** (2001) 123513

[32] M. Fujii, K. Hamaguchi, *Nonthermal dark matter via Affleck-Dine baryogenesis and its detection possibility*, Phys. Rev. **D 66** (2002) 083501

[33] A. D. Dolgov, *Non-GUT baryogenesis*, Phys. Rep. **C222** (1992) 309

[34] G. Lazarides, Q. Shafi, *Origin of matter in the inflationary cosmology*, Phys. Lett. **B 258** (1991) 305

[35] M. Fujii, K. Hamaguchi, T. Yanagida, *Leptogenesis with almost degenerate Majorana neutrinos*, hep-ph/0202210

[36] S. Davidson, M. Losada, A. Riotto, *Baryogenesis at low reheating temperatures*, Phys. Rev. Lett. **84** (2000) 4284

[37] M. Li, X. Wang, B. Feng, X. Zhang, *Quintessence and spontaneous leptogenesis*, Phys. Rev. **D 65** (2002) 103511

[38] L. Bento, Z. Berezhiani, *Baryogenesis: the lepton leaking mechanism*, hep-ph/0111116

[39] M. Bastero-Gil, E. J. Copeland, J. Gray, A. Lukas, M. Plümacher, *Baryogenesis by brane-collisions*, hep-th/0201040

[40] B. Garbrecht, T. Prokopec, M. G. Schmidt, *Coherent baryogenesis*, Phys. Rev. Lett. **92** (2004) 061303-1

SCALAR FIELDS IN PARTICLE PHYSICS IN COSMOLOGY

Pierre Binetruy
Laboratoire de Physique Théorique,
Université Paris XI, 91405 Orsay cedex
Pierre.Binetruy@th.u-psud.fr

Abstract We review the role of scalar fields in theoretical models of high energy and cosmology. We discuss two general classes of light scalar fields: i) Goldstone bosons (and pseudo-Goldstone bosons) associated with the spontaneous breaking of symmetry, ii) moduli fields associated with the flat directions of the scalar potential in supersymmetric theorie. The latter are discussed in particular in the context of extra dimensions. Gravitational constraints are discussed. The potential role of a scalar fields in cosmology investigated: a quantitative analysis of their in path in the inflation scenario as well as in the understanding of dark energy is presented.

Keywords: Scalar fields,unitarity,triviality, naturalness, supersymmetry, sponta neous symmetry breaking,Goldstone bosons, moduli, flat directions, extra dimensions, gravitation, inflation.

I. Introduction

Although a fundamental scalar particle (i.e. a fundamental particle of spin 0) has yet to be discovered, scalar fields have nowadays become ubiquitous. They play a central role in the symmetry breaking sector of the Standard Model (Higgs particle). The vacuum energy stored in their vacuum leads to an exponential growth of the universe in inflation models. And they might be responsible for the late reacceleration of the universe.

In many cases, the scalar field is light i.e. its mass is much lighter than the fundamental scale of the theory (often by many orders of magnitude).

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 181–234.

For example, in the context of the Standard Model, the Higgs particle is believed to be light on the scale of electroweak symmetry breaking (see Figure 1).

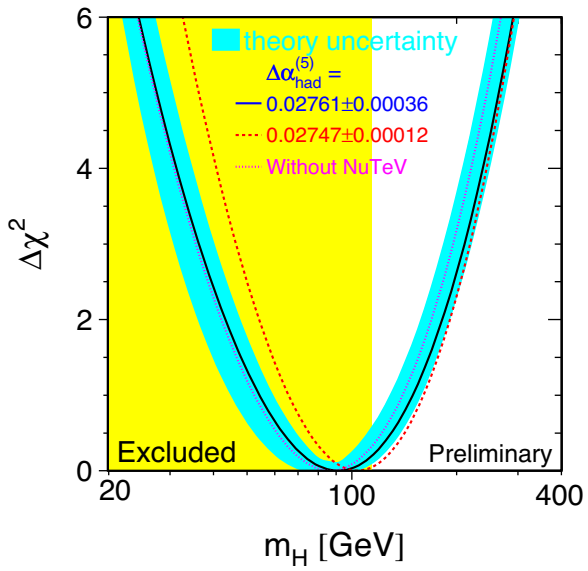

Figure 1. $\Delta\chi^2$ for the Higgs mass after LEP experiments (LEP Electroweak Working Group)

On the other hand, the Standard Model is believed to be the effective theory of a fundamental theory whose typical scale is much higher than the TeV scale. Indications in this direction come from the unification of gauge couplings or the lightness of neutrino masses (explained naturally by the seesaw mechanism which typically involves a large scale). In such contexts where the typical mass scale of the theory is very large, it remains to explain why the Higgs mass is much smaller. Indeed, as we will recall in Section II, scalar fields are special from the point of view of quantum corrections, which tend to destabilize light masses. This is the hierarchy problem.

There are basically two known ways of getting light scalars. One uses the spontaneous breaking of continuous global symmetries: this yields massless scalars known as Goldstone bosons (the Goldstone theorem is recalled in Section III). A small addition of an interaction which breaks

explicitly the symmetry turns them into pseudo-Goldstone bosons: their mass is then proportional to the amount of breaking.

The other one uses supersymmetry: in supersymmetric theories, the scalar potential often has flat directions that are valleys of vanishing potential energy that extend to infinity in field space. Such directions remain flat to all orders of perturbation theory and one has to resort to non-perturbative breaking of supersymmetry to lift the corresponding degeneracy. The fields corresponding to such flat directions are known as moduli fields. After breaking of supersymmetry they acquire a mass which is often small because of the non-perturbative nature of the breaking. What distinguishes moduli from Goldstone bosons is the fact that physics depends on the value of the moduli fields.

In what follows, we recall a few facts about naturalness and the problem of hierarchy in Section II: this leads us to the introduction of supersymmetry. We then discuss the physics of Golstone and pseudo-Goldstone bosons in Section III. We then return to supersymmetry in Section IV to consider the role of moduli fields, especially in the context of string theory and more generally of higher-dimensional theories where they appear extensively. Cosmological aspects are reviewed in Section V.

II. Naturalness and the problem of hierarchy

The central question that we will address in this section is the existence of quadratic divergences associated with the presence of a fundamental scalar field, such as the Higgs field in the Standard Model. Before dealing with this, we must have a short presentation of the notion of effective theory.

II-1 Effective theories

In the modern point of view, a given theory (e.g. the Standard Model) is always the effective theory of a more complete underlying theory, which adequately describes physics at a energy scale higher than a threshold M. This threshold is physical in the sense that the complete physical spectrum includes particles with a mass of order M.

The description in terms of an effective theory, restricted to the light states, is obviously valid only up to the scale M. The heavy fields (of mass M or larger) regulate the theory and therefore the scale M acts as a cut-off Λ on loop momenta.

In quantum field theory, the renormalisation procedure allows to deal with infinities, i.e. contributions that diverge when the cut-off is sent to infinity. However, the cut-offs that we consider here are physical and

thus cannot be sent to arbitrary values. There is then the possibility that the corrections due to the heavy fields (of mass M) destabilize the low energy theory. As we will see in the next subsection, this is indeed a possibility when we are working with fundamental scalars.

In some theories, we may infer some upper bound on the physical cut-off Λ (which we identify from now on with the scale of new physics M) from the value of the low-energy parameters. We will discuss briefly two standard methods used (unitarity and triviality) and illustrate them on the example of a complex scalar field.

More precisely, we consider, as in the Standard Model, a complex scalar field ϕ with Lagrangian

$$\begin{aligned} \mathcal{L} &= \partial^\mu \phi^\dagger \partial_\mu \phi - V\left(\phi^\dagger \phi\right) \\ V\left(\phi^\dagger \phi\right) &= -m^2 \phi^\dagger \phi + \frac{\lambda}{2}\left(\phi^\dagger \phi\right)^2 \end{aligned} \tag{II-1.1}$$

The minimization of this potential gives the background value $< \phi^\dagger \phi >= v^2$ with

$$v^2 \equiv m^2/\lambda \tag{II-1.2}$$

We thus parametrize ϕ as:

$$\phi = \begin{pmatrix} \varphi^+ \\ v + h + i\varphi^0 \end{pmatrix} \tag{II-1.3}$$

The fields φ^+ and φ^0 are massless fields (Goldstone bosons: see Section III) whereas the mass of the h field is

$$m_h^2 = 2m^2 = 2\lambda v^2 \tag{II-1.4}$$

This scalar field may have extra couplings to gauge fields or the top quark for example.

Unitarity[1, 2]

Unitarity of the S-matrix, which is a consequence of the conservation of probabilities at the quantum level, imposes some constraints on scattering cross sections, especially on their high-energy behavior. This is usually expressed in terms of partial-wave expansion: if $\mathcal{M}(s,\theta)$ is the amplitude for a $2 \to 2$ scattering process with center of mass energy $\sqrt{s}$ and diffusion angle θ, one defines the J-th partial wave as:

$$a_J(s) = \frac{1}{32\pi} \int d\cos\theta \; P_J(\cos\theta) \mathcal{M}(s,\theta) \tag{II-1.5}$$

where P_J is the J-th Legendre polynomial. The constraint coming from unitarity reads $\text{Im } a_J \geq |a_J|^2 = (\text{Re } a_J)^2 + (\text{Im } a_J)^2$, from which we obtain $(\text{Re } a_J)^2 \leq \text{Im } a_J\,(1 - \text{Im } a_J)$ Since the right-hand side of this equation is bounded by 1/4, it implies

$$|\text{Re } a_J| \leq \frac{1}{2} \tag{II-1.6}$$

Such limits were considered in the context of the Fermi model of weak interactions to introduce an intermediate vector boson. They may also be applied to the physics of the Standard Model. For example, the $J = 0$ tree level amplitude for $W_L^+ W_L^- \to Z_L Z_L$ ($W_L^\pm$, Z_L are the longitudinal components of $W^\pm$ and Z; using the equivalence theorem [1, 3], they can be identified respectively with $\varphi^\pm$ and φ^0) simply reads $a_0(s) = -G_{_F}\sqrt{2}m_h^2/(16\pi)$, in the limit $s \gg m_h^2$. The unitarity constraint (II-1.6) thus gives a constraint on the Higgs mass. It turns out that the most stringent constraint comes from the mixed zero-isospin channel $2W_L^+W_L^- + Z_L Z_L$ and reads, in terms of the electroweak breaking scale $v = \left(G_{_F}\sqrt{2}\right)^{-1/2}$,

$$m_h < \sqrt{\frac{16\pi}{5}}v = 780 \text{ GeV} \tag{II-1.7}$$

Triviality[4, 5]

In the renormalization group approach, the scalar self-coupling λ is turned into a running coupling $\lambda(\mu)$ varying with the momentum scale μ characteristic of the process considered. The study of one-loop radiative corrections allows to compute to lowest order (in λ) the evolution of $\lambda(\mu)$ with the scale μ, i.e. its so-called beta function:

$$\mu\frac{d\lambda}{d\mu} = \frac{3}{2\pi^2}\lambda^2 + \cdots \tag{II-1.8}$$

where we give here only the dominant one-loop contribution.

We see that the coupling $\lambda(\mu)$ is monotonically increasing. If we want that the theory described by the Lagrangian (II-1.1) makes sense all the way up to the scale Λ, we must impose that $\lambda(\mu) < \infty$ for scales $\mu < \Lambda$. If Λ is known, this imposes some bound on the value of λ at low energy, say $\lambda(v)$. For example, if we send Λ to infinity, this imposes $\lambda(v) = 0$. This is why a theory described by an action (II-1.1) which would be valid at *all* energy scales is known as *trivial*, i.e. is a free field theory in the infrared (low energy) regime. In practice, this only means that at some

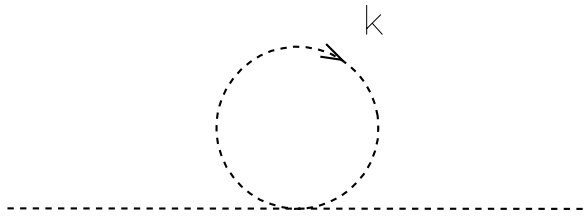

Figure 2. The scalar one-loop diagram giving rise to a quadratic divergence

scale Λ smaller than the scale $\Lambda_{\rm Landau}$ where the coupling would explode (known as the Landau pole, see below), some new physics appears.

The exact value of the Landau pole requires a non-perturbative computation since the running coupling explodes at this scale. Complete calculations show that it is not unreasonable to use the one-loop result (II-1.8) to obtain an order of magnitude for the Landau pole. Thus, solving for λ the differential equation (II-1.8),

$$\lambda^{-1}(\mu) = \lambda^{-1} - \frac{3}{2\pi^2} \ln \frac{\mu}{v} \tag{II-1.9}$$

where $\lambda \equiv \lambda(v)$, one obtains, using $\lambda^{-1}(\Lambda_{\rm Landau}) = 0$

$$\Lambda_{\rm Landau} \sim v \; e^{2\pi^2/(3\lambda)} \tag{II-1.10}$$

This is used to put an upper bound, a triviality bound, on the scale of new physics: $\Lambda < \Lambda_{\rm Landau}$. Since λ can be expressed in terms of m_h itself through (II-1.4), one may alternatively say that, for a given value of Λ, the Higgs mass is bounded by

$$m_h^2 < \frac{4\pi^2 v^2}{3 \ln(\Lambda/v)} \tag{II-1.11}$$

a decreasing function of Λ.

II-2 The concept of naturalness

The presence of fundamental scalar fields leads to the well-known problem [6–8] of quadratic divergences as soon as one introduces a finite cut-off Λ in the theory. Indeed a diagram of the type given in Fig. 2 generically contributes to the scalar mass-squared m^2

$$\delta m^2 = \lambda \int^{\Lambda} \frac{d^4k}{(2\pi)^4} \frac{1}{k^2} \sim \frac{\lambda}{16\pi^2} \int^{\Lambda} dk^2,$$

which is of order $\lambda\Lambda^2/16\pi^2$, .

Denoting by m_0 the bare mass (which, in this context, is the mass of the scalar field in the absence of underlying physics), we obtain at the one-loop level a scalar mass-squared

$$m^2 = m_0^2 + \alpha\lambda\frac{\Lambda^2}{16\pi^2}$$

where α is a positive or negative number of order one. Taking Λ as a fundamental mass unit,

$$\frac{m_0^2}{\Lambda^2} = \frac{m^2}{\Lambda^2} - \alpha\frac{\lambda}{16\pi^2} \tag{II-2.12}$$

Plugging typical numbers, say $m \sim 100$ GeV and $\Lambda \sim M_P \sim 10^{19}$ GeV, we see that m_0^2/Λ^2 must be adjusted to more than 30 orders of magnitude. This is to most people an intolerable fine tuning.

Let us be more precise in the case of the Standard Model. Then, the Higgs mass receives the following one-loop corrections:

$$\delta m_h^2 = \frac{3\Lambda^2}{8\pi^2 v^2}\left[\left(4m_t^2 - 2M_W^2 - 4M_Z^2 - m_h^2\right) + O\left(\log\frac{\Lambda}{\mu}\right)\right] \tag{II-2.13}$$

where we recognize the contribution of the diagram of Fig. 2, proportional to m_h^2, as well as the contribution of the top quark loop, proportional to m_t^2. The latter is leading in the case of a light Higgs. One may also note that the one-loop quadratically divergent contribution vanishes if we have the following relation between the masses:

$$4m_t^2 = 2M_W^2 + 4M_Z^2 + m_h^2 \tag{II-2.14}$$

This relation, known as the Veltman condition [9], is obviously not ensured to hold at higher orders.

One may define the amount f of fine tuning discussed above by

$$\frac{\delta m_h^2}{m_h^2} \equiv \frac{1}{f} \tag{II-2.15}$$

Indeed, if $\delta m_h^2 = 100 m_h^2$, then one needs to fine tune the Higgs bare mass m_0^2 to the per cent level in order to recover the right physical Higgs mass m_h^2. This amount of fine-tuning is represented on a plot (m_h,Λ) [10] in Fig. 3 for values of Λ smaller than 100 TeV. The region forbidden by the triviality bound discussed in the preceding subsection is also presented. One may note that, in the region corresponding to the Veltman condition (II-2.14), there is less need for fine tuning: this

region does not extend however to very large values of Λ because of the higher order contributions.

In any case, it is clear that fine tunings more severe than the per cent level are necessary as soon as the scale of new physics Λ is larger than 100 TeV.

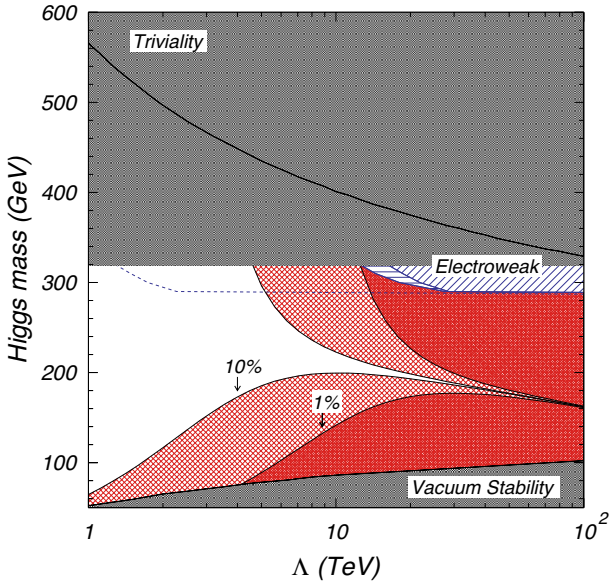

Figure 3. Plot in the $m_h - \Lambda$ plane showing the triviality (dark region at top) and stability (dark region at bottom) constraints, as well as the tuning contours. The darkly hatched region marked "1%" represents tunings of greater than 1 part in 100; the "10%" region means greater than 1 part in 10. The empty region has less than 1 part in 10 finetuning. [10]

Such a fine tuning goes against the prejudice that the observable properties of a theory (masses, charges,...) are stable under small variations of the fundamental parameters (the bare parameters). One talks of the *naturalness* of a theory to describe such a behaviour [11].

This does not mean that there cannot be small parameters in a natural theory. A parameter is *naturally* small if setting it to zero enhances the symmetries of the theory. Indeed, in such a case, the symmetry controls the quantum corrections to the parameter. The best known example is the mass m_e of the electron in quantum electrodynamics: in the limit $m_e \to 0$, the symmetry of the system is enhanced to include chiral symmetry, i.e. invariance under $\psi_e \to e^{i\alpha\gamma_5}\psi_e$ where ψ_e is the Dirac

spinor describing the electron. The presence of this symmetry imposes that the corrections to the electron mass are themselves proportional to m_e.

In the case of a complex scalar field with self-coupling λ the fundamental high-energy scale of the theory is given by its Landau pole Λ_{Landau} (II-1.10). But m^2/Λ is not naturally small because $m^2 = 0$ does not correspond to any symmetry enhancement at the quantum level (see [11] for a more precise statement).

II-3 Supersymmetry as a solution to the problem of naturalness

We have seen that:

i) setting the mass of a scalar field to zero does not enhance the symmetry.

ii) setting the mass of a fermion field to zero enhances the symmetry (chiral symmetry).

The idea is therefore to relate under a new symmetry a scalar field with a fermion field. This symmetry – supersymmetry – must be such that the masses of the scalar and of the fermion fields be equal. It is therefore related, in some sense to be defined, to the invariance under the Poincaré group since it connects representations of different spin. In such a scheme, the relation $m_s/\Lambda \ll 1$ is natural because $m_f/\Lambda \ll 1$ is natural and because the scalar mass m_s is related to the fermion mass m_f.

Technically, the contribution of fermions cancels the contribution of bosons to the quadratic divergence. We will check this result explicitly, on a model known as the Wess-Zumino model. This model might seem at first rather contrived. You will see the beauty of it in the lectures by D. Kazakov. For the time being, it will be an opportunity to familiarize ourselves with some typical supersymmetric interactions.

The model contains:

- a complex scalar field (2 degrees of freedom) : $\phi = (A + iB)/\sqrt{2}$,
- a fermion field described by a Majorana spinor ψ (two degrees of freedom) : $\psi^C = C\bar{\psi}^T = \psi$.

The Lagrangian decomposes into

i) a kinetic term

$$\mathcal{L}_{\rm kin} = \partial^\mu \phi^* \partial_\mu \phi + \frac{i}{2}\ \bar{\psi}\partial\!\!\!/\psi = \frac{1}{2}\ \partial^\mu A \partial_\mu A + \frac{1}{2}\ \partial^\mu B \partial_\mu B + \frac{i}{2}\bar{\psi}\partial\!\!\!/\psi \quad \text{(II-3.16)}$$

ii) an interaction term which is expressed in terms of a single function $W(\phi)$ known as the superpotential

$$\mathcal{L}_W = -\left|\frac{dW}{d\phi}\right|^2 - \frac{1}{2}\left(\frac{d^2W}{d\phi^2}\ \bar{\psi}_{_R}\psi_{_L} + \frac{d^2W^*}{d\phi^{*2}}\ \bar{\psi}_{_L}\psi_{_R}\right) \quad \text{(II-3.17)}$$

Explicitly we will take

$$W(\phi) = \frac{1}{2}m\phi^2 + \frac{1}{3}\lambda\phi^3 \quad \text{(II-3.18)}$$

$$\begin{aligned}
\mathcal{L}_W &= -\underbrace{|m\phi + \lambda\phi^2|^2}_{V(\phi)} - \frac{1}{2}\left[m\left(\bar{\psi}_{_R}\psi_{_L} + \bar{\psi}_{_L}\psi_{_R}\right)...\right. \\
&\quad \left. ... + 2\lambda\left(\phi\ \bar{\psi}_{_R}\psi_{_L} + \phi^*\ \bar{\psi}_{_L}\psi_{_R}\right)\right] \\
&= -\frac{1}{2}m^2\left(A^2+B^2\right) - \frac{m\lambda}{\sqrt{2}}\ A\left(A^2+B^2\right) - \frac{\lambda^2}{4}\ \left(A^2+B^2\right)^2 \\
&\quad -\frac{1}{2}m\ \bar{\psi}\psi - \frac{\lambda}{\sqrt{2}}\ \bar{\psi}\left(A - iB\ \gamma_5\right)\psi \qquad \text{(II-3.19)}
\end{aligned}$$

We see that the form (II-3.17) ensures that scalar bosons (A, B) and spinor fermions (ψ) have the same mass (m).

The self-energy diagrams for the scalar field A which contribute at one loop to the quadratic divergence are given in Fig. 4.

Exercise: Show that the quadratic divergences cancel among the 3 diagrams of Figure 4 (assume that the integrals are properly regularized and thus that there exists a regularization procedure that respects supersymmetry).

One may show that the cancellation is even larger and that logarithmic divergences are only present in wave function renormalisation. There are no mass counterterms and, more generally, the parameters of the superpotential are not renormalized (no finite nor infinite quantum corrections). This is an example of the famous non-renormalisation theorems that have made the success of supersymmetry.

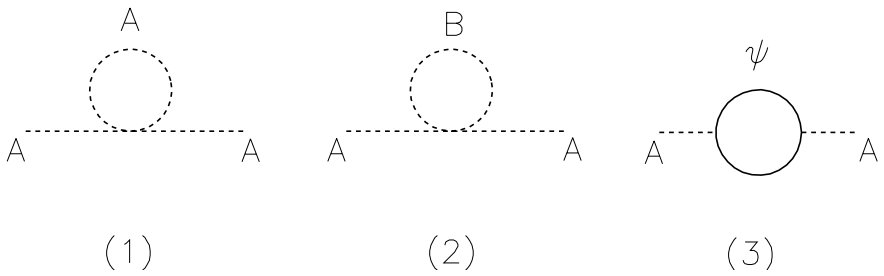

Figure 4. Quadratically divergent self-energy diagrams for the A scalar field

III. Spontaneous breaking of symmetry

We discuss in this section the appearance of Goldstone bosons in the case of the spontaneous breaking of a global continuous symmetry. In presence of small explicit breaking terms, this leads to naturally light scalars, the pseudo-Goldstone bosons (PGB). We review the scenario of the little Higgs where the Higgs is such a PGB.

III-1 Example of a global symmetry

We start with the simplest example of a global abelian internal symmetry and consider a complex scalar field $\phi(x)$ with a Lagrangian density

$$\mathcal{L} = \partial^\mu \phi^\dagger \partial_\mu \phi - V(\phi^\dagger \phi) \qquad \text{(III-1.20)}$$

This has a global phase invariance $\phi(x) \to \phi'(x) = e^{-i\theta}\phi(x)$. We take, as in the last Section, the standard Mexican hat potential of Figure 5

$$V(\phi^\dagger \phi) = -m^2\ \phi^\dagger \phi + \frac{\lambda}{2}(\phi^\dagger \phi)^2 \qquad \text{(III-1.21)}$$

where $\lambda > 0$ to avoid instability at large values of the field.

The ground state corresponds to a non-vanishing value for the scalar field: $\phi_0^\dagger \phi_0 = m^2/\lambda$. The equation of motion reads

$$(\Box - m^2)\phi = j(\phi) \qquad \text{(III-1.22)}$$

where $j(\phi) \equiv -\lambda\phi(\phi^\dagger\phi)$ describes the self-interaction of the field ϕ. In the limit of vanishing self-interaction ($j \to 0$), the equation is not the standard Klein-Gordon equation because of the sign in front of m^2 and we thus do not know how to quantize the theory. Writing $-m^2 = (im)^2$, we may identify the origin of this sign as the instability of the system

for small values of the field, i.e. around $\phi = 0$ (see Figure 5). In other words, we should evaluate the scalar field ϕ around one of its stable values defined by $\phi_0^\dagger \phi_0 = m^2/\lambda$.

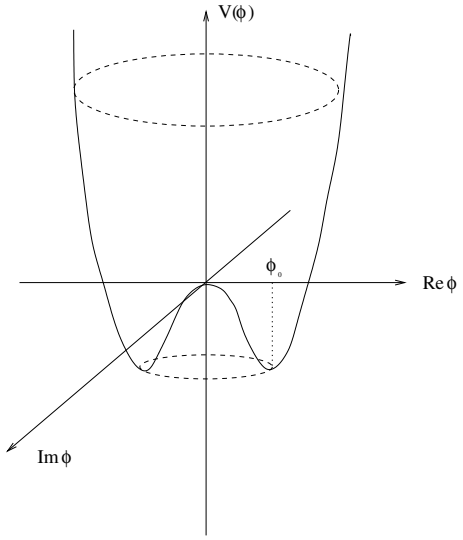

Figure 5.

We note that the potential has a rotation symmetry corresponding to $\phi \to e^{-i\theta}\phi$. On the other hand, the ground state of the system (which corresponds to the choice of one point on the set of degenerate minima) necessarily breaks this symmetry. We are in a situation of spontaneous symmetry breaking:

- the Lagrangian (the Hamiltonian) is invariant under the symmetry
- the ground state (the vacuum) is not left invariant by the symmetry.

Now, following our remark above, let us express the scalar field around its ground value, say $e^{i\theta_0}\sqrt{m^2/\lambda}$: we write

$$\phi(x) = e^{i(\gamma(x)+\theta_0)} \left[\rho(x) + \sqrt{\frac{m^2}{\lambda}}\right] \tag{III-1.23}$$

Then

$$V(\phi^\dagger\phi) = V([\rho + \sqrt{m^2/\lambda}]^2) = -\frac{m^4}{2\lambda} + 2m^2\rho^2 + 2\rho^3\sqrt{\lambda m^2} + \frac{\lambda}{2}\rho^4 \tag{III-1.24}$$

We conclude that the field ρ has mass squared $2m^2$ (note the normalization of the kinetic term $\partial^\mu\rho\partial_\mu\rho$) whereas γ is massless. This is a reflection of the spontaneous breaking of the symmetry. Indeed, it is a general theorem due to Goldstone [12] that *to every continuous global symmetry spontaneously broken, there corresponds a massless particle.* In the case of internal symmetries such as considered here, it is a boson called Goldstone boson.

Let us stress one important property of a Goldstone boson. If we write the global phase transformation $\phi(x) \to \phi'(x) = e^{-i\theta}\phi(x)$ in the parametrization (III-1.23), we obtain

$$\rho'(x) = \rho(x) \ , \quad \gamma'(x) = \gamma(x) - \theta \tag{III-1.25}$$

The Goldstone transformation is nonlinear and is characterized by a constant term [1]. This is why a Goldstone boson associated with an internal symmetry only has derivative interactions.

III-2 Goldstone theorem

We now consider the general case of a group G of continuous global transformations. The scalar fields are taken to be real and transform under G as:

$$\phi'(x) = e^{-i\alpha^a t^a} \ \phi(x) \tag{III-2.26}$$

where the t^a are the generators of the group, in a n-dimensional representation[2]. Infinitesimally

$$\phi'_m(x) = \phi_m(x) - i\alpha^a t^a_{mn}\phi_n(x) \tag{III-2.27}$$

where $m = 1, \cdots, n$.

Since the potential $V(\phi)$ is invariant, we have

$$\frac{\partial V}{\partial \phi_m} \ \alpha^a \ t^a_{mn} \ \phi_n = 0 \tag{III-2.28}$$

from which we obtain, by differentiating with respect to ϕ_ℓ,

$$\frac{\partial^2 V}{\partial\phi_m\partial\phi_\ell} \ \alpha^a \ t^a_{mn} \ \phi_n + \frac{\partial V}{\partial\phi_m} \ \alpha^a \ t^a_{m\ell} = 0 \tag{III-2.29}$$

At the minimum ϕ_0 of the potential, we thus have

$$M^2_{\ell m} \ \alpha^a \ t^a_{mn} \ \phi_{0n} = 0 \tag{III-2.30}$$

[1] When the symmetry is local ($\theta(x)$), this allows to gauge away the corresponding degree of freedom.

[2] The t^a are taken to be Hermitian, $t^{a\dagger} = t^a$. Since ϕ is real, $(it^a)^* = it^a$. Hence $t^{a*} = -t^a$ and $t^{aT} = -t^a$: the generators are antisymmetric.

where

$$M^2_{\ell m} = \left.\frac{\partial^2 V}{\partial\phi_\ell \partial\phi_m}\right|_{\phi_0} \qquad \text{(III-2.31)}$$

is the scalar squared mass matrix. In the case of spontaneous breaking, the vacuum ϕ_0 is not invariant and thus $\alpha^a t^a_{mn}\phi_{0n}$ is a non-zero eigenvector of the squared mass matrix with vanishing eigenvalue: it corresponds to a Goldstone boson.

It is possible to separate the generators t^a ($a = 1 \cdots \dim G$) into:

• generators t^i which leave the vacuum invariant: $t^i\phi_0 = 0$. These generators correspond to the residual symmetry of gauge group H and $i = 1 \cdots \dim H$.

• generators θ^α which do not leave the vacuum invariant: $\theta^\alpha\phi_0 \neq 0$ ($\alpha = 1, \cdots, \dim G - \dim H$).

Goldstone bosons are a mixed blessing: as massless particles they mediate infinite range forces. Since there are no known long range forces besides gravity and electromagnetism this would mean that a Goldstone boson has to be extremely weakly coupled to ordinary matter.

Otherwise there are two ways out. If the symmetry is local, the Goldstone boson does not appear in the physical spectrum. It provides instead the longitudinal degree of freedom to the gauge boson which becomes massive through the Higgs mechanism. Otherwise, a small term *explicitly* breaking the symmetry generates a mass for the Goldstone boson, which becomes what is called a *pseudo-Goldstone boson.*

III-3 An example of Goldstone boson: the dilaton field

We illustrate the notion of Goldstone boson on the example of a spacetime symmetry[3] which will play a special role in the discussion of the cosmological constant problem: dilatation or scale invariance.

Scale invariance (and its violations as described by the renormalisation group approach) plays an important role in some of the applications that follow. Scale invariance is not observed in our universe. It could be that this global symmetry is spontaneously broken, in which case we expect a Goldstone boson. This is the dilaton field.

A dilatation or scaling transformation is a spacetime transformation of the form

$$x \to x' = e^{-\alpha}x \qquad \text{(III-3.32)}$$

[3]Beware, so far our discussion of spontaneous breaking has only considered internal symmetries, i.e. symmetries that leave spacetime invariant.

It acts linearly on the fields:

$$\Phi(x) \to \Phi'(x') = e^{\alpha d}\Phi(x) \tag{III-3.33}$$

which we may write by keeping spacetime fixed

$$\Phi(x) \to \Phi'(x) = e^{\alpha d}\Phi(e^{\alpha}x) \tag{III-3.34}$$

The number d is characteristic of the field Φ and is called its scaling dimension. Infinitesimally,

$$\delta\Phi(x) = \alpha\left(d + x^{\mu}\partial_{\mu}\right)\Phi(x) \tag{III-3.35}$$

At the classical level, the scaling dimension coincides with the canonical dimension: $d = 1$ for the scalar fields, $3/2$ for spin $\frac{1}{2}$ or $\frac{3}{2}$ fermions.

We may consider as an example the following action involving a scalar field $\phi(x)$ and a Dirac spinor field $\Psi(x)$:

$$\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_1 = \int d^4x\mathcal{L}_0(x) + \int d^4x\mathcal{L}_1(x) \tag{III-3.36}$$

$$\mathcal{L}_0 = \frac{1}{2}\partial^{\mu}\phi\partial_{\mu}\phi + \frac{i}{2}\bar{\Psi}\gamma^{\mu}\partial_{\mu}\Psi - \lambda_{_Y}\phi\bar{\Psi}\Psi - \frac{\lambda_0}{4!}\phi^4, \tag{III-3.37}$$

$$\mathcal{L}_1 = -\frac{1}{2}m_0^2\phi^2. \tag{III-3.38}$$

Under the dilatation transformation (III-3.35), we have $\delta\mathcal{L}_0 = \alpha\left(4 + x^{\mu}\partial_{\mu}\right)\mathcal{L}_0$ and thus, by integration by parts,

$$\delta\mathcal{S}_0 = \alpha\int d^4x\partial_{\mu}\left(x^{\mu}\mathcal{L}_0\right) \tag{III-3.39}$$

Thus $\mathcal{S}_0$ is invariant by dilatation. But not $\mathcal{S}_1$ and

$$\delta\mathcal{S} = \alpha\int d^4x\ \Delta \quad \text{with} \quad \Delta = m_0^2\phi^2 \tag{III-3.40}$$

It follows that the dilatation current, or scaling current, D_{μ} is not conserved.

Using standard Noether formulas, one obtains from (III-3.32) and (III-3.33) the general explicit form for the dilatation current

$$D_{\mu} = x^{\rho}T_{\rho\mu} + \frac{\delta\mathcal{L}}{\delta\left(\partial^{\mu}\Phi(x)\right)}\ d\Phi(x) \tag{III-3.41}$$

where $T_{\mu\nu}$ is the canonical energy-momentum tensor: [4]

$$T_{\mu\nu} = \frac{\delta\mathcal{L}}{\delta\left(\partial^{\mu}\Phi(x)\right)}\partial_{\nu}\Phi - g_{\mu\nu}\mathcal{L} \tag{III-3.42}$$

[4]Under some conditions, it is actually possible to define [13] (without affecting the construction of the Lorentz generators P_{μ} and $M_{\mu\nu}$) a symmetric energy-momentum tensor $\Theta_{\mu\nu}$ such that $D_{\mu} = x^{\rho}\Theta_{\rho\mu}$

It is possible to modify the Lagrangian in order to break dilatation symmetry spontaneously rather than explicitly, as with a mass term. One must make the Lagrangian invariant: only the fundamental state breaks the invariance. According to the Goldstone theorem, a massless boson appears in the spectrum. This Goldstone boson associated with dilatations is called dilaton. We denote it by $\sigma(x)$.

In order to construct an invariant Lagrangian describing the interactions of the dilaton with other fields, one may use a method which is inspired from the construction of chiral Lagrangians describing the pion interactions (the pion is the Goldstone boson associated with chiral symmetry breaking).

We consider a scalar field $\Phi(x)$ which we write:

$$\Phi(x) = f\ e^{\sigma(x)/f} \tag{III-3.43}$$

The field $\sigma(x)$ corresponds to fluctuations of $\Phi(x)$ around its vacuum value f, which determines the scale of spontaneous breaking of dilatation symmetry. Since under dilatations, the scalar field $\Phi(x)$ transforms as (III-3.34), the field $\sigma(x)$ transforms as:

$$\sigma(x) \to \sigma'(x) = \sigma\left(e^{\alpha}x\right) + \alpha f \tag{III-3.44}$$

or infinitesimally

$$\delta\sigma(x) = \alpha\left(f + x^{\mu}\partial_{\mu}\sigma\right) \tag{III-3.45}$$

We note that the symmetry is realized non-linearly: as we have seen in subsection III-1, the constant term is a sign that $\sigma(x)$ is a Goldstone boson.

One may use factors $e^{\sigma(x)/f}$ to make dilatation-breaking terms invariant. Thus $m_0^2\phi^2$ is replaced by $m_0^2\phi^2 e^{2\sigma/f}$ since $\delta\left(\phi^2 e^{2\sigma/f}\right) = (4 + x^{\mu}\partial_{\mu})\left(\phi^2 e^{2\sigma/f}\right)$. One also introduces a kinetic term for $\sigma(x)$:

$$\mathcal{L}_{kin}^{(\sigma)} = \frac{1}{2}\partial^{\mu}\Phi\partial_{\mu}\Phi = \frac{1}{2}e^{2\sigma/f}\partial^{\mu}\sigma\partial_{\mu}\sigma \tag{III-3.46}$$

Then the action corresponding to the Lagrangian

$$\mathcal{L} = \mathcal{L}_0 + \frac{1}{2}e^{2\sigma/f}\partial^{\mu}\sigma\partial_{\mu}\sigma - \frac{1}{2}m_0^2\phi^2 e^{2\sigma/f} \tag{III-3.47}$$

with $\mathcal{L}_0$, given in (III-3.36), is dilatation invariant. Under these conditions, the dilatation current is conserved:

$$\partial^{\mu}D_{\mu} = 0 \tag{III-3.48}$$

Note that the couplings to standard matter are of order $1/f$. Since the massless dilaton exchange should not generate an observable long range force, f has to be very large.

III-4 The Higgs as a pseudo-Goldstone boson: the little Higgs scenario

Could the Higgs be a Goldstone boson of an internal symmetry? Problems immediately arise because of the vanishing mass and the non-derivative couplings which are the trademark of a Goldstone boson. The solution is to use explicit *collective* breaking :

• explicit breaking

$$\mathcal{L} = \mathcal{L}_0 + \epsilon\ \mathcal{L}_1 \tag{III-4.49}$$

where the breaking term ϵ is much smaller than 1. As we have seen earlier, we expect a pseudo-Goldstone. Typically, if its mass appears at one loop, we have

$$\delta m^2 \sim \frac{\epsilon^2}{16\pi^2}\ \Lambda^2 \tag{III-4.50}$$

For $\Lambda \sim 10$ TeV, ϵ must be small.

We also expect non-derivative couplings of order ϵ or ϵ^2. But the couplings of the Higgs are of order 1, e.g. $\lambda_{\rm top}$.

• collective breaking

$$\mathcal{L} = \mathcal{L}_0 + \epsilon_1\ \mathcal{L}_1 + \epsilon_2\ \mathcal{L}_2 \tag{III-4.51}$$

One considers the following situation. In the limit $\epsilon_1 \to 0$, $\mathcal{L}_0 + \epsilon_2\ \mathcal{L}_2$ has a global symmetry which keeps the Higgs massless. Similarly, in the limit $\epsilon_2 \to 0$, $\mathcal{L}_0 + \epsilon_1\ \mathcal{L}_1$ has a global symmetry which keeps the Higgs massless. One thus expects

$$\delta m^2 \sim \frac{\epsilon_1^2}{16\pi^2}\frac{\epsilon_2^2}{16\pi^2}\ \Lambda^2 \tag{III-4.52}$$

but non-derivative couplings of order ϵ_1 or ϵ_2.

How to compensate the quadratic divergences? In the model that we will describe below, the top loop quadratic divergence is cancelled by the fermion loops given in the second row of Figure 6.

The new colored fermions χ_L and χ_R have the following couplings to the Higgs:

$$\lambda_t f \left(1 - \frac{h^* h}{2f^2}\right) \bar{\chi}_R \chi_L + \text{h.c.} \tag{III-4.53}$$

Hence it is necessary to introduce new particles at a mass scale f of the order of a TeV. Similarly to cancel gauge and Higgs loops (see Figure 7).

Hence we have three mass scales in the theory:

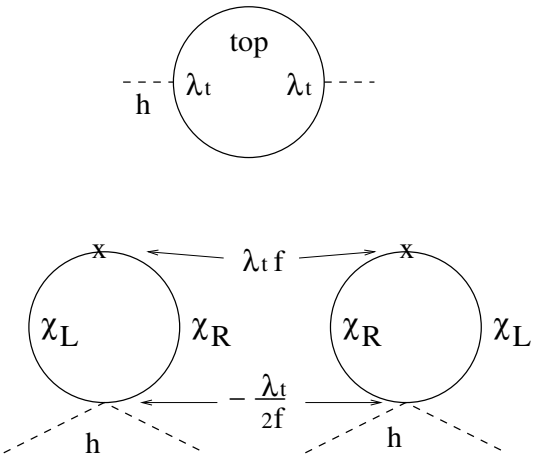

Figure 6.

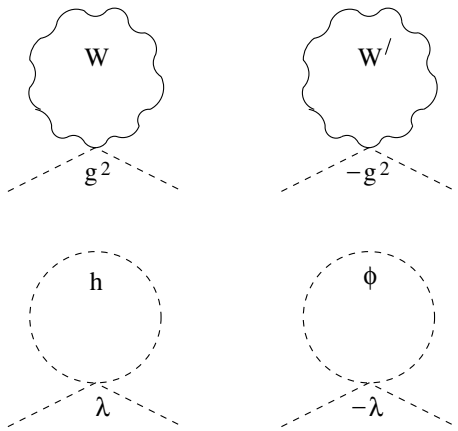

Figure 7.

• $v = 250$ GeV where we have the scalar degrees of freedom of the Standard Model

• $f \sim 1$ TeV where there are new colored fermions related to the top quark, new gauge bosons and new scalars

• $\Lambda \sim 10$ TeV where the effective theory should be replaced by an underlying fundamental theory still to be determined.

In the little Higgs model of Arkani-Hamed, Cohen, Katz and Nelson [14], a global $SU(5)_{\rm global}$ symmetry of dimension 24 is broken into a global $SO(5)_{\rm global}$ of dimension 10. This leaves us with 14 Goldstone bosons.

For example, a symmetric tensor field $\Sigma_{ij} = \Sigma_{ji}$, $i, j = 1 \cdots 5$ transforms under $SU(5)$ as $\Sigma \to U\Sigma U^T$, where U is a matrix of $SU(5)$. If this field has a vacuum expectation value,

$$\langle \Sigma \rangle = f\delta_{ij} \tag{III-4.54}$$

the residual symmetry is given by matrices U such that $UU^T = \mathbf{1}$ i.e. orthogonal matrices.

One may parametrize the Goldstone bosons as (similar to (III-1.23))

$$\Sigma(x) = e^{i\Pi^a(x)T^a} \langle \Sigma \rangle e^{-i\Pi^a(x)T^a} \tag{III-4.55}$$

The local gauge symmetry consists of two copies of $SU(2) \times U(1)$ (see the lecture on deconstruction by S. Pokorski for motivations) $(SU(2)_1 \times U(1)_1 \times SU(2)_2 \times U(1)_2)_{\rm local}$ broken into $(SU(2) \times U(1))_{\rm local}$ (the symmetry of the Standard Model!). Four Goldstone bosons are eaten up in the process. The Higgs H is found among the 10 remaining Goldstone bosons:

$$\Pi^a T^a \equiv \begin{pmatrix} 0 & H^\dagger & \Phi^\dagger \\ H & 0 & H^* \\ \Phi & H^T & 0 \end{pmatrix} \tag{III-4.56}$$

If one shuts off the gauge interactions of $SU(2)_2 \times U(1)_2$, a global $SU(3)_2$ protects the Higgs from getting a mass. Hence the gauge couplings g_1, g_1' (of $SU(2)_1$ and $U(1)_1$), g_2, g_2' (of $SU(2)_2$ and $U(1)_2$) play the roles of the small parameters ϵ_1, ϵ_2 in collective breaking.

Figures 8 and 9 give some constraints on the parameters of the theory: $c = g/g_2$ and $c' = g'/g_2'$, with g (resp. g') coupling of $SU(2)$ (resp. $U(1)$) [15].

IV. Supersymmetry and moduli fields

In this Section, we explain the concept of modulus field in the framework of supersymmetric theories and we present the most common moduli fields encountered in the context of higher-dimensional field theories and in string theory.

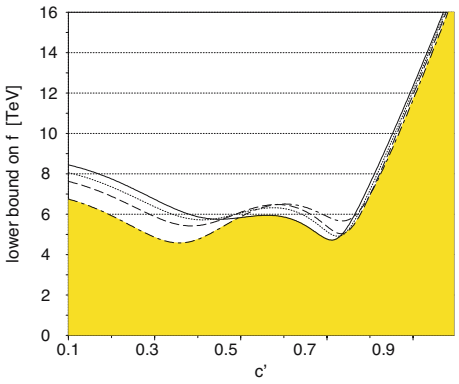

Figure 8. Lower bound on f in TeV versus $c' = g'/g'_2$ [15].

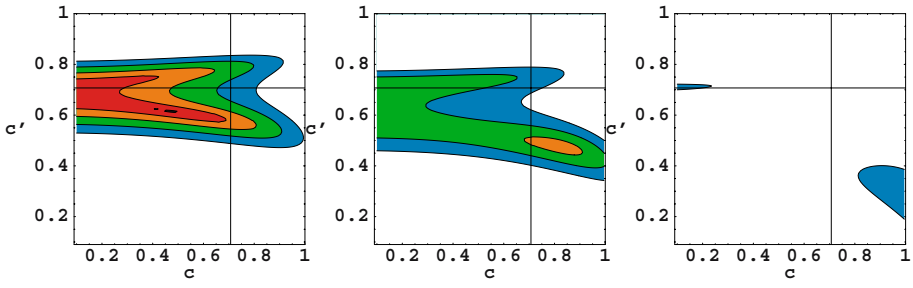

Figure 9. Limits in the (c, c') plane respectively for $Y_1 = 3Y/5$, $Y_1 = 4Y/5$ and $Y_1 = Y$, and $f < 2$ TeV (red), $f < 23$ TeV (orange), $f < 4$ TeV (green), $f < 5$ TeV (blue) [15].

IV-1 Spontaneous supersymmetry breaking, flat directions and moduli

We first complete our basic knowledge of supersymmetry. The Lagrangian of the Wess and Zumino model presented in subsection III-1 may be written as

$$\begin{aligned}\mathcal{L} &= \partial^\mu \phi^* \partial_\mu \phi + \frac{1}{2}\overline{\Psi} i\gamma^\mu \partial_\mu \Psi \\ &\quad -\frac{1}{2}\left[\frac{d^2W}{d\phi^2}\overline{\Psi}_R \Psi_L + \frac{d^2W^*}{d\phi^{*2}}\overline{\Psi}_L \Psi_R\right]\end{aligned}$$

$$+\left[F^*F + F\frac{dW}{d\phi} + F^*\frac{dW}{d\phi^*}\right] \qquad \text{(IV-1.57)}$$

Indeed, solving for the auxiliary field F yields

$$F = -\frac{dW}{d\phi^*} \qquad \text{(IV-1.58)}$$

and the standard potential

$$V = \left|\frac{dW}{d\phi}\right|^2 = F^*F \qquad \text{(IV-1.59)}$$

The introduction of the auxiliary field $F = (F_1(x) + iF_2(x))/\sqrt{2}$ simplifies the discussion of spontaneous supersymmetry breaking. Indeed, the Lagrangian (IV-1.56) is invariant under the infinitesimal supersymmetry transformations:

$$\begin{aligned}
&\delta_S A = \bar{\varepsilon}\Psi \quad , \quad \delta_S B = i\bar{\varepsilon}\gamma_5\Psi \\
&\delta_S \Psi = \left[-i\gamma^\mu\partial_\mu\left(A + iB\gamma_5\right) + F_1 - iF_2\gamma_5\right]\varepsilon \quad , \\
&\delta_S F_1 = -i\bar{\varepsilon}\gamma^\mu\partial_\mu\Psi \quad , \quad \delta_S F_2 = -\bar{\varepsilon}\gamma_5\gamma^\mu\partial_\mu\Psi
\end{aligned} \qquad \text{(IV-1.60)}$$

where ε is the spinor (supersymmetry transforms a scalar into a spinor) parameter of the supersymmetry transformation.

If, in the ground state, the auxiliary field takes a non-zero value, for example $< F_1 > \neq 0$, then

$$\delta\Psi = < F_1 > \varepsilon + \cdots \qquad \text{(IV-1.61)}$$

This constant term is typical of a Goldstone particle (cf. (III-1.25)). Indeed, in this case, it is an indication that supersymmetry is spontaneously broken: the Goldstone field is the spinor Ψ(the spinor nature of the Goldstone field is related to the spinor nature of the parameter of the transformation); one often refers to it as the goldstino.

We then conclude from (IV-1.59) that supersymmetry is spontaneously broken if (and only if) the vacuum energy is non-vanishing. The result is very general in global supersymmetry and can be derived from the supersymmetry algebra[5].

All this can easily be extended to n such supermultiplets with scalar components (ϕ_i, ψ_i, $i = 1 \cdots n$) and a superpotential $W(\phi_i)$, analytic in the fields ϕ_i. The potential reads:

$$V(\phi_i) = \sum_i \left|\frac{\partial W}{\partial \phi_i}\right|^2 = \sum_i |F_i|^2 \qquad \text{(IV-1.62)}$$

[5]See D. Kazakov lectures.

We are now ready to present the notions of flat directions of the scalar potential and associated modulus field. Let us take as an example

$$W(\chi,\phi)=\frac{1}{2}\mu\chi\phi^2+\frac{1}{3}\lambda\phi^3 \qquad \text{(IV-1.63)}$$

The scalar potential reads

$$V(\chi,\phi)=\frac{1}{4}|\mu|^2|\phi|^4+|\phi|^2|\mu\chi+\lambda\phi|^2 \qquad \text{(IV-1.64)}$$

and the ground state lies at $\langle\phi\rangle=0$, whereas the value of χ remains undetermined. This is an example of a flat direction of the scalar potential, i.e. a direction in field space, extending to infinite values, and along which $V=0$.

Flat directions are frequent in supersymmetric theories and they play a specific role. Indeed, the non-renormalization theorems alluded to at the end of Section II-3 allow to prove that flat directions remain flat to all orders of perturbation theory (since the parameters of the potential are not renormalized). Hence the scalar fields corresponding to these directions remain massless. Such fields are called moduli. Only non-perturbative effects can lift the degeneracy associated to these flat directions, and give a smal mass to these moduli fields.

Let us stress an important difference between the Goldstone bosons discussed in the preceding section and moduli fields. Whereas physical results do not depend on the actual value of a Goldstone field, they do depend on the value of moduli. For example, in our example, noticing that, along the flat direction

$$\langle\frac{\partial^2 W}{\partial\phi^2}\rangle=\mu\langle\chi\rangle \quad \langle\frac{\partial^2 W}{\partial\chi^2}\rangle=0 \quad \langle\frac{\partial^2 W}{\partial\chi\partial\phi}\rangle=0 \qquad \text{(IV-1.65)}$$

we conclude that the value of the modulus field χ along the flat direction determines the mass of the ϕ field (and its fermion partner).

To be complete, we also introduce gauge interactions under the form of abelian symmetries labelled by a (the corresponding charge of field ϕ_i is q_{ai}). The complete scalar potential then takes the form

$$\begin{aligned} V &= \sum_i |F_i|^2+\frac{1}{2}\sum_a D_a^2 \\ &= \sum_i\left|\frac{\partial W}{\partial\phi_i}\right|^2+\frac{1}{2}\sum_a\left(\sum_i q_{ai}\phi_i^*\phi_i-\xi_a\right)^2 \end{aligned} \qquad \text{(IV-1.66)}$$

where ξ_a is a real parameter called the Fayet-Iliopoulos coupling. The real field D_a is the auxiliary field of the vector supermultiplet describing

the abelian gauge field A^a_μ and its supersymmetric partner the gaugino (a Majorana fermion). A non-vanishing D_a is also a sign of spontaneous supersymmetry breaking (the golstino being then the corresponding gaugino field).

Moduli fields are often encountered in higher-dimensional field theories and in string theories. We will discuss in the next subsections some of the most commonly used examples of moduli fields. Since the gravitational couplings of these fields play an important role, we close this introduction by reviewing the coupling of chiral supermultiplets $\Phi^i = (\phi^i, \Psi^i, F^i)$ to supergravity (i.e. local supersymmetry)[6] . This coupling is described by three basic functions:

• the superpotential which is an analytic function of the scalar fields $W(\phi^i)$: it determines the self-interactions of the scalar fields as well as their Yukawa interactions to fermions.

• the Kähler potential $K(\phi^i, \bar{\phi^j})$ determines in particular the kinetic term of scalar fields:

$$\mathcal{L}_{\rm kin} = g_{i\bar{j}}(\phi^i, \bar{\phi^j})\ \partial^\mu \phi^i \partial_\mu \bar{\phi^j} \tag{IV-1.67}$$

where the so-called Kähler metric $g_{i\bar{j}}(\phi^i \bar{\phi^j})$ is defined as

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial \phi^i \partial \bar{\phi^j}} \tag{IV-1.68}$$

This is interpreted in a geometric way: the scalar fields ϕ^i, $\bar{\phi^j}$ parametrize a complex manifold and the Kähler metric is the metric on this manifold. One speaks of a flat Kähler metric if $g_{i\bar{j}} = \delta_{ij}$. If the metric is non-flat, the non-normalized kinetic term is obviously not renormalizable. Non-renormalizable interaction terms are also generated besides the kinetic term, for example a 4-fermion interaction: they are proportional to derivatives of the Kähler metric.

• a kinetic function $f(\phi^i)$ for the gauge fields which determines the gauge field kinetic term:

$$\mathcal{L}_{\rm kin} = -\frac{1}{4}{\rm Re} f_{ab}(\phi^i)\ F^{a\mu\nu} F^b_{\mu\nu} - \frac{1}{4}{\rm Im} f_{ab}(\phi^i)\ F^{a\mu\nu} \tilde{F}^b_{\mu\nu} \tag{IV-1.69}$$

The scalar potential then reads[7]

$$V = e^{K/m_P^2} \left[D_i W g^{i\bar{j}} D_{\bar{j}} \bar{W} - 3\frac{|W|^2}{m_P^2} \right] + \frac{g^2}{2}{\rm Re} f^{-1}_{ab} D^a D^b \tag{IV-1.70}$$

[6]Their complex conjugates are written $\bar{\Phi^{\bar{\imath}}} = (\bar{\phi^{\bar{\imath}}}, \bar{\Psi^{\bar{\imath}}}, \bar{F^{\bar{\imath}}})$.

[7]m_P is the reduced Planck mass scale defined from Newton's constant G_N: $m_P = \sqrt{\hbar c/(8\pi G_N)} = 2.4 \times 10^{18}$ GeV/c^2

where

$$D_i W = \frac{\partial W}{\partial \phi^i} + \frac{1}{m_P^2}\frac{\partial K}{\partial \phi^i} W \qquad \text{(IV-1.71)}$$

and

$$D^a = \frac{\partial K}{\partial \phi^i}(t^a)^i{}_j \phi^j \qquad \text{(IV-1.72)}$$

The inclusion of a Fayet-Iliopoulos term, in the case of a $U(1)$ symmetry, requires special care in supergravity [16].

IV-2 Moduli fields from the compactification of higher-dimensional theories

We start our discussion of moduli fields with theories with more than 3 spatial dimensions. Our illustrative example will be the historical model of Kaluza and Klein [17, 18] which dates back to the 20's.

T. Kaluza and O. Klein [17, 18] proposed to unify geometrically electromagnetism with gravitation by introducing the electromagnetic field as a component of the metric of a 5-dimensional spacetime. To be slightly more general, let us consider a theory of gravity in $(D \equiv d+1)$-dimensional spacetime, described by the coordinates x^M, $M = 0, \cdots, d$. The metric is g_{MN} and the signature $(+ - - - ...)$.

We single out the spatial coordinate $x^d \equiv y$ and assume that the corresponding dimension is compact of size $L = 2\pi R$; the other d dimensions corresponding to x^μ, $\mu = 0, \cdots, d-1$ are noncompact. From the point of view of d dimensions, $g_{\mu\nu}$ is a symmetric tensor interpreted as the metric, $g_{\mu d}$ is a vector field and g_{dd} a scalar field. Correspondingly, we may write the D-dimensional metric as the following ansatz matrix:

$$g_{MN} \equiv \begin{pmatrix} g_{\mu\nu}(x) & \mathcal{A}_\mu(x) \\ \mathcal{A}_\nu(x) & -e^{2\sigma(x)} \end{pmatrix} \qquad \text{(IV-2.73)}$$

where we have restricted the spacetime dependence of the components to the non-compact dimensions (x^μ).

It turns out that part of the reparametrization invariance of the original D-dimensional theory may be interpreted as a gauge invariance associated with the vector field $\mathcal{A}_\mu(x) \sim g_{\mu d}$. Thus the D-dimensional gravitational theory provides a geometric unification of d-dimensional gravity and of an abelian gauge symmetry such as the one present in the theory of electrodynamics.

As for the component g_{dd} of the metric, it measures distances in the compact dimension in (higher-dimensional) Planck units. Thus the d-dimensional field $e^{\sigma(x)}$, or more precisely its vacuum value, provides a

measure of the size of the compact dimension[8]. It is often called a breathing mode. From this point of view, its x-dependence reflects the fluctuations of the compact dimension (along y) in directions transverse to it (measured by x^μ). As we will see below, the purely gravitational D-dimensional action yields a vanishing potential for this scalar field: it corresponds to a flat direction of the scalar potential. In a supersymmetric set up this real field becomes part of a complex scalar field associated with this flat direction. Such a complex field is thus a modulus and is traditionally written as T in 4 dimensions. Of course, compactification requires the size of the compact dimension to be determined: some extra dynamics must be included in order to lift the degeneracy associated with the flat direction and to determine the vacuum expectation value of the modulus field.

To be more explicit, let us consider the action of gravity in $D \equiv d+1$ dimensions:

$$\mathcal{S} = -\frac{1}{16\pi G_{(D)}} \int d^D x \sqrt{|g|}\ \mathcal{R}^{(D)} \qquad \text{(IV-2.74)}$$

where $\mathcal{R}^{(D)}$ is the curvature scalar associated with the D-dimensional metric. Following (IV-2.73), we write the D-dimensional line element as ($y \equiv x^d$):

$$ds^2 = g_{MN} dx^M dx^N = g^{(d)}_{\mu\nu} dx^\mu dx^\nu - e^{2\sigma(x)} \left(dy + A_\mu dx^\mu\right)^2 \qquad \text{(IV-2.75)}$$

Reparametrizations of the form $y \to y + \alpha(x^\mu)$ lead to gauge transformations for the field $A_\mu(x)$: $A_\mu(x) \to A_\mu(x) - \partial_\mu \alpha$.

With the ansatz (IV-2.75), we have

$$\mathcal{R}^{(D)} = \mathcal{R}^{(d)} - 2D^\mu \left(\partial_\mu \sigma\right) - 2\partial^\mu \sigma \partial_\mu \sigma + \frac{1}{4} e^{2\sigma} F^{\mu\nu} F_{\mu\nu} \qquad \text{(IV-2.76)}$$

where $\mathcal{R}^{(d)}$ is the curvature scalar built out of the metric $g^{(d)}_{\mu\nu}$. Hence, introducing

$$\frac{1}{G_{(d)}} \equiv \frac{1}{G_{(D)}} \int_0^L dy = \frac{L}{G_{(D)}} \qquad \text{(IV-2.77)}$$

we obtain after integrating by parts

$$\mathcal{S}_{\text{eff}} = -\frac{1}{16\pi G_{(d)}} \int d^d x\ \sqrt{|g^{(d)}|}\ e^\sigma\ \left[\mathcal{R}^{(d)} + \frac{1}{4} e^{2\sigma} F^{\mu\nu} F_{\mu\nu}\right] \qquad \text{(IV-2.78)}$$

[8]To be accurate, in our notations, the radius of the compact dimension is $R\langle e^\sigma \rangle$. One may alternatively set $R = 1$ in Planck units, in which case the metric coefficient fixes the radius; or normalize $\langle e^\sigma \rangle$ to 1: R is then the radius.

We see that no potential is generated for the breathing mode e^{σ}: it corresponds to a flat direction of the scalar potential. [9]. The presence of the modulus flat direction may be traced back to the dilatation symmetry $y \to e^{-\lambda}y$, $A_\mu(x) \to e^{-\lambda}A_\mu(x)$, $\sigma(x) \to \sigma(x) + \lambda$, which leaves the line element (IV-2.75) invariant.

We may also introduce matter fields in these higher-dimensional spacetimes. Since particles may be associated with waves, and extra compact dimensions with a multi-dimensional box, we expect standing waves in this higher-dimensional box. These are called Kaluza-Klein modes; they should be observed in our 4-dimensional world as particles.

Let us illustrate this on the case of a particle of zero spin and mass m_0 in $D = 5$ dimensions (we use the same notations as above with $d = 4$). It is described by a scalar field $\Phi(x^M)$ with equation of motion:

$$\partial^M \partial_M \Phi = \partial^\mu \partial_\mu \Phi + \partial^y \partial_y \Phi = -m_0^2 \Phi \tag{IV-2.79}$$

We may decompose $\Phi(x^\mu, y)$ on the basis of plane waves in the compact dimension $e^{ip_5 y}$: because of the identification $y \equiv y + 2\pi R$, we have $p_5 = n/R$, $n \in \mathbf{Z}$. Thus

$$\Phi(x^\mu, y) = \frac{1}{2\pi R} \sum_{n \in Z} \phi_n(x^\mu) e^{iny/R} \tag{IV-2.80}$$

and the equation of motion yields (we set here $g^{55} = 1$ for simplicity)

$$\partial^\mu \partial_\mu \phi_n = -\left(m_0^2 + \frac{n^2}{R^2} \right) \phi_n \tag{IV-2.81}$$

Thus the 5-dimensional field is seen as a tower of 4-dimensional particles with a spectrum characteristic of extra dimensions: particles with the same quantum numbers at regular mass-squared intervals. The spectrum has a typical mass scale $M_{_C}$ which is given by the inverse of the radius of the compact manifold

$$M_{_C} \sim R^{-1} \tag{IV-2.82}$$

Clearly a multidimensional manifold may have several "radii" and thus several T moduli associated.

The fact that no sign of extra dimensions has been found experimentally indicates that $M_{_C}$ is large. In this context, it is thus important to determine the effective theory at energies much smaller than $M_{_C}$. Since

[9]The absence of a kinetic term for the modulus field $\sigma(x)$ in (IV-2.78) does not mean that it is non-dynamical: the kinetic term is present in the Einstein frame i.e. in a frame with a standard Einstein term $\sqrt{g}R$, obtained by a Weyl rescaling of the metric ($g^{(d)}_{\mu\nu} = e^{2\sigma/(2-d)} g_{\mu\nu}$).

all massive fields ($m \sim M_C$) decouple, the task is to find out which are the "massless" fields ($m_0 = 0$ or $m_0 \ll M_C$) which are present at low energy. In the simple example that we have chosen, a massless ($m_0 = 0$) scalar field in D dimensions yields one and only one massless scalar field in 4 dimensions.[10] In the presence of several extra dimensions, there may be several radii and associated moduli fields.

IV-3 String dualities and branes

Theorists have been working for the last 30 years on the idea that string theory may provide the long awaited quantum theory of gravity. Much progress has been made since the early days, although one does not seem to have yet a complete understanding of string theory, of its concepts and methods.

We recall here that the fundamental objects are, at the string energy scale M_S (or distance scale M_S^{-1}), one-dimensional objects: open and closed strings. For the time being, we will center our attention on the closed string since its oscillations provide gravitationally coupled fields (including the graviton itself), among which we will find our moduli fields.

The oscillation modes in one direction along a closed string (say left-moving) decouple from the modes oscillating in the other direction (right-moving). If one introduces the creation operators α^I_{-1} and $\tilde{\alpha}^I_{-1}$ corresponding respectively to the left-moving and right-moving oscillations, one finds, among the massless modes of the closed string:

- $h^{IJ} = \frac{1}{2}\left(\alpha^I_{-1}\tilde{\alpha}^J_{-1} + \alpha^J_{-1}\tilde{\alpha}^I_{-1} - \frac{2}{D-2}\delta^{IJ}\sum_K \alpha^K_{-1}\tilde{\alpha}^K_{-1}\right)|0\rangle$, a symmetric traceless tensor (spin 2) field interpreted as the graviton field ($I, J = 1, ..., D-2$, D being the spacetime dimension);
- $b^{IJ} = \frac{1}{2}\left(\alpha^I_{-1}\tilde{\alpha}^J_{-1} - \alpha^J_{-1}\tilde{\alpha}^I_{-1}\right)|0\rangle$, an antisymmetric tensor field;
- $e^{\phi} = \sum_K \alpha^K_{-1}\tilde{\alpha}^K_{-1}|0\rangle$, a scalar field known as the string dilaton, which, as we will now see, is a modulus field. Its ground state value determines the string coupling (see (IV-3.85) below).

We recall that quantum consistency requires to consider supersymmetric string i.e. superstring theories in $D = 10$ dimensional spacetime. There are actually five distinct superstring theories in 10 dimensions, which we identify briefly.

[10]The higher-dimensional graviton also has Kaluza-Klein modes. The ansatz that we have chosen for the metric in (IV-2.73) corresponds to neglecting the massive modes – they have a non-trivial y dependence – and restricting our attention to the zero modes.

Left-moving fermions may have the same or the opposite chirality as the right-moving fermions. Depending on this choice, the closed superstring theory is called a type IIA or type IIB superstring. One may also consider open strings together with closed strings (for consistency and in order to obtain a graviton among the massless modes). In this case only one supersymmetry charge is allowed and the corresponding theory is refered to as type I. Open strings may carry gauge charges at their ends, which allows them to describe gauge theories. Finally, since left-movers and right-movers may be quantized independently, it has been realized that one can describe simultaneously the right-movers by a superstring theory (and thus obtain $N = 1$ spacetime supersymmetry) in 10 dimensions and the left-movers by a standard bosonic string theory in 26 dimensions. Spacetime obviously has only the standard 10 dimensions. The extra 16 compact dimensions found in the right-movers are considered as internal and provide gauge degrees of freedom (limited to $SO(32)$ or $E_8 \times E_8$ by the cancellation of quantum anomalies). This gives the heterotic string theory, which has provided the first candidates for a theory of all fundamental interactions.

If, for the sake of illustration, we write the 10-dimensional effective supergravity action of the weakly coupled heterotic string, we find that it includes the following terms:

$$\mathcal{S} = -\int \frac{d^{10}x}{(2\pi)^7}\sqrt{|g|}e^{-2\phi}\left[M_S^8\left(\mathcal{R}^{(10)} + 4\partial^\mu\phi\partial_\mu\phi\right) + M_S^6\frac{1}{4}\mathrm{Tr}F^2 + \cdots\right] \tag{IV-3.83}$$

We note the presence of the dilaton field e^ϕ with a vanishing potential (hence a flat direction): it is a modulus field.

Once one compactifies on a 6-dimensional manifold of volume R^6, one obtains

$$\begin{aligned}\mathcal{S} &= -\int \frac{d^4x}{(2\pi)^7}\sqrt{|g|}\left(M_S^8R^6e^{-2\phi}\mathcal{R}^{(4)} + M_S^6R^6e^{-2\phi}\frac{1}{4}\mathrm{Tr}F^2 + \cdots\right), \\ &= -\int d^4x\sqrt{|g|}\left(\frac{1}{2}m_P^2\mathcal{R}^{(4)} + \frac{1}{16\pi\alpha_U}\mathrm{Tr}F^2 + \cdots\right)\end{aligned} \tag{IV-3.84}$$

from which we read the 4-dimensional Planck scale as well as the value of the gauge coupling α_U at the string scale. Introducing the string the compactification scale $M_C \equiv R^{-1}$, we obtain an expression for the string scale and the string coupling λ_S:

$$M_S^2 = 2\pi\alpha_U m_P^2 \quad , \quad \lambda_S \equiv \langle e^\phi\rangle = \frac{\alpha_U^2}{2\pi^{3/2}}\frac{m_P^3}{M_C^3} \tag{IV-3.85}$$

This shows that the string scale is of the order of the Planck scale (taking for α_U the value at unification: 1/24). Moreover, if the compact manifold

is isotropic, M_C represents, to a first approximation, the scale where the theory becomes truly unified and is thus interpreted as the gauge coupling unification scale M_U. It is clear in this context that (IV-3.85) implies a large string scale M_S. One obtains different relations for the other string theories which allow lower string scales.

Our discussion of mass scales shows the pre-eminent role played by fundamental scalar fields in string theory: the dilaton fixes the string coupling, other moduli fields determine the radius (and shape) of the compact manifold (as we have seen in the preceding subsection). Obviously the low energy physics depends on the value of these moduli fields.

Before we explain why dilaton and radii correspond to flat directions, we have to show how these *real* scalar fields fit into supersymmetric multiplets. The antisymmetric tensor b^{IJ} which is present among the massless modes of the closed string plays a crucial role to provide the missing bosonic degrees of freedom (remember for example that the scalar component of a chiral supermultiplet is complex).

For example, to form the complex modulus field T, the radius-squared R^2 of the compact manifold is paired up with an imaginary part which is related to the antisymmetric tensor field b^{kl} (with k and l 6-dimensional *compact* indices; hence the corresponding components are 4-dimensional scalars). Similar interpretations apply to the other radii moduli, known as Kähler moduli. The gauge invariance of the antisymmetric tensor ($\delta b^{IJ} = \partial^I \Lambda^J - \partial^J \Lambda^I$) induces a Peccei-Quinn symmetry for Im T (Im $T \to$ Im $T+$ constant) which has only derivative couplings, just like the axion. Hence the superpotential cannot depend on Im T, and being analytic in the fields, cannot depend on T as a whole [19].

Through supersymmetry, the string dilaton ϕ is related to the antisymmetric tensor $b^{\mu\nu}$ (this time with 4-dimensional indices). Together with a Majorana fermion, the dilatino, they form what is known as a linear supermultiplet L, which is real. The superpotential, being analytic in the fields, cannot depend on L. This is related again to the gauge invariance associated with the antisymmetric tensor. This in turn ensures that the superpotential cannot depend on ϕ.

The latter result may be interpreted from the point of view of standard nonrenormalisation theorems [20, 21]. Indeed, since e^{ϕ} is the string coupling, it ensures that the superpotential is not renormalized, to all orders of string perturbation theory.

Before we proceed, let us be more explicit on the way the string dilaton and the T modulus appear in 4 dimensions [11]. We work here in the string frame and the corresponding fundamental mass scale is the the string scale M_S. In order to introduce the degree of freedom associated with the overall size of the compact manifold, we introduce the "breathing mode" e^σ through the compact space part of the metric:

$$g_{kl}(x,y) = e^{2\sigma(x)} g^{(0)}_{kl}(y) \ , \ k,l = 4,\cdots,9, \quad \int d^6y \sqrt{|g^{(0)}|} = M_S^{-6} \tag{IV-3.86}$$

Thus the volume of the 6-dimensional compact manifold is $R^6 = \int d^6y \sqrt{|g|} = M_S^{-6} \langle e^{6\sigma} \rangle$. Thus $\langle e^{2\sigma} \rangle$ measures R^2 in string units.

If we consider specifically the weakly coupled heterotic string, then the terms in IV-3.83 give, after compactification,

$$\mathcal{S} = \int d^4x \sqrt{|g^{(4)}|} \ \frac{1}{(2\pi)^7} e^{-2\phi+6\sigma} ...$$
$$... \left[M_S^2 \left(-\mathcal{R}^{(4)} + 12 D^\mu \partial_\mu \sigma + 42 \partial^\mu \sigma \partial_\mu \sigma - 4 \partial^\mu \sigma \partial_\mu \phi \right) ... \right.$$
$$\left. ... - \frac{1}{4} \mathrm{Tr} F^{\mu\nu} F_{\mu\nu} \right] \tag{IV-3.87}$$

In terms of the real fields

$$s = \frac{1}{(2\pi)^7} e^{-2\phi+6\sigma} \quad , \quad t = \frac{1}{(2\pi)^7} e^{2\sigma} \tag{IV-3.88}$$

the action reads, after integrating by parts,

$$\mathcal{S} = \int d^4x \sqrt{|g^{(4)}|} \, s \left[M_S^2 \left(-\mathcal{R}^{(4)} + \frac{3}{2} \frac{\partial^\mu t \partial_\mu t}{t^2} - \frac{\partial^\mu s \partial_\mu s}{s^2} \right) - \frac{1}{4} \mathrm{Tr} F^{\mu\nu} F_{\mu\nu} \right] \tag{IV-3.89}$$

We conclude that

$$m_P^2 = 2 \langle s \rangle M_S^2 \quad , \quad \frac{1}{g^2} = \langle s \rangle \tag{IV-3.90}$$

The couplings of the s field are reminiscent of the Wess-Zumino terms which restore scale invariance through a dilaton field (see subsection III-3).

The sign of the kinetic term for s is not a problem because s is coupled to the spacetime curvature. If we go to the Einstein frame by performing

[11] The reader who is not interested in these technical details may go directly below Eq. (IV-3.95).

a Weyl transformation on the 4-dimensional metric:

$$g^{(4)}_{\mu\nu} \equiv \frac{1}{2s}\frac{m_P^2}{M_S^2} g_{\mu\nu} \tag{IV-3.91}$$

the action takes the standard form

$$\mathcal{S} = \int d^4x\sqrt{|g|}\left[-\frac{1}{2}m_P^2\mathcal{R} + \frac{1}{4}\frac{\partial^\mu s\partial_\mu s}{s^2} + \frac{3}{4}\frac{\partial^\mu t\partial_\mu t}{t^2} - \frac{1}{4}\, s\ \mathrm{Tr}F^{\mu\nu}F_{\mu\nu}\right] \tag{IV-3.92}$$

The fields s and t appear to be the real parts of complex scalar fields S and T with Kähler potential

$$K(S,T) = -\ln\left(S+\bar{S}\right) - 3\ln\left(T+\bar{T}\right) \tag{IV-3.93}$$

One readily checks, using (IV-1.67) and (IV-1.68) that the corresponding kinetic terms yield the kinetic terms for s and t just found. The imaginary part of S is provided by the pseudoscalar field, or string axion a, which is equivalent to the antisymmetric tensor field in four dimensions ($\partial^\mu a \sim \epsilon^{\mu\nu\rho\sigma}\partial_\nu b_{\rho\sigma}$).

The Lagrangian (IV-3.92) is invariant under the group $SL(2,\mathbf{Z})$ of modular transformations

$$T \to \frac{aT-ib}{icT+d}\ , \quad ad-bc=1\ , \quad a,b,c,d\in\mathbf{Z} \tag{IV-3.94}$$

among which we recognize T-duality ($T \to 1/T$). Indeed, such a transformation corresponds to a Kähler transformation for K:

$$K \to K + F + \bar{F}\ , \quad F = 3\ln\left(icT+d\right) \tag{IV-3.95}$$

Type IIA, IIB, I, heterotic $SO(32)$ and $E_8\times E_8$ form the five known types of superstring theories. There are some unexpected equivalences between them, associated with the presence of the moduli fields just discussed. These equivalences are basically of 2 types:

- large/small compactification radius duality

  When we go from a point particle to a string, the physics of compactification becomes incomparably richer. We still find that momenta in the compact dimensions are quantized in units of $1/R$: this yields *Kaluza-Klein modes* with energies proportional to $1/R$. But compact dimensions allow the possibility of the string winding around them (think of the string wrapped N times around

a circle of radius R). The stable configuration thus obtained is called a *winding mode*. It has an energy proportional to R, since it would vanish for zero radius, and to the number m of wrappings: $E \sim mR$.

Hence, when R is large (with respect to M_S^{-1}), the Kaluza-Klein modes are light whereas the winding modes are heavy. An effective low energy theory would include only the Kaluza-Klein states. It is the contrary when R is small.

It turns out that the two corresponding theories are equivalent. This can be expressed as a symmetry of the theory associated with the transformation of the modulus field $T \leftrightarrow 1/T$ (a special case of (IV-3.94)). Under the large/small compactification radius duality, a theory A with a large compact dimension is equivalent to a theory B with a small compact dimension. Such a duality is known as T-duality.

- strong/weak coupling duality (or S-duality).

  The five string theories discussed above have been defined in their perturbative regimes. In other words, if λ_S is the string coupling for one of these theories, the theory is defined by its perturbative expansion: corresponding amplitudes are expressed as power series in λ_S. Non-perturbative effects appear to vanish at small coupling, as for example an instanton contribution of order e^{-1/g^2}. Strong/weak coupling duality relates a theory $\mathcal{T}_1$ in its strong coupling regime to a theory $\mathcal{T}_2$ in its weak coupling regime: an amplitude $\mathcal{M}^{(1)}(\lambda_S^{(1)})$ in theory $\mathcal{T}_1$ can be understood as amplitude $\mathcal{M}^{(2)}(\lambda_S^{(2)} = 1/\lambda_S^{(1)})$. Such a type of duality relates for example type I superstring to the $SO(32)$ heterotic string theory whereas type IIB string theory is self-dual.

  The name S-duality refers to the 4-dimensional scalar field which has been discussed above. Since $\langle S \rangle \sim 1/g^2$, strong/weak duality corresponds to the duality $S \to 1/S$.

Duality relations allow to have access to non-perturbative effects in a given superstring theory by studying perturbatively its dual theory. The most striking discovery in this respect was the realization that there exists an eleventh dimension. More precisely, the spectrum of type IIA superstring theory was found to include states with mass M_S/λ_S at weak coupling. Supersymmetry helped to solve the problem of the bound state of N such particles which was found to have precisely a mass of NM_S/λ_S. This is strongly reminiscent of the mass spectrum of Kaluza-Klein modes with a radius of compactification $R_{(11)} \equiv \lambda_S/M_S$.

The reason why this eleventh dimension was not found in the perturbative string approach is that perturbation precisely means an expansion around $\lambda_{_S} = 0$ and thus, at fixed string scale $M_{_S}$, around $R_{_{(11)}} = 0$. This is why a perturbative expansion does not "see" this eleventh dimension. In all generality, one expects the ultimate M-theory to be eleven-dimensional (at least). It is connected with the five known superstring theories.

V. Gravitational and cosmological aspects

V-1 Cosmological relevance of scalar fields

We start with a quick reminder on cosmology. The fundamental equations are Einstein's equations:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G_{_N} T_{\mu\nu} + \lambda g_{\mu\nu} \tag{V-1.96}$$

where $T_{\mu\nu}$ is the energy-momentum tensor of matter and $R_{\mu\nu}$ the Ricci tensor, which is obtained from the Riemann tensor (constructed from the metric tensor) measuring the curvature of spacetime. The cosmological constant λ is of the dimension of an inverse length squared. It was introduced by Einstein [22] in order to build a static universe model, its repulsive effect compensating the gravitational attraction.

Assuming that the Universe is homogeneous and isotropic on large scales, one may take as an ansatz for the metric the Robertson-Walker form:

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2 \theta d\phi^2 \right) \right] \tag{V-1.97}$$

where $a(t)$ is the cosmic scale factor and $k = \pm 1$ or 0, depending on whether the universe is spatially closed, open or flat. We may also assume that the dominant component of the Universe (radiation or non-relativistic matter depending on the epoch) is a perfect fluid of energy density ρ and pressure p.

Inserting the ansatz (II-1.15) into the Einstein's equations (V-1.96) yields the Friedmann equation ($\dot{a} = da/dt$)

$$H^2 \equiv \frac{\dot{a}^2(t)}{a^2(t)} = \frac{1}{3} \left(\lambda + \frac{\rho}{m_{_P}^2} \right) - \frac{k}{a^2} \tag{V-1.98}$$

The Hubble parameter H gives an estimate of the rate of expansion.

Observation seems to indicate that the Universe is spatially flat: $k = 0$. In other words, we may write (V-1.98) at present time t_0 as follows:

$$\Omega_\Lambda + \Omega_M = 1 \ , \quad \Omega_\Lambda \equiv \frac{\lambda}{3H_0^2} \ , \quad \Omega_M \equiv \frac{\rho}{\rho_c} \tag{V-1.99}$$

where we have introduced the critical density $\rho_c \equiv 3m_P^2 H_0^2$ (H_0 is the present value of the Hubble parameter: $H_0 = h_0 \times 100$ km.s^{-1}Mpc^{-1} with $h_0 \sim 0.7$).

The standard explanation for the flatness of the Universe is that it has undergone at an early stage a period of exponential expansion, known as inflation [23]. In the standard scenario, this expansion is due to the presence of a dynamical component whose energy density ρ dominates and is almost constant. The expansion rate measured by the Hubble parameter (V-1.98) (with $k \sim 0$) is then constant , which yields an exponential behaviour for the cosmic scale factor: $a(t) \propto e^{Ht}$.

The energy density of relativistic particles (resp. non-relativistic) particles goes as a^{-4} (resp. a^{-3}) and thus does not provide the desired behaviour. On the other hand, since bosons can be in the same quantum state, they can form a coherent superposition in the form of a macroscopic field. Isotropy requires to consider only scalar fields. This is why scalar field play a central role in cosmology!

To be more precise, associated with a scalar field ϕ of mass m, one can have the following two realizations:

- a Bose gas at non-zero temperature (with the standard Bose-Einstein distribution $n_{\mathbf{k}} = \left[\exp\left(\sqrt{\mathbf{k}^2 + m^2}/T\right) - 1\right]^{-1}$)
- a Bose condensate ϕ_0 of non-interacting scalars with vanishing momentum $\mathbf{k}$ ($n_{\mathbf{k}} = (2\pi)^3 \phi_0 m \delta^3(\mathbf{k})$)

The Bose condensate corresponds to infinitely large occupation number (at $\mathbf{k} = \mathbf{0}$) and thus behaves as a classical field. Indeed, it has been shown [24] that gravitational production of scalar fields amplifies ("squeezes") the zero-point quantum fluctuations: a gravitationally produced scalar field would be left in a squeezed state that acts like a scalar field.

We will see in what follows that the moduli fields discussed in the preceding Section tend to be extremely light ($m < 10^{-26}$ eV). This means that the associated de Broglie wave length ($\hbar/p$) could reach cosmological distances. The condensate would span the whole observable Universe. This allows to use classical scalar field condensates to solve some of the cosmological problems, as we will see in more details in what follows.

Recent cosmological observation expressed in the plane $(\Omega_M, \Omega_\Lambda)$ singles out a region in parameter space centered around the values: $\Omega_M \sim 0.3$ and $\Omega_\Lambda \sim 0.7$. According to (V-1.99), this implies that $|\lambda| \leq H_0^2$. In other words, the length scale $\ell_\Lambda \equiv |\lambda|^{-1/2}$ associated with the cosmological constant must be larger than the Hubble length $\ell_{H_0} \equiv cH_0^{-1} = h_0^{-1}.10^{26}$ m, and thus be a cosmological distance.

This is not a problem as long as one remains classical: ℓ_{H_0} provides a natural cosmological scale for our present Universe. The problem arises when one tries to combine gravity with the quantum theory of matter. Indeed, even if one sets λ to zero, one expects, in the context of the quantum theory, a non-vanishing vacuum (*i.e.* ground state) energy: $< T_{\mu\nu} >= \rho_{\rm vac} g_{\mu\nu}$; then the Einstein equations (V-1.96) read

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G_{_N} T_{\mu\nu} + 8\pi G_{_N} \rho_{\rm vac} g_{\mu\nu} \qquad \text{(V-1.100)}$$

The last term is interpreted as an effective cosmological constant:

$$\lambda_{\rm eff} = 8\pi G_{_N} \rho_{\rm vac} \equiv \frac{\Lambda^4}{m_{_P}^2} \qquad \text{(V-1.101)}$$

Generically, $\rho_{\rm vac}$ receives a non-zero contribution from symmetry breaking: for instance, the scale Λ would be typically of the order of 100 GeV in the case of the electroweak gauge symmetry breaking or 1 TeV in the case of supersymmetry breaking. But the constraint $|\lambda_{\rm eff}| \leq H_0^2$ now reads:

$$\Lambda \leq 10^{-30}\ m_{_P} \sim 10^{-3}\ \text{eV} \qquad \text{(V-1.102)}$$

It is this very unnatural fine-tuning of parameters (in explicit cases $\rho_{\rm vac}$ and thus Λ are functions of the parameters of the theory) that is referred to as the *cosmological constant problem*, or more accurately the vacuum energy problem.

The most natural reason why vacuum energy would be vanishing is a symmetry argument. Global supersymmetry indeed provides such a rationale. The problem is that, at the same time, supersymmetry predicts equal boson and fermion masses and therefore needs to be broken. The amount of breaking necessary to push the supersymmetric partners high enough not to have been observed yet, is incompatible with the limit (V-1.102).

Moreover, in the context of cosmology, we should consider supersymmetry in a gravity context and thus work with its local version, supergravity. In this context, the criterion of vanishing vacuum energy is traded for one of vanishing gravitino mass. Local supersymmetry is compatible with a non-vanishing vacuum energy, preferably a negative one (although possibly also a positive one).

Dilaton, scalar-tensor theories and tests of general relativity. We have encountered in the preceding Section moduli fields which couple to matter with gravitational strength. If these fields remain light they induce a long rang force similar to gravity which might lead to difficulties when confronting observation.

One of the stringent constraints on gravitational-type interactions comes from the high accuracy at which the equivalence principle has been tested[12]. In its weak form, the equivalence principle states the universality of free fall: two test bodies at the same location and at rest with respect to each other, fall in the same way in an external gravitational field, independently of their mass and composition (hence inertial and gravitational masses are identical). In the Einstein formulation, at every point of an arbitrary gravitational field, it is possible to define locally a coordinate system such that the laws of nature take the same form as in special relativity (see the book by Will [25] for a more detailed formulation).

Let us consider for example the string dilaton coupling to gauge fields, as obtained in (IV-3.89),

$$\mathcal{S} = -\frac{1}{4}\int d^4x\sqrt{|g^{(4)}|}\ s\ F^{\mu\nu}F_{\mu\nu} \qquad \text{(V-1.103)}$$

As long as the dilaton s is not stabilized, the gauge coupling constants depend on space and time ($1/g^2 = s$). Since the mass of hadrons is mostly gluon field energy, it follows that these masses also depend on space and time and we lose the universality of free fall.

It should be noted that the scalar field dependence in (V-1.103) cannot be absorbed in a Weyl transformation of the metric,

$$g^{(4)}_{\mu\nu} = A^2(\phi)g_{\mu\nu} \qquad \text{(V-1.104)}$$

because $\sqrt{|g^{(4)}|}g^{(4)\mu\rho}g^{(4)\nu\sigma}$ is Weyl invariant. The easiest way to satisfy the stringent constraints imposed by the apparent absence of violations of the equivalence principle is to consider a scalar-tensor theory for which the matter fields couple to a universal metric of the form (V-1.104) where ϕ stands for one or several (ϕ^a, $a = 1, \cdots, n$) scalar fields: lengths and times are measured by rods and clocks in the frame defined by this unique metric.

It is possible to appeal to the cosmological evolution to account for the smallness of such coefficients in scalar-tensor theories. For example, Damour and Nordtvedt [26] have found an attractor mechanism towards General Relativity.

Time variation of fundamental constants. Since couplings and scales are often given in terms of moduli fields, it is tempting to consider that, since moduli may not all have been stabilized, some of these

[12]To give an idea of the orders of magnitude involved, the relative difference in acceleration $|\Delta\mathbf{a}|/|\mathbf{a}|$ between two bodies of different composition in the Earth gravitational field is presently measured to be smaller than 10^{-12}.

quantities are still presently varying with time or have been doing so in the course of the cosmological evolution. This leads to the fascinating possibility that some of the fundamental constants of nature are time-dependent.

Such an idea was put forward by Dirac [27, 28]. According to him, a fundamental theory should not involve fundamental dimensionless parameters (i.e. dimensionless ratios of fundamental parameters) which are very large numbers. Such numbers should instead be considered as resulting from the evolution of the universe and the corresponding dimensionless parameters be variables characterizing the evolving state of the universe. Obviously this leads to some time-dependent fundamental parameters.

In the context of supersymmetric theories where many of these dimensionless ratios are fixed by the values of moduli fields, one may expect some time dependence. For example, in heterotic string theory we have seen that the Planck scale (hence Newton's constant) is given in terms of the string scale by the vacuum expectation value of the string dilaton. Similarly for the 4-dimensional coupling, evaluated at the string scale (close to unification scale). If the dilaton is not properly stabilized at low energy, that is if the flat direction is not lifted or if its minimum remains too shallow, one thus expects a possible time dependence of the dimensionless ratio M_P/M_S or of the fine structure constant.

There are however some stringent bounds on the possible time evolution of fundamental constants [29]. For example, present limits on $\left|\dot{G}_N/G_N\right|$ are in the 10^{-12} yr^{-1} region whereas the presence in the Oklo uranium in Gabon of a natural fission reactor which operated some 10^9 yr ago puts a limit [30] on $|\dot{\alpha}/\alpha|$ in the 10^{-17} yr^{-1} region.

Moduli problem. Because moduli are light and have gravitational interactions, they are long lived. There are then two potential dangers. If their lifetime is smaller than the age of our universe, their decay might have released a very large amount of entropy in the universe and diluted its content. If their lifetime is larger than the age of our universe, they might presently still be oscillating around their minimum and the energy stored in these oscillations may overclose the universe. One refers to these problems as the moduli problem [31–33]. Taken at their face values, such constraints forbid any modulus field which is not superlight or very heavy. We now proceed to make these statements quantitative.

We first define two quantities which play a central role in this discussion. A modulus field ϕ has typically gravitational interactions and thus its decay constant Γ_ϕ scales like m_P^{-2}. Since the only available scale

is the scalar field mass m_ϕ, one infers from simple dimensional analysis that

$$\Gamma_\phi = \frac{m_\phi^3}{m_{_P}^2} \tag{V-1.105}$$

Since the age of the Universe is of order H_0^{-1}, one deduces that the modulus will decay at present times if $\Gamma_\phi \sim H_0$, that is if its mass m_ϕ is of order $\left(H_0 m_{_P}^2\right)^{1/3} \sim 20$ MeV.

The other relevant quantity is the initial value f_ϕ of the scalar field with respect to its ground state value ϕ_0. Presumably at very high energy (that is above the phase transition associated with dynamical supersymmetry breaking) the flat direction is restored, one expects generically that $f_\phi \sim m_{_P}$ since this is the only scale available.

Let us first consider the case where $m_\phi < 20$ MeV, that is a field which has not yet decayed at present time. The equation of evolution for the field ϕ reads

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = -\Gamma_\phi \dot{\phi} \tag{V-1.106}$$

where the friction term (proportional to the Hubble parameter H) accounts for the expansion of the universe and the last term comes from particle creation due to the time variation of ϕ.

As long as $H > m_\phi$, the friction term $3H\dot{\phi}$ dominates in the equation of motion and the field ϕ remains frozen at its initial value f_ϕ. When $H \sim m_\phi$, i.e. for $T_I \sim (m_\phi m_{_P})^{1/2}$ (since $H \sim T^2/m_{_P}$), the field ϕ starts oscillating around the minimum ϕ_0 of its potential. According to a standard result, coherent oscillations behave like non-relativistic matter i.e.

$$\rho_\phi(T) = \rho_\phi(T_I)\left(\frac{T}{T_I}\right)^3 \sim m_\phi^2 f_\phi^2 \left(\frac{T}{T_I}\right)^3 \tag{V-1.107}$$

Since the radiation energy density $\rho_R(T)$ behaves as T^4, ρ_ϕ/ρ_R increases as the temperature of the universe decreases and one reaches a time where the energy of the scalar field oscillations dominates the energy density of the universe. One should then make sure that $\rho_\phi(T_0) < \rho_c$. Using (V-1.107) and $T_I = (m_\phi m_{_P})^{1/2}$, one may write this condition as

$$m_\phi < m_{_P}\left(\frac{\rho_c m_{_P}}{f_\phi^2 T_0^3}\right)^2 \sim 10^{-26}\ \text{eV} \tag{V-1.108}$$

where we have set $f_\phi \sim m_{_P}$. Thus, if 10^{-26} eV $< m_\phi < 20$ MeV, there is too much energy stored in the ϕ field (which has not yet decayed at present times).

We next consider the case where $m_\phi > 20$ MeV, that is the scalar field has already decayed at present times. Decay occurs at a temperature T_D when $H(T_D) \sim \Gamma_\phi$ i.e.

$$\Gamma_\phi^2 \sim \frac{\rho_\phi(T_D)}{m_P^2} = \frac{\rho_\phi(T_I)}{m_P^2}\left(\frac{T_D}{T_I}\right)^3 \tag{V-1.109}$$

where we have used (V-1.107), assuming that, at T_D, the scalar field energy density dominates over radiation; $\rho_\phi(T_I) \sim m_\phi^2 f_\phi^2$. At decay, all energy density is transfered into radiation. Thus, the reheating temperature T_{RH}, that is the temperature of radiation issued from the decay, is given by the condition $\rho_\phi(T_D) \sim T_{RH}^4$. Using (V-1.109) to express $\rho_\phi(T_D)$, we obtain

$$T_{RH} \sim m_{_P}^{1/2}\Gamma_\phi^{1/2} \sim \frac{m_\phi^{3/2}}{m_{_P}^{1/2}} \tag{V-1.110}$$

The entropy release is

$$\sigma \equiv \frac{S_{RH}}{S_D} = \left(\frac{T_{RH}}{T_D}\right)^3 \sim \frac{1}{m_{_P}^{1/2}\Gamma_\phi^{1/2}}\frac{\rho_\phi(T_I)}{T_I^3} \tag{V-1.111}$$

This gives, using $T_I \sim m_\phi^{1/2} m_{_P}^{1/2}$ and $\rho_\phi(T_I) \sim m_\phi^2 f_\phi^2$, $\sigma \sim f_\phi^3/(m_\phi m_{_P}^2)$. With $f_\phi \sim m_{_P}$, this gives a very large entropy release as long as the modulus mass remains much smaller than the Planck scale.

This entropy release must necessarily precede nucleosynthesis since otherwise it would dilute away its effects. This condition, namely $T_{RH} > 1$ MeV, gives $m_\phi > 10$ TeV. Thus for 20 MeV $< m_\phi <$ 10 TeV, the entropy release following the decay of the modulus field is too large to be consistent with present observations.

In the absence of other effects, we are left with only superlight moduli fields ($m_\phi < 10^{-26}$ eV) or heavy ones ($m_\phi > 10$ TeV).

V-2 Inflation scenarios

As we have seen above, a standard scenario for inflation involves a scalar field ϕ evolving slowly in its potential $V(\phi)$. The equation of motion is simply

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \tag{V-2.112}$$

Slow evolution of the scalar field requires the friction term to dominate on the right-hand side: $3H\dot{\phi} \sim -V'$.

The energy density and pressure stored in the scalar field are (we recognize the kinetic energy $\dot{\phi}^2/2$ and the potential energy $V(\phi)$)

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\ , \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{V-2.113}$$

Differentiating with respect to time the Friedmann equation (V-1.98), which reads here $3m_P^2H^2 = \rho_\phi$, one obtains $\dot{H} = -\dot{\phi}^2/(2m_P^2)$. Hence, the condition of almost constancy of H i.e. $|\dot{H}| \ll H^2$ amounts to the condition $\dot{\phi}^2/2 \ll V(\phi)$ i.e. kinetic energy for the scalar field much smaller than its potential energy. Using the Friedmann equation and (V-2.112), this can be written as the following slowroll constraints:

$$\varepsilon \equiv \frac{1}{2}\left(\frac{m_P V'}{V}\right)^2 \ll 1 \ , \quad \eta \equiv \frac{m_P^2 V''}{V} \ll 1 \tag{V-2.114}$$

In the de Sitter phase i.e. in the phase of exponential growth of the cosmic scale factor, quantum fluctuations of the scalar field value are transmitted to the metric. Because the size of the horizon is fixed (to H^{-1}) in this phase, the comoving scale a/k associated with these fluctuations eventually outgrows the horizon, at which time the fluctuations become frozen. It is only much later when the universe has recovered a radiation or matter dominated regime that these scales reenter the horizon and evolve again. They have thus been protected from any type of evolution throughout most of the evolution of the universe (this is in particular the case for the fluctuations on a scale which reenters the horizon *now*). Fluctuations in the cosmic microwave background provide detailed information on the fluctuations of the metric. In particular, the observation by the COBE satellite of the largest scales puts a important constraint on inflationary models. Specifically, in terms of the scalar potential, this constraint known as COBE normalization, reads:

$$\frac{1}{m_P^3}\frac{V^{3/2}}{V'} = 5.3 \times 10^{-4} \tag{V-2.115}$$

Using the slowroll parameter introduced above, this can be written as

$$V^{1/4} \sim \varepsilon^{1/4}\ 6.7 \times 10^{16}\ \text{GeV} \tag{V-2.116}$$

In most of the models that we will be discussing, ε is very small. However as long as $\varepsilon \gg 10^{-52}$, we have $V^{1/4} \gg 1$ TeV. In other words, the typical scale associated with inflation is then much larger than the TeV, in which case it makes little sense to work outside a supersymmetric context.

From the point of view of supersymmetry, one might expect that the presence of numerous flat directions may ease the search for an inflating potential[13]. One possible difficulty arises from the condition (V-2.114)

[13]although we have seen that this leads to new problems, the solution of which may require late inflation.

on η which may be written as a condition on the mass of the inflaton field

$$m^2 \ll H^2 \tag{V-2.117}$$

Since supersymmetry breaking is expected to set the scale that characterizes departures from flatness, it should control both m and $V^{1/4}$. For example, in the case of gravity mediation, we expect both m^2 and $H^2 \sim V/m_P^2$ to be of the order of $m_{3/2}^2$. If one does not want to be playing with numbers of order one to explain the $N = 50$ e-foldings of exponential evolution necessary to a satisfactory inflation scenario, one should be ready to introduce a second scale into the theory.

Since supersymmetric scalar potentials consist of F-terms and D-terms, the discussion of suitable potentials for inflation naturally follows this classification. As we will see in the following, they naturally provide models for what is known as hybrid inflation which involves two directions in field space: one is slow-rolling whereas the other ensures the exit from inflation (and is fixed during slowroll). We conclude this subsection by returning to the case of moduli fields in this context of inflation.

***F* term inflation.** Let us start with a simple illustrative model [34]. We consider two chiral supermultiplets of respective scalar components σ and χ with superpotential

$$W(\sigma, \chi) = \sigma \left(\lambda \psi^2 - \mu^2\right) \tag{V-2.118}$$

Writing $|\sigma| \equiv \phi/\sqrt{2}$, one obtains for the scalar potential

$$V = 2\lambda^2 \phi^2 \left|\psi\right|^2 + \left|\lambda \psi^2 - \mu^2\right|^2 \tag{V-2.119}$$

The global supersymmetric minimum is found for $\psi^2 = \mu^2$ and $\phi = 0$ but, for fixed ϕ, we may write the potential as ($\psi \equiv A + iB$)

$$V = \mu^4 + 2\lambda(\lambda\phi^2 - \mu^2)A^2 + 2\lambda(\lambda\phi^2 + \mu^2)B^2 + \lambda^2(A^2 + B^2)^2 \tag{V-2.120}$$

We conclude that, for $\phi^2 > \phi_c^2 \equiv \mu^2/\lambda$, there is a local minimum at $A = B = 0$ for which $V = \mu^4$. In other words, the ϕ direction is flat for $\phi > \phi_c$ with a non-vanishing potential energy. This may lead to inflation if one is trapped there. Since global supersymmetry is broken along this direction, one expects that loop corrections yield some slope which allows slowroll. Once ϕ reaches ϕ_c, ψ starts picking up a vacuum expectation value and one quickly falls into the global minimum.

This simple example of F-term hybrid inflation may easily be generalized. However F-term inflation suffers from a major drawback when one

tries to consider it in the context of supergravity [34, 35]. We recall the form of the scalar potential in supergravity, as written in (IV-1.70). In what follows the crucial role is played by the exponential factor e^{K/m_P^2} in front of the F-terms. Thus, inflation necessarily breaks supersymmetry. Let us assume for a moment that the inflation is dominated by some of the F-terms and that the D-terms are vanishing or negligible. Then the slow roll conditions (V-2.114) can be written as

$$\varepsilon = \frac{1}{2}\left(\frac{K_I}{m_P} + \cdots\right)^2 << 1 \quad \eta = K_{I\bar{I}} + \cdots << 1 \tag{V-2.121}$$

Here the subscript I denotes a derivative with respect to the inflaton noted ϕ^I. The latter condition is difficult to satisfy. The quantity $K_{I\bar{I}}$ stands in front of the kinetic term and therefore in the true vacuum it should be normalized to one. Then it is very unlikely to expect it to be much smaller during inflation. Indeed, this condition can be written as (V-2.117) since the mass of the inflaton m^2 receives a contribution $K_{I\bar{I}}V/m_P^2 \sim K_{I\bar{I}}H^2$.

These arguments indicate that it is not easy to implement F-type inflation in supergravity theories. All the solutions proposed involve specific non-minimal forms of the Kähler potential [35].

***D* term inflation.** What is interesting about inflation supported by D-terms is that the problems discussed above can be automatically avoided because of the absence of a factor e^{K/m_P^2} in front of them. Indeed for inflation dominated by some of the D-terms the slow roll conditions can be easily satisfied.

Let us show how such a scenario can naturally emerge in a theory with a $U(1)$ gauge symmetry [36, 37]. We consider an example with global supersymmetry. We introduce three chiral superfields ϕ_0, ϕ_+ and ϕ_- with charges equal to 0, +1 and −1 respectively. The superpotential has the form

$$W = \lambda\phi_0\phi_+\phi_- \tag{V-2.122}$$

which can be justified by several choices of discrete or continuous symmetries. The scalar potential in the global supersymmetry limit reads:

$$V = \lambda^2|\phi_0|^2\left(|\phi_-|^2 + |\phi_+|^2\right) + \lambda^2|\phi_+\phi_-|^2 + \frac{g^2}{2}\left(|\phi_+|^2 - |\phi_-|^2 - \xi\right)^2 \tag{V-2.123}$$

where g is the gauge coupling and ξ is a Fayet-Iliopoulos D-term (which we choose to be positive). This system has a unique supersymmetric vacuum with broken gauge symmetry

$$\phi_0 = \phi_- = 0\ , \quad |\phi_+| = \sqrt{\xi} \tag{V-2.124}$$

Minimizing the potential, for fixed values of ϕ_0, with respect to other fields, we find that for $|\phi_0| > \phi_c \equiv g\sqrt{\xi}/\lambda$, the minimum is at $\phi_+ = \phi_- = 0$. Thus, for $|\phi_0| > \phi_c$ and $\phi_+ = \phi_- = 0$ the tree level potential has a vanishing curvature in the ϕ_0 direction and large positive curvature in the remaining two directions ($m_\pm^2 = \lambda^2|\phi_0|^2 \mp g^2\xi$). Along the ϕ_0 direction ($|\phi_0| > \phi_c$, $\phi_+ = \phi_- = 0$), the tree level value of the potential remains constant: $V = g^2\xi^2/2 \equiv V_0$. Thus ϕ_0 provides a natural candidate for the inflaton field.

Along the inflationary trajectory all the F-terms vanish and the universe is dominated by the D-term which splits the masses of the Fermi-Bose components in the ϕ_+ and ϕ_- superfields. Such splitting results in a one-loop effective potential. In the present case this potential can be easily evaluated and for large ϕ_0 it behaves as

$$V_{eff} = \frac{g^2}{2}\xi^2\left(1 + \frac{g^2}{16\pi^2}\ln\frac{\lambda^2|\phi_0|^2}{\Lambda^2}\right) \equiv V_0\left(1 + \frac{Cg^2}{8\pi^2}\ln\frac{\lambda\varphi}{\Lambda}\right) \quad \text{(V-2.125)}$$

where $\varphi \equiv |\phi_0|$ and $C \sim 1$. Along this potential, the value of φ that leads to the right number $N \sim 50$ of e-foldings is:

$$\frac{\varphi}{m_{_P}} = \sqrt{\frac{NCg^2}{4\pi^2}} \quad \text{(V-2.126)}$$

This is safely of order g in the model that we consider. The values of the slowroll parameters (V-2.114) are correspondingly

$$\varepsilon = \frac{Cg^2}{32N\pi^2}\ , \quad \eta = -\frac{1}{2N} \quad \text{(V-2.127)}$$

Finally, the COBE normalisation (V-2.116) fixes the overall scale:

$$\xi^{1/2} \sim \left(\frac{C}{N}\right)^{1/4} \times 1.9\ 10^{16}\text{GeV} \quad \text{(V-2.128)}$$

The model can be generalized to supergravity [16].

V-3 Cosmological constant and dark energy

Late acceleration of the expansion. Over the last years, there has been an increasing number of indications that the Universe is presently undergoing accelerated expansion. This appears to be a strong departure from the standard picture of a matter-dominated Universe. Indeed, the equation for the conservation of energy,

$$\dot{\rho} = -3(p + \rho)H \quad \text{(V-3.129)}$$

allows to derive from the Friedmann equation (V-1.98), written in the case of a universe dominated by a component with energy density ρ and pressure p:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho + 3p) \qquad \text{(V-3.130)}$$

Obviously, a matter-dominated ($p_M \sim 0$) universe is decelerating. One needs instead a component with a negative pressure.

A cosmological constant is associated with a contribution to the energy-momentum tensor as in (V-1.100)(V-1.101):

$$T^\mu_\nu = -\Lambda^4 \delta^\mu_\nu = (-\rho, p, p, p) \qquad \text{(V-3.131)}$$

The associated equation of motion is therefore

$$p = -\rho \qquad \text{(V-3.132)}$$

It follows from (V-3.130) that a cosmological constant tends to accelerate expansion.

We have seen above that the energy density of the Universe seems to be dominated by a component of the type of a cosmological constant. This raises a new problem. Since matter and a cosmological constant evolve very differently, why should they be of the same order at present times? Indeed, for a component of equation of state $p = w\rho$, we may rewrite (V-3.129) as

$$\frac{\dot{\rho}}{\rho} = -3\frac{\dot{a}}{a}(1 + w) \quad . \qquad \text{(V-3.133)}$$

Thus matter ($p \sim 0$) energy density evolves as a^{-3} whereas a cosmological constant stays constant, as expected. Why should they be presently of similar magnitude (see Figure 10)? This is known as the *cosmic coincidence problem.*

This problem seems to indicate that there should be a dynamical origin to the solution of this late acceleration of the expansion. This may or may not be related with the solution of the cosmological constant problem. In the absence of any real clue for this latter problem, one has assumed that there is an unknown mechanisms that relaxes the vacuum energy. The reaccelaration of the expansion is then due to a new dynamical component of the energy density of the universe that has only emerged recently. This component is called *dark energy.* Just as in the case of inflation (which corresponds also to an acceleration), a natural candidate is provided by scalar fields.

Dark energy. We thus consider a dynamical component X with negative pressure (to provide acceleration; see (V-3.130)). Its equation

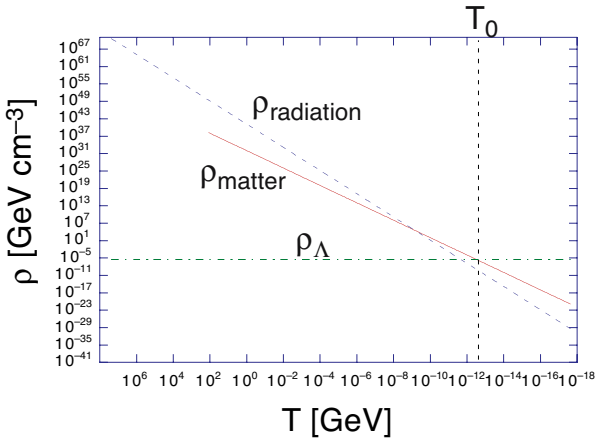

Figure 10. Evolution od radiation (dashed), matter (full) and cosmological constant (dot-dashed) energy density with the temperature T of the Universe.

of state thus reads [14]:

$$p_X = w_X \rho_X, \quad w < 0 \tag{V-3.134}$$

Experimental data may constrain such a dynamical component, just as it did with the cosmological constant. For example, in a spatially flat Universe with only matter and an unknown component X with equation of state $p_X = w_X \rho_X$, one obtains from (V-3.130) with $\rho = \rho_M + \rho_X$, $p = w_X \rho_X$ the following form for the deceleration parameter

$$q_0 \equiv -\frac{1}{H_0^2}\left(\frac{\ddot{a}}{a}\right)_{t=t_0} = \frac{\Omega_M}{2} + (1+3w_X)\frac{\Omega_X}{2} \tag{V-3.135}$$

where $\Omega_X = \rho_X/\rho_c$.

Another important property of dark energy is that it does not appear to be clustered (just as a cosmological constant). Otherwise, its effects would have been detected locally, as for the case of dark matter.

[14] We recall that non-relativistic matter (dust) has an equation of state $p \sim 0$ whereas $p = \rho/3$ corresponds to radiation.

We consider here a scalar field ϕ slowly evolving in its potential $V(\phi)$. Indeed, using (V-2.113) we have

$$w_\phi = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} \tag{V-3.136}$$

If the kinetic energy is subdominant ($\dot{\phi}^2/2 \ll V(\phi)$), we clearly obtain $-1 \leq w_\phi \leq 0$. A consequence is that the corresponding speed of sound $c_s = \delta p/\delta\rho$ is of the order of the speed of light: the scalar field pressure resists gravitational clustering.

We will see below that the scalar field must be extremely light. We therefore have two possible situations:

- a scalar potential slowly decreasing to zero as ϕ goes to infinity [38–40]. This is often referred to as *quintessence* or runaway qintessence.
- a very light field (pseudo-Goldstone boson) which is presently relaxing to its vacuum state [41].

In both cases one is relaxing to a position where the vacuum energy is zero. This is associated with our assumption that some unknown mechanism wipes the cosmological constant out. We discuss the two cases in turn.

i) Runaway quintessence

A runaway potential is frequently present in models where supersymmetry is dynamically broken. We have seen that supersymmetric theories are characterized by a scalar potential with many flat directions, *i.e.* directions ϕ in field space for which the potential vanishes. The corresponding degeneracy is lifted through dynamical supersymmetry breaking. In some instances (dilaton or compactification radius), the field expectation value $< \phi >$ actually provides the value of the strong interaction coupling. Then at infinite ϕ value, the coupling effectively goes to zero together with the supersymmetry breaking effects and the flat direction is restored: the potential decreases monotonically to zero as ϕ goes to infinity.

Let us take the example of supersymmetry breaking by gaugino condensation in effective superstring theories. The value g_0 of the gauge coupling at the string scale $M_{_S}$ is provided by the vacuum expectation value of the dilaton field s (taken to be dimensionless by dividing by $m_{_P}$) present among the massless string modes: $g_0^2 =< s >^{-1}$. If the gauge group has a one-loop beta function coefficient $b > 0$, then the running

gauge coupling becomes strong at the scale

$$\Lambda \sim M_S e^{-8\pi^2/(bg_0^2)} = M_S e^{-8\pi^2 s/b} \tag{V-3.137}$$

At this scale, the gaugino fields are expected to condense. Through dimensional analysis, the gaugino condensate $< \bar{\lambda}\lambda >$ is expected to be of order Λ^3. Terms quadratic in the gaugino fields thus yield in the effective theory below condensation scale a potential for the dilaton:

$$V \sim \left| < \bar{\lambda}\lambda > \right|^2 \propto e^{-48\pi^2 s/b} \tag{V-3.138}$$

The s-dependence of the potential is of course more complicated and one usually looks for stable minima with vanishing cosmological constant. But the behavior (V-3.137) is characteristic of the large s region and provides a potential slopping down to zero at infinity as required in the quintessence approach. A similar behavior is observed for moduli fields whose *vev* describes the radius of the compact manifolds which appear from the compactification from 10 or 11 dimensions to 4 in superstring theories.

Let us take therefore the example of an exponentially decreasing potential. More explicitly, we consider the following action

$$\mathcal{S} = \int d^4x\sqrt{g}\left[-\frac{m_P^2}{2}R + \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi)\right] \tag{V-3.139}$$

which describes a real scalar field ϕ minimally coupled with gravity and the self-interactions of which are described by the potential:

$$V(\phi) = V_0 e^{-\lambda\phi/m_P} \tag{V-3.140}$$

where V_0 is a positive constant.

The energy density and pressure stored in the scalar field are given by (V-2.113). We will assume that the background (matter and radiation) energy density ρ_B and pressure p_B obey a standard equation of state $p_B = w_B\rho_B$. The equation of motion for ϕ is again (V-2.112) with

$$H^2 = \frac{1}{3m_P^2}(\rho_B + \rho_\phi) \tag{V-3.141}$$

It can be rewritten as

$$\dot{\rho}_\phi = -3H\dot{\phi}^2 \tag{V-3.142}$$

We are looking for *scaling solutions i.e.* solutions where the ϕ energy density scales as a power of the cosmic scale factor: $\rho_\phi \propto a^{-n_\phi}$ or $\dot{\rho}_\phi/\rho_\phi = -n_\phi H$. In this case, one easily obtains from (V-2.113) and

(V-3.142) that the ϕ field obeys a standard equation of state $p_\phi = w_\phi \rho_\phi$, with

$$w_\phi = \frac{n_\phi}{3} - 1 \tag{V-3.143}$$

Hence

$$\rho_\phi \propto a^{-3(1+w_\phi)} \tag{V-3.144}$$

If one can neglect the background energy ρ_B, then (V-3.141) yields a simple differential equation for $a(t)$ which is solved as:

$$a \propto t^{2/[3(1+w_\phi)]} \tag{V-3.145}$$

Since $\dot{\phi}^2 = (1 + w_\phi)\rho_\phi \sim t^{-2}$, one deduces that ϕ varies logarithmically with time. One then easily obtains from (V-2.112,V-3.141) that

$$\phi = \phi_0 + \frac{2}{\lambda} m_{_P} \ \ln(t/t_0). \tag{V-3.146}$$

and[15]

$$w_\phi = \frac{\lambda^2}{3} - 1 \ , \tag{V-3.147}$$

It is clear from (V-3.147) that, for λ sufficiently small, the field ϕ can play the role of quintessence. We note that, even if we started with a small value ϕ_o, ϕ reaches a value of order $m_{_P}$.

But the successes of the standard big-bang scenario indicate that clearly ρ_ϕ cannot have always dominated: it must have emerged from the background energy density ρ_B. If we consider initial conditions where ρ_B dominates, the attractor solution will generally yield $w_\phi = w_B$ and thus no dark energy solution.

Ways to obtain a quintessence component have been proposed however. Let us sketch some of them in turn.

One is the notion of *tracker field* [42]. This idea also rests on the existence of scaling solutions of the equations of motion which play the role of late time attractors, as illustrated above. An example is provided by a scalar field described by the action (V-3.139) with a potential

$$V(\phi) = \lambda \ \frac{\Lambda^{4+\alpha}}{\phi^\alpha} \tag{V-3.148}$$

with $\alpha > 0$. In the case where the background density dominates, one finds an attractor scaling solution [39] $\phi \propto a^{3(1+w_B)/(2+\alpha)}$, $\rho_\phi \propto$

[15] under the condition $\lambda^2 \leq 6$ ($w_\phi \leq 1$ since $V(\phi) \geq 0$).

$a^{-3\alpha(1+w_B)/(2+\alpha)}$. Thus ρ_ϕ decreases at a slower rate than the background density ($\rho_B \propto a^{-3(1+w_B)}$) and tracks it until it becomes of the same order at a given value a_Q. We thus have:

$$\frac{\phi}{m_{_P}} \sim \left(\frac{a}{a_Q}\right)^{3(1+w_B)/(2+\alpha)}, \tag{V-3.149}$$

$$\frac{\rho_\phi}{\rho_B} \sim \left(\frac{a}{a_Q}\right)^{6(1+w_B)/(2+\alpha)} \tag{V-3.150}$$

One finds

$$w_\phi = -1 + \frac{\alpha(1+w_B)}{2+\alpha} \tag{V-3.151}$$

Shortly after ϕ has reached for $a = a_Q$ a value of order m_P, it satisfies the standard slow roll conditions and therefore (V-3.151) provides a good approximation to the present value of w_ϕ. Thus, at the end of the matter-dominated era, this field may provide the quintessence component that we are looking for.

Two features are interesting in this respect. One is that this scaling solution is reached for rather general initial conditions, *i.e.* whether ρ_ϕ starts of the same order or much smaller than the background energy density [42]. Regarding the cosmic coincidence problem, it can be rephrased here as follows (since ϕ is of order $m_{_P}$ in this scenario): why is $V(m_{_P})$ of the order of the critical energy density ρ_c? It is thus the scale Λ which determines the time when the scalar field starts to emerge and the universe expansion reaccelerates. Indeed, using (V-3.150), the constraint reads:

$$\Lambda \sim \left(H_0^2 m_{_P}^{2+\alpha}\right)^{1/(4+\alpha)} \tag{V-3.152}$$

We may note that this gives for $\alpha = 2$, $\Lambda \sim 10$ MeV, not such an atypical scale for high energy physics.

A model [43] has been proposed which goes one step further: the dynamical component, a scalar field, is called k-essence and the model is based on the property observed in string models that scalar kinetic terms may have a non-trivial structure. Tracking occurs only in the radiation-dominated era; a new attractor solution where quintessence acts as a cosmological constant is activated by the onset of matter domination.

Models of dynamical supersymmetry breaking easily provide a model of the tracker field type just discussed [44]. The corresponding scalar field is then a condensate of fermions.

Quintessential problems

However appealing, the quintessence idea is difficult to implement in the context of realistic models [45, 46]. The main problem lies in the

fact that the quintessence field must be extremely weakly coupled to ordinary matter. This problem can take several forms:

• we have assumed until now that the quintessence potential monotonically decreases to zero at infinity. In realistic cases, this is difficult to achieve because the couplings of the field to ordinary matter generate higher order corrections that are increasing with larger field values, unless forbidden by a symmetry argument. Also, because the *vev* of ϕ is of order $m_{_P}$, one must take into account the full supergravity corrections. One may then argue [47] that this could put in jeopardy the positive definiteness of the scalar potential, a key property of the quintessence potential. This may point towards models where $< W >= 0$ or to no-scale type models.

• the quintessence field must be very light. If we return to our example in (V-3.148), $V''(m_P)$ provides an order of magnitude for the mass-squared of the quintessence component:

$$m_\phi \sim \Lambda \left(\frac{\Lambda}{m_P}\right)^{1+\alpha/2} \sim H_0 \sim 10^{-33} \text{ eV} \qquad \text{(V-3.153)}$$

using (V-3.152). This might argue for a pseudo-Goldstone boson nature of the scalar field that plays the rôle of quintessence. This field must in any case be very weakly coupled to matter; otherwise its exchange would generate observable long range forces. Eötvös-type experiments put very severe constraints on such couplings.

• it is difficult to find a symmetry that would prevent any coupling of the form $\beta(\phi/m_P)^n F^{\mu\nu}F_{\mu\nu}$ to the gauge field kinetic term. Since the quintessence behavior is associated with time-dependent values of the field of order m_P, this would generate, in the absence of fine tuning, corrections of order one to the gauge coupling. But we have seen in subsection V-1 that the time dependence of the fine structure constant for example is very strongly constrained: $|\dot{\alpha}/\alpha| < 5 \times 10^{-17}\text{yr}^{-1}$. This yields a limit [45]:

$$|\beta| \leq 10^{-6} \frac{m_P H_0}{< \dot{\phi} >} \qquad \text{(V-3.154)}$$

where $< \dot{\phi} >$ is the average over the last 2×10^9 years.

ii) Pseudo-Goldstone boson

There exists a class of models [41] very close in spirit to the case of runaway quintessence: they correspond to a situation where a scalar

field has not yet reached its stable groundstate and is still evolving in its potential.

More specifically, let us consider a potential of the form:

$$V(\phi) = M^4 v\left(\frac{\phi}{f}\right) \tag{V-3.155}$$

where M is the overall scale, f is the vacuum expectation value $<\phi>$ and the function v is expected to have coefficients of order one. If we want the potential energy of the field (assumed to be close to its *vev* f) to give a substantial fraction of the energy density at present time, we must set

$$M^4 \sim \rho_c \sim H_0^2 m_P^2 \tag{V-3.156}$$

However, requiring that the evolution of the field ϕ around its minimum has been overdamped by the expansion of the Universe until recently imposes

$$m_\phi^2 = \frac{1}{2}V''(f) \sim \frac{M^4}{f^2} \leq H_0^2. \tag{V-3.157}$$

Let us note that this is again one of the slowroll conditions familiar to the inflation scenarios.

From (V-3.156) and (V-3.157), we conclude that f is of order $m_{_P}$ (as the value of the field ϕ in runaway quintessence) and that $M \sim 10^{-3}$ eV (not surprisingly, this is the scale Λ typical of the cosmological constant, see (V-1.102)). As we have seen, the field ϕ must be very light: $m_\phi \sim h_0 \times 10^{-60} m_P \sim h_0 \times 10^{-33}$ eV. Such a small value is only natural in the context of an approximate symmetry: the field ϕ is then a pseudo-Goldstone boson. A typical example of such a field is provided by the string axion field. In this case, the potential simply reads:

$$V(\phi) = M^4\left[1 + \cos(\phi/f)\right] \tag{V-3.158}$$

All the preceding shows that there is extreme fine tuning in the couplings of the quintessence field to matter, unless they are forbidden by some symmetry. This is somewhat reminiscent of the fine tuning associated with the cosmological constant. In fact, and as already stressed, the quintessence solution does not claim to solve the cosmological constant (vacuum energy) problem.

To conclude, we hope to have convinced the reader of the wealth of possibilities offered by fundamental scalar fields. Obviously we are still waiting for the experimental discovery of such a fundamental scalar.

But the probable discovery of a Higgs particle at the forthcoming LHC should provide the decisive argument for the relevance of fundamental scalars in nature.

References

[1] B. Lee and C. Quigg and H. Thacker, Phys. Rev. D16 (1977) 1519

[2] W. Marciano and G. Valencia and S. Willenbrock, Phys. Rev., D40 (1989) 1725-1729

[3] M. Chanowitz and M.K. Gaillard, Nucl. Phys. B261 (1985) 379-431

[4] N. Cabibbo and L. Maiani and G. Parisi and R. Petronzio, Nucl. Phys. B158 (1979) 295-305

[5] M. Lindner, Z. Phys., C31 (1986) 295-300

[6] E. Gildener, Phys. Rev. D14 (1976) 167

[7] S. Weinberg, Phys. Rev. D19 (1979) 1277

[8] L. Susskind, Phys. Rev. D20 (1979)

[9] M. Veltman, Nucl. Phys., (1981)

[10] C. Kolda and H. Murayama, JHEP 0007 (2000) 035

[11] G. 't Hooft, Recent developments in gauge theories (NATO ASI Series B: Physics Vol. 59 (1979) Plenum Press

[12] J. Goldstone, Nuovo Cimento, 19 (1961) 154

[13] C.G. Callan and S. Coleman and R. Jackiw, Ann. Phys. (N.Y.) 59 (1970) 42

[14] N. Arkani-Hamed and A.G. Cohen and E. Katz and A.E. Nelson, JHEP 0207 (2002) 034

[15] C. Csaki and Hubisz and Kribs and Meadle and J. Terning, Phys. Rev. D68 (2003) 035009 [arXiv:hep-ph/0303236]

[16] P. Binétruy and G. Dvali and R. Kallosh and A. van Proeyen, Class. Quant. Grev., 21 (2004) 3137

[17] T. Kaluza, Sitzungsber. Preuss Akad. Wiss. Berlin Math. Phys. (1921)

[18] O. Klein, Z. Phys., 37 (1926) 895

[19] E. Witten, Nucl. Phys., B268 (1986) 79

[20] E. Martinec, Phys. Lett., B171 (1986) 189

[21] M. Dine and N. Seiberg, Phys. Rev. Lett., 57 (1986) 2625

[22] A. Einstein, Sitzungsber. Preuss. Akad. Wiss., "phus.-math. Klasse VI", (1917) 142

[23] A.H. Guth, Phys. Rev. D23 (1981) 347

[24] L.P. Grishchuk and Y.V. Sidorov, Phys. Rev., D42 (1990) 3413

[25] C.M. Will, Theory and experiment in gravitational physics, Cambridge University Press, 1993

[26] T. Damour and K. Nordtvedt, Phys. Rev. Lett., 70 (1993) 2217-2219

[27] P.A.M. Dirac, Nature (London), 139 (1937) 323

[28] P.A.M. Dirac, Proc. Roy. Soc. London, A165 (1938) 198

[29] J.-Ph. Uzan, Rev. Mod. Phys., 75 (2003) 403

[30] T. Damour and F.J. Dyson, Nucl. Phys., B480 (1996) 37

[31] G.D. Coughlan and W. Fischler and E.W. Kolb and S. Raby and G.G. Ross, Phys. Lett. B131 (1983) 59-64

[32] A.S. Goncharov and A.D. Linde and M.J. Visotsky, Phys. Lett., B147 (1984) 279-283

[33] J. Ellis and D.V. Nanopoulos and M. Quiros, Phys. Lett. B174 (1986) 176-182

[34] A. Cohen and D. Kaplan and A. Nelson, Phys. Lett. B388 (1996) 588-598

[35] E.D. Stewart, Phys. Rev. D51 (1995) 6847-6853

[36] P. Binétruy and G. Dvali, Phys. Lett. B388 (1996) 241

[37] E. Halyo, Phys. Lett., B387 (1996) 43

[38] C. Wetterich, Nucl. Phys., B302 (1988) 668

[39] B. Ratra and P.J.E. Peebles, Phys. Rev., D37 (1988) 3406

[40] R.R. Caldwell and R. Dave and P.J. Steinhardt, Phys. Rev. Lett., 75 (1995) 2077

[41] J. Frieman and C. Hill and A. Stebbins and I. Waga, Phys. Rev. Lett., 75 (1995) 2077

[42] I. Zlatev and L. Wang and P.J. Steinhardt, Phys. Rev. Lett., 82 (1999) 896

[43] C. Armendariz-Pico and V. Mukhanov and P.J. Steinhardt, Phys. Rev. Lett. 85 (2000) 4438-4431

[44] P. Binétruy, Phys. Rev., D60 (1999) 063502

[45] S.M. Carroll, Phys. Rev. Lett., 81 (1998) 3067

[46] C. Kolda and D. Lyth, Phys. Lett., B458 (1999) 197

[47] P. Brax and J. Martin, Phys. Lett. B468 (1999) 40

INFLATION, QUANTUM FLUCTUATIONS AND COSMOLOGICAL PERTURBATIONS

David Langlois
GRECO, Institut d'Astrophysique de Paris (CNRS)
98bis Boulevard Arago, 75014 Paris, France
langlois@iap.fr

Abstract These lectures are intended to give a pedagogical introduction to the main current picture of the very early universe. After elementary reviews of general relativity and of the standard Big Bang model, the following subjects are discussed: inflation, the classical relativistic theory of cosmological perturbations and the generation of perturbations from scalar field quantum fluctuations during inflation.

Keywords: Cosmological scalar fields, inflation, slow-roll, potentials, cosmological perturbations, Bardeen potentials, initial conditions, quantum fluctuations, power spectra.

I. Introduction

The purpose of these lectures is to give an introduction to the present *standard picture of the early universe*, which *complements* the older standard Big Bang model. These notes are intended for non-experts on this subject. They start with a very short introduction to General Relativity, on which modern cosmology is based, followed by an elementary review of the standard Big Bang model. We then discuss the limitations of this model and enter into the main subject of these lectures: *inflation*.

Inflation was initially invented to solve some of the problems of the standard Big Bang model and to get rid of unwanted relics generically predicted by high energy models. It turned out that inflation, as it was realized later, could also solve an additional puzzle of the standard model, that of the generation of the cosmological perturbations. This welcome surprise put inflation on a rather firm footing, about twenty years ago. Twenty years later, inflation is still alive, in a stronger position than ever because its few competitors have been eliminated as new cosmological observations have accumulated during the last few years.

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 235–278.

II. A few elements on general relativity and cosmology

Modern cosmology is based on Einstein's theory of general relativity. It is thus useful, before discussing the early universe, to recall a few notions and useful formulas from this theory. Details can be found in standard textbooks on general relativity (see e.g. [1]). In the framework of general relativity, the spacetime geometry is defined by a *metric*, a symmetric tensor with two indices, whose components in a coordinate system $\{x^\mu\}$ ($\mu = 0, 1, 2, 3$) will be denoted $g_{\mu\nu}$. The square of the "distance" between two neighbouring points of spacetime is given by the expression

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{II-0.1}$$

We will use the signature $(-, +, +, +)$.

In a coordinate change $x^\mu \to \tilde{x}^\mu$, the new components of the metric are obtained by using the standard tensor transformation formulas, namely

$$\tilde{g}_{\mu\nu} = \frac{\partial x^\rho}{\partial \tilde{x}^\mu} \frac{\partial x^\sigma}{\partial \tilde{x}^\nu} g_{\rho\sigma}. \tag{II-0.2}$$

One can define a *covariant derivative* associated to this metric, denoted D_μ, whose action on a tensor with, for example, one covariant index and one contravariant index will be given by

$$D_\lambda T^\mu{}_\nu = \partial_\lambda T^\mu{}_\nu + \Gamma^\mu_{\lambda\sigma} T^\sigma{}_\nu - \Gamma^\sigma_{\lambda\nu} T^\mu{}_\sigma \tag{II-0.3}$$

(a similar term must be added for each additional covariant or contravariant index), where the Γ are the *Christoffel symbols* (they are not tensors), defined by

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \right). \tag{II-0.4}$$

We have used the notation $g^{\mu\nu}$ which corresponds, for the metric (and only for the metric), to the inverse of $g_{\mu\nu}$ in a matricial sense, i.e. $g_{\mu\sigma} g^{\sigma\nu} = \delta^\nu_\mu$.

The "curvature" of spacetime is characterized by the *Riemann* tensor, whose components can be expressed in terms of the Christoffel symbols according to the expression

$$R_{\lambda\mu\nu}{}^\rho = \partial_\mu \Gamma^\rho_{\lambda\nu} - \partial_\lambda \Gamma^\rho_{\mu\nu} + \Gamma^\sigma_{\lambda\nu} \Gamma^\rho_{\sigma\mu} - \Gamma^\sigma_{\mu\nu} \Gamma^\rho_{\sigma\lambda}. \tag{II-0.5}$$

Einstein's equations relate the spacetime geometry to its matter content. The geometry appears in Einstein's equations via the *Ricci tensor*, defined by

$$R_{\mu\nu} = R_{\mu\sigma\nu}{}^\sigma, \tag{II-0.6}$$

and the *scalar curvature*, which is the trace of the Ricci tensor, i.e.

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{II-0.7}$$

The matter enters Einstein's equations via the *energy-momentum tensor*, denoted $T_{\mu\nu}$, whose time/time component corresponds to the energy density, the time/space components to the momentum density and the space/space component to the stress tensor. Einstein's equations then read

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = 8\pi G\, T_{\mu\nu}, \tag{II-0.8}$$

where the tensor $G_{\mu\nu}$ is called the *Einstein tensor*. Since, by construction, the Einstein tensor satisfies the identity $D_\mu G^\mu_\nu = 0$, any energy-momentum on the right-hand side of Einstein's equation must necessarily satisfy the relation

$$D_\mu T^\mu_\nu = 0, \tag{II-0.9}$$

which can be interpreted as a generalization, in the context of a curved spacetime, of the familiar conservation laws for energy and momentum.

The motion of a particule is described by its trajectory in spacetime, $x^\mu(\lambda)$, where λ is a parameter. A free particle, i.e. which does not feel any force (other than gravity), satisfies the *geodesic equation*, which reads

$$t^\sigma D_\sigma t^\mu = 0, \tag{II-0.10}$$

where $t^\mu = dx^\mu/d\lambda$ is the vector field tangent to the trajectory (note that the geodesic equation written in this form assumes that the parameter λ is affine). Equivalently, the geodesic can be rewritten as

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma} \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0. \tag{II-0.11}$$

The geodesic equation applies both to

- massive particles, in which case one usually takes as the parameter λ the so-called *proper time* so that the corresponding tangent vector u^μ is normalized: $g_{\mu\nu} u^\mu u^\nu = -1$;
- massless particles, in particular the photon, in which case the tangent vector, usually denoted k^μ is light-like, i.e. $g_{\mu\nu} k^\mu k^\nu = 0$.

Einstein's equations can also be obtained from a variational principle. The corresponding action reads

$$\mathcal{S} = \frac{1}{16\pi G} \int d^4x \sqrt{-g}\, (R - 2\Lambda) + \int d^4x \sqrt{-g} \mathcal{L}_{mat}. \tag{II-0.12}$$

One can check that the variation of this action with respect to the metric $g_{\mu\nu}$, upon using the definition

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta\left(\sqrt{-g}\mathcal{L}_{mat}\right)}{\delta g_{\mu\nu}}, \tag{II-0.13}$$

indeed gives Einstein's equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu}. \tag{II-0.14}$$

This is a slight generalization of Einstein's equations (II-0.8) that includes a *cosmological constant* Λ. It is worth noticing that the cosmological constant can also be interpreted as a particular energy-momentum tensor of the form $T_{\mu\nu} = -(8\pi G)^{-1}\Lambda g_{\mu\nu}$.

II-1 Review of standard cosmology

In this subsection, the foundations of modern cosmology are briefly recalled. They follow from Einstein's equations introduced above and from a few hypotheses concerning spacetime and its matter content. One of the essential assumptions of cosmology (so far confirmed by observations) is to consider, as a first approximation, the universe as being homogeneous and isotropic. Note that these symmetries define implicitly a particular "slicing" of spacetime, the corresponding space-like hypersurfaces being homogeneous and isotropic. A different slicing of the *same* spacetime will give in general space-like hypersurfaces that are not homogeneous and isotropic.

The above hypothesis turns out to be very restrictive and the only metrics compatible with this requirement reduce to the so-called *Robertson-Walker* metrics, which read in an appropriate coordinate system

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{II-1.15}$$

with $\kappa = 0, -1, 1$ depending on the curvature of spatial hypersurfaces: respectively flat, elliptic or hyperbolic.

The matter content compatible with the spacetime symmetries of homogeneity and isotropy is necessarily described by an energy-momentum tensor of the form (in the same coordinate system as for the metric (II-1.15)):

$$T^{\mu}_{\nu} = \mathrm{Diag}\left(-\rho(t), p(t), p(t), p(t)\right). \tag{II-1.16}$$

The quantity ρ corresponds to an energy density and P to a pressure.

One can show that the so-called comoving particles, i.e. those particles whose spatial coordinates are constant in time, satisfy the geodesic equation (II-0.11).

II-2 Friedmann-Lematre equations

Substituting the Robertson-Walker metric (II-1.15) in Einstein's equations (II-0.8), one gets the so-called *Friedmann-Lematre equations*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{\kappa}{a^2}, \tag{II-2.17}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3P\right). \tag{II-2.18}$$

An immediate consequence of these two equations is the *continuity equation*

$$\dot{\rho} + 3H\left(\rho + p\right) = 0, \tag{II-2.19}$$

where $H \equiv \dot{a}/a$ is the *Hubble parameter*. The continuity equation can be also obtained directly from the energy-momentum conservation $D_\mu T^\mu_\nu = 0$, as mentioned before.

In order to determine the cosmological evolution, it is easier to combine (II-2.17) with (II-2.19). Let us assume an equation of state for the cosmological matter of the form $p = w\rho$ with w constant. This includes the two main types of matter that play an important rôle in cosmology:

- gas of relativistic particles, $w = 1/3$;
- non relativistic matter, $w \simeq 0$.

In these cases, the conservation equation (II-2.19) can be integrated to give

$$\rho \propto a^{-3(1+w)}. \tag{II-2.20}$$

Substituting in (II-2.17), one finds, for $\kappa = 0$,

$$3\frac{\dot{a}^2}{a^2} = 8\pi G\rho_0 \left(\frac{a}{a_0}\right)^{-3(1+w)}, \tag{II-2.21}$$

where, by convention, the subscript '0' stands for *present* quantities. One thus finds $\dot{a}^2 \propto a^{2-3(1+w)}$, which gives for the evolution of the scale factor

- in a universe dominated by non relativistic matter

$$a(t) \propto t^{2/3}, \tag{II-2.22}$$

- and in a universe dominated by radiation

$$a(t) \propto t^{1/2}. \tag{II-2.23}$$

One can also mention the case of a *cosmological constant*, which corresponds to an equation of state $w = -1$ and thus implies an exponential evolution for the scale factor

$$a(t) \propto \exp(Ht). \tag{II-2.24}$$

More generally, when several types of matter coexist with respectively $p_{(i)} = w_{(i)}\rho_{(i)}$, it is convenient to introduce the dimensionless parameters

$$\Omega_{(i)} = \frac{8\pi G\rho_0^{(i)}}{3H_0^2}, \tag{II-2.25}$$

which express the *present* ratio of the energy density of some given species with respect to the so-called *critical energy density* $\rho_{crit} = 3H_0^2/(8\pi G)$, which corresponds to the total energy density for a flat universe.

One can then rewrite the first Friedmann equation (II-2.17) as

$$\left(\frac{H}{H_0}\right)^2 = \sum_i \Omega_{(i)} \left(\frac{a}{a_0}\right)^{-3(1+w_{(i)})} + \Omega_\kappa \left(\frac{a}{a_0}\right)^{-2}, \tag{II-2.26}$$

with $\Omega_\kappa = -\kappa/a_0^2H_0^2$, which implies that the cosmological parameters must satisfy the consistency relation

$$\sum_i \Omega_{(i)} + \Omega_\kappa = 1. \tag{II-2.27}$$

As for the second Friedmann equation (II-2.18), it implies

$$\frac{\ddot{a}_0}{a_0H_0^2} = -\frac{1}{2}\sum_i \Omega_{(i)}(1 + w_{(i)}). \tag{II-2.28}$$

Present cosmological observations yield for the various parameters

- Baryons: $\Omega_b \simeq 0.05$,
- Dark matter: $\Omega_d \simeq 0.25$,
- Dark energy (compatible with a cosmological constant): $\Omega_\Lambda \simeq 0.7$,
- Photons: $\Omega_\gamma \simeq 5 \times 10^{-5}$.

Observations have not detected so far any deviation from flatness. Radiation is very subdominant today but extrapolating backwards in time, radiation was dominant in the past since its energy density scales as $\rho_\gamma \propto a^{-4}$ in contrast with non relativistic matter ($\rho_m \propto a^{-3}$). Moreover, since the present matter content seems dominated by dark energy similar to a cosmological constant ($w_\Lambda = -1$), this indicates that our universe is presently accelerating.

II-3 The cosmological redshift

An important consequence of the expansion of the universe is the *cosmological redshift* of photons. This is in fact how the expansion of the universe was discovered initially.

Let us consider two light signals emitted by a comoving object at two successive instants t_e and $t_e + \delta t_e$, then received later at respectively t_o and $t_o + \delta t_o$ by a (comoving) observer. One can always set the observer at the center of the coordinate system. All light trajectories reaching the observer are then radial and one can write, using (II-1.15)

$$\int_0^{r_e} \frac{dr}{\sqrt{1-\kappa r^2}} = \int_{t_e}^{t_o} \frac{dt}{a(t)}. \tag{II-3.29}$$

The left-hand side being identical for the two successive trajectories, the right-hand side must vanish, which yields

$$\frac{\delta t_o}{a_o} - \frac{\delta t_e}{a_e} = 0. \tag{II-3.30}$$

This implies for the frequencies measured at emission and at reception a redshift given by

$$1 + z \equiv \frac{\nu_e}{\nu_o} = \frac{a_o}{a_e}. \tag{II-3.31}$$

II-4 Thermal history of the universe

To go beyond a simply geometrical description of cosmology, it is very fruitful to apply thermodynamics to the matter content of the universe. One can then define a temperature T for the cosmological photons, not only when they are strongly interacting with ordinary matter but also after they have decoupled because, with the expansion, the thermal distribution for the gas of photons is unchanged except for a global rescaling of the temperature so that T essentially evolves as

$$T(t) \propto \frac{1}{a(t)}. \tag{II-4.32}$$

This means that, going backwards in time, the universe was hotter and hotter. This is the essence of the hot Big Bang scenario.

As the universe evolves, the reaction rates between the various species are modified. A detailed analysis of these changes enables to reconstruct the past thermal history of the universe. Two events in particular play an essential rôle because of their observational consequences:

- Primordial nucleosynthesis

Nucleosynthesis occured at a temperature around 0.1 MeV, when the average kinetic energy became sufficiently low so that nuclear binding was possible. Protons and neutrons could then combine, which lead to the production of light elements, such that Helium, Deuterium, Lithium, etc... Within the observational uncertainties, this scenario is remarkably confirmed by the present measurements.

- Decoupling of baryons and photons (or last scattering)

  A more recent event is the so-called "recombination" of nuclei and electrons to form atoms. This occured at a temperature of the order of the eV. Free electrons thus almost disappeared, which entailed an effective decoupling of the cosmological photons and ordinary matter. What we see today as the Cosmic Microwave Background (CMB) is made of the fossil photons, which interacted for the last time with matter at the last scattering epoch. The CMB represents a remarkable observational tool for analysing the perturbations of the early universe, as well as for measuring the cosmological parameters introduced above.

II-5 Puzzles of the standard Big Bang model

The standard Big Bang model has encountered remarkable successes, in particular with the nucleosynthesis scenario and the prediction of the CMB, and it remains today a cornerstone in our understanding of the present and past universe. However, a few intriguing facts remain unexplained in the strict scenario of the standard Big Bang model and seem to necessitate a larger framework. We review below the main problems:

- Homogeneity problem

  A first question is why the approximation of homogeneity and isotropy turns out to be so good. Indeed, inhomogeneities are unstable, because of gravitation, and they tend to grow with time. It can be verified for instance with the CMB that inhomogeneities were much smaller at the last scattering epoch than today. One thus expects that these homogeneities were still smaller further back in time. How to explain a universe so smooth in its past ?

- Flatness problem

Another puzzle lies in the (spatial) flatness of our universe. Indeed, Friedmann's equation (II-2.17) implies

$$\Omega - 1 \equiv \frac{8\pi G\rho}{3H^2} - 1 = \frac{\kappa}{a^2 H^2}. \tag{II-5.33}$$

In standard cosmology, the scale factor behaves like $a \sim t^p$ with $p < 1$ ($p = 1/2$ for radiation and $p = 2/3$ for non-relativistic matter). As a consequence, $(aH)^{-2}$ grows with time and $|\Omega - 1|$ must thus diverge with time. Therefore, in the context of the standard model, the quasi-flatness observed today requires an extreme fine-tuning of Ω near 1 in the early universe.

- Horizon problem

  One of the most fundamental problems in standard cosmology is certainly the *horizon problem*. The (particle) *horizon* is the maximal distance that can be covered by a light ray. For a light-like radial trajectory $dr = a(t)dt$ and the horizon is thus given by

$$d_H(t) = a(t) \int_{t_i}^{t} \frac{dt'}{a(t')} = a(t) \frac{t^{1-q} - t_i^{1-q}}{1-q}, \tag{II-5.34}$$

  where the last equality is obtained by assuming $a(t) \sim t^q$ and t_i is some initial time.

  In standard cosmology ($q < 1$), the integral converges in the limit $t_i = 0$ and the horizon has a finite size, of the order of the so-called Hubble radius H^{-1}:

$$d_H(t) = \frac{q}{1-q} H^{-1}. \tag{II-5.35}$$

  It also useful to consider the *comoving Hubble radius*, $(aH)^{-1}$, which represents the fraction of comoving space in causal contact. One finds that it *grows* with time, which means that the *fraction of the universe in causal contact increases with time* in the context of standard cosmology. But the CMB tells us that the Universe was quasi-homogeneous at the time of last scattering on a scale encompassing many regions a priori causally independent. How to explain this ?

A solution to the horizon problem and to the other puzzles is provided by the inflationary scenario, which we will examine in the next section. The basic idea is to invert the behaviour of the comoving Hubble radius,

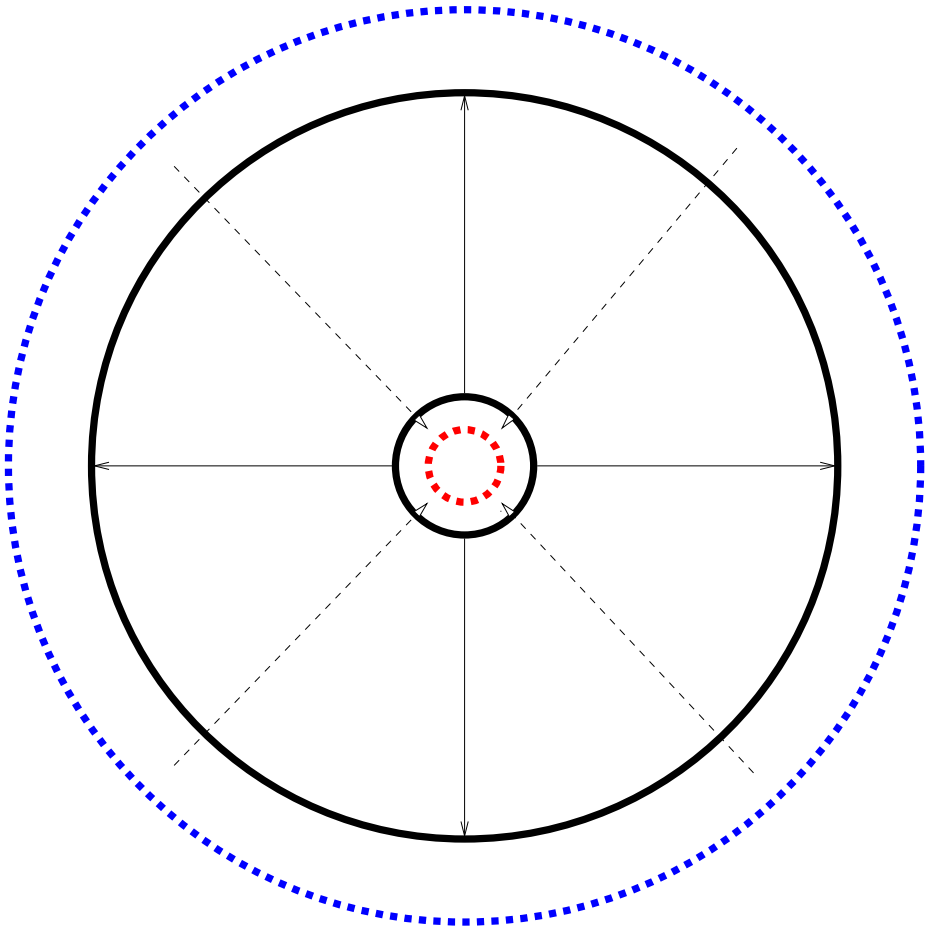

Figure 1. Evolution of the comoving Hubble radius $\lambda_H = (aH)^{-1}$: during standard cosmology, λ_H increases (continuous lines), whereas during inflation λ_H shrinks (dashed lines).

that is to make him *decrease* sufficiently in the very early universe. The corresponding condition is that

$$\ddot{a} > 0, \tag{II-5.36}$$

i.e. that the Universe must undergo a *phase of acceleration*.

III. Inflation

The broadest definition of inflation is that it corresponds to a phase of acceleration of the universe,

$$\ddot{a} > 0. \tag{III-0.37}$$

In this broad sense, the current cosmological observations, if correctly interpreted, mean that our present universe is undergoing an inflationary phase. We are however interested here in an inflationary phase taking place in the very early universe, with different energy scales.

The Friedmann equations (II-2.17) tell us that one can get acceleration only if the equation of state satisfies the condition

$$P < -\frac{1}{3}\rho, \tag{III-0.38}$$

condition which looks at first view rather exotic.

A very simple example giving such an equation of state is a cosmological constant, corresponding to a cosmological fluid with the equation of state

$$P = -\rho. \tag{III-0.39}$$

However, a strict cosmological constant leads to exponential inflation *forever* which cannot be followed by a radiation or matter era. Another possibility is a scalar field, which we discuss now in some details.

III-1 Cosmological scalar fields

The dynamics of a scalar field coupled to gravity is governed by the action

$$S_\phi = \int d^4x\sqrt{-g}\left(-\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi)\right). \tag{III-1.40}$$

The corresponding energy-momentum tensor, which can be derived using (II-0.13), is given by

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\partial^\sigma\phi\partial_\sigma\phi + V(\phi)\right). \tag{III-1.41}$$

If one assumes the geometry, and thus the matter, to be homogeneous and isotropic, then the energy-momentum tensor reduces to the perfect fluid form with the energy density

$$\rho = -T_0^0 = \frac{1}{2}\dot{\phi}^2 + V(\phi), \tag{III-1.42}$$

where one recognizes the sum of a kinetic energy and of a potential energy, and the pressure

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{III-1.43}$$

The equation of motion for the scalar field is the Klein-Gordon equation, which is obtained by taking the variation of the above action (III-1.40) with respect to the scalar field and which reads

$$D^\mu D_\mu \phi = V', \tag{III-1.44}$$

in general and

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0 \tag{III-1.45}$$

in the particular case of a FLRW (Friedmann-Lematre-Robertson-Walker) universe.

The system of equations governing the dynamics of the scalar field and of the geometry in a FLRW universe is thus given by

$$H^2 = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right), \tag{III-1.46}$$

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0, \tag{III-1.47}$$

$$\dot{H} = -4\pi G\dot{\phi}^2. \tag{III-1.48}$$

The last equation can be derived from the first two and is therefore redundant.

III-2 The slow-roll regime

The dynamical system (III-1.46-III-1.48) does not always give an accelerated expansion but it does so in the so-called *slow-roll regime* when the potential energy of the scalar field dominates over its kinetic energy.

More specifically, the so-called *slow roll* approximation consists in neglecting the kinetic energy of the scalar field , $\dot{\phi}^2/2$ in (III-1.46) and the acceleration $\ddot{\phi}$ in the Klein-Gordon equation (III-1.47). One then gets the simplified system

$$H^2 \simeq \frac{8\pi G}{3}V, \tag{III-2.49}$$

$$3H\dot{\phi} + V' \simeq 0. \tag{III-2.50}$$

Let us now examine in which regime this approximation is valid. From (III-2.50), the velocity of the scalar field is given by

$$\dot{\phi} \simeq -\frac{V'}{3H}. \tag{III-2.51}$$

Substituting this relation in the condition $\dot{\phi}^2/2 \ll V$ yields the requirement:

$$\epsilon_V \equiv \frac{m_P^2}{2}\left(\frac{V'}{V}\right)^2 \ll 1, \tag{III-2.52}$$

where we have introduced the *reduced Planck mass*

$$m_P \equiv \frac{1}{\sqrt{8\pi G}}. \tag{III-2.53}$$

Similarly, the time derivative of (III-2.51), using the time derivative of (III-2.49), gives, after substitution in $\ddot{\phi} \ll V'$, the condition

$$\eta_V \equiv m_P^2 \frac{V''}{V} \ll 1. \tag{III-2.54}$$

In summary, the slow-roll approximation is valid when the two conditions $\epsilon_V, \eta_V \ll 1$ are satisfied, which means that the slope and the curvature of the potential, in Planck units, must be sufficiently small.

III-3 Number of e-folds

When working with a specific inflationary model, it is important to be able to relate the cosmological scales observed at the present time with the scales during inflation. For this purpose, one usally introduces the *number of e-foldings before the end of inflation*, denoted N, and simply defined by

$$N = \ln \frac{a_{end}}{a}, \tag{III-3.55}$$

where a_{end} is the value of the scale factor at the end of inflation and a is a fiducial value for the scale factor during inflation. By definition, N *decreases* during the inflationary phase and reaches zero at its end. In the slow-roll approximation, it is possible to express N as a function of the scalar field. Since $dN = -d\ln a = -Hdt = -(H/\dot{\phi})d\phi$, one easily finds, using (III-2.51) and (III-2.49), that

$$N(\phi) \simeq \int_{\phi}^{\phi_{end}} \frac{V}{m_P^2 V'} d\phi. \tag{III-3.56}$$

Given an explicit potential $V(\phi)$, one can in principle integrate the above expression to obtain N in terms of ϕ. This will be illustrated below for some specific models.

Let us now discuss the link between N and the present cosmological scales. Let us consider a given scale characterized by its comoving wavenumber $k = 2\pi/\lambda$. This scale crossed outside the Hubble radius, during inflation, at an instant $t_*(k)$ defined by

$$k = a(t_*)H(t_*). \tag{III-3.57}$$

To get a rough estimate of the number of e-foldings of inflation that are needed to solve the horizon problem, let us first ignore the transition from a radiation era to a matter era and assume for simplicity that the inflationary phase was followed instantaneously by a radiation phase that has lasted until now. During the radiation phase, the comoving Hubble radius $(aH)^{-1}$ increases like a. In order to solve the horizon problem, the increase of the comoving Hubble radius during the standard evolution must be compensated by *at least* a decrease of the same amount during inflation. Since the comoving Hubble radius roughly scales like a^{-1} during inflation, the minimum amount of inflation is simply given by the number of e-folds between the end of inflation and today $\ln(a_0/a_{end}) = \ln(T_{end}/T_0) \sim \ln(10^{29}(T_{end}/10^{16}\mathrm{GeV}))$, i.e. around 60 e-folds for a temperature $T \sim 10^{16}$Gev at the beginning of the radiation era. As we will see later, this energy scale is typical of inflation in the simplest models.

This determines roughly the number of e-folds $N(k_0)$ between the moment when the scale corresponding to our present Hubble radius $k_0 = a_0H_0$ exited the Hubble radius during inflation and the end of inflation. The other lengthscales of cosmological interest are *smaller* than k_0^{-1} and therefore exited the Hubble radius during inflation *after* the scale k_0, whereas they entered the Hubble radius during the standard cosmological phase (either in the radiation era for the smaller scales or in the matter era for the larger scales) *before* the scale k_0.

A more detailed calculation, which distinguishes between the energy scales at the end of inflation and after the reheating, gives for the number of e-folds between the exit of the mode k and the end of inflation

$$N(k) \simeq 62 - \ln\frac{k}{a_0H_0} + \ln\frac{V_k^{1/4}}{10^{16}\mathrm{GeV}} + \ln\frac{V_k^{1/4}}{V_{\mathrm{end}}^{1/4}} + \frac{1}{3}\ln\frac{\rho_{\mathrm{reh}}^{1/4}}{V_{\mathrm{end}}^{1/4}}. \tag{III-3.58}$$

Since the smallest scale of cosmological relevance is of the order of 1 Mpc, the range of cosmological scales covers about 9 e-folds.

The above number of e-folds is altered if one changes the thermal history of the universe between inflation and the present time by including for instance a period of so-called thermal inflation.

III-4 Power-law potentials

It is now time to illustrate all the points discussed above with some specific potential. We consider first the case of power-law monomial potentials, of the form

$$V(\phi) = \frac{\lambda}{p} m_P^4 \left(\frac{\phi}{m_P}\right)^p, \tag{III-4.59}$$

which have been abundantly used in the literature. In particular, the above potential includes the case of a free massive scalar field, $V(\phi) = m^2\phi/2$. The slow-roll conditions $\epsilon \ll 1$ and $\eta \ll 1$ both imply

$$\phi \gg p\, m_P, \tag{III-4.60}$$

which means that the scalar field amplitude must be above the Planck mass during inflation.

After subsituting the potential (III-4.59) into the slow-roll equations of motion (III-2.49-III-2.50), one can integrate them explicitly to get

$$\phi^{2-\frac{p}{2}} - \phi_i^{2-\frac{p}{2}} = -\frac{2}{4-p}\sqrt{\frac{p\lambda}{3}}\, m_P^{3-\frac{p}{2}}\,(t-t_i) \tag{III-4.61}$$

for $p \neq 4$ and

$$\phi = \phi_i \exp\left[-\sqrt{\frac{4\lambda}{3}}\, m_P (t-t_i)\right] \tag{III-4.62}$$

for $p = 4$.

One can also express the scale factor as a function of the scalar field (and thus as a function of time by substituting the above expression for $\phi(t)$) by using $d\ln a/d\phi = H/\dot{\phi} \simeq -\phi/(pm_P^2)$. One finds

$$a = a_{end} \exp\left[-\frac{\left(\phi^2 - \phi_{end}^2\right)}{2p\, m_P^2}\right]. \tag{III-4.63}$$

Defining the end of inflation by $\epsilon_V = 1$, one gets $\phi_{end} = p\, m_P/\sqrt{2}$ and the number of e-folds is thus given by

$$N(\phi) \simeq \frac{\phi^2}{2pm_P^2} - \frac{p}{4}. \tag{III-4.64}$$

This can be inverted, so that

$$\phi(N) \simeq \sqrt{2Np}\, m_P, \tag{III-4.65}$$

where we have ignored the second term of the right hand side of (III-4.64), in agreement with the condition (III-4.60).

III-5 Exponential potential

If one considers a potential of the form

$$V = V_0 \exp\left(-\sqrt{\frac{2}{q}}\,\frac{\phi}{m_P}\right), \tag{III-5.66}$$

then it is possible to find an *exact* solution (i.e. valid beyond the slow-roll approximation) of the system (III-1.46-III-1.48), with a power-law scale factor, i.e.

$$a(t) \propto t^q. \tag{III-5.67}$$

The evolution of the scalar field is given by the expression

$$\phi(t) = \sqrt{2q}\, m_P \ln\left[\sqrt{\frac{V_0}{q(3q-1)}}\,\frac{t}{m_P}\right]. \tag{III-5.68}$$

Note that one recovers the slow-roll approximation in the limit $q \gg 1$, since the slow-roll parameters are given by $\epsilon_V = 1/q$ and $\eta_V = 2/q$.

III-6 Brief history of the inflationary models

Let us now try to summarize in a few lines the history of inflationary models. The first model of inflation is usually traced back to Alan Guth [2] in 1981, although one can see a precursor in the model of Alexei Starobinsky [3]. Guth's model, which is named today *old inflation* is based on a first-order phase transition, from a false vacuum with non zero energy, which generates an exponential inflationary phase, into a true vacuum with zero energy density. The true vacuum phase appears in the shape of bubbles via quantum tunneling. The problem with this inflationary model is that, in order to get sufficient inflation to solve the problems of the standard model mentioned earlier, the nucleation rate must be sufficiently small; but, then, the bubbles never coalesce because the space that separates the bubbles undergoes inflation and expands too rapidly. Therefore, the first model of inflation is not phenomenologically viable.

After this first and unsuccessful attempt, a new generation of inflationary models appeared, usually denoted *new inflation* models [4]. They rely on a second order phase transition, based on thermal corrections of the effective potential and thus assume that the scalar field is in thermal equilibrium.

This hypothesis of thermal equilibrium was given up in the third generation of models, initiated by Andrei Linde, and whose generic name is *chaotic inflation* [5]. This allows to use extremely simple potentials,

quadratic or quartic, which lead to inflationary phases when the scalar field is displaced from the origin with values of the order of several Planck masses. This is sometimes considered to be problematic from a particle physics point of view, as discussed briefly later.

During the last few years, there has been a revival of the inflation model building based on high energy theories, in particular in the context of supersymmetry. In these models, the value of the scalar field is much smaller than the Planck mass.

III-7 The inflationary "zoology"

There exist many models of inflation. As far as single-field models are concerned (or at least *effectively* single field during inflation, the hybrid models requiring a second field to end inflation as discussed below), it is convenient to regroup them into three broad categories:

- Large field models ($0 < \eta \leq \epsilon$)

  The scalar field is displaced from its stable minimum by $\Delta\phi \sim m_P$. This includes the so-called chaotic type models with monomial potentials

  $$V(\phi) = \Lambda^4 \left(\frac{\phi}{\mu}\right)^p , \tag{III-7.69}$$

  or the exponential potential

  $$V(\phi) = \Lambda^4 \exp\left(\phi/\mu\right), \tag{III-7.70}$$

  which have already been discussed.

  This category of models is widely used in the literature because of their computational simplicity. It has been claimed, although it is still a debated issue, that these models are not so well motivated by particle physics, based on the argument that all non-renormalizable terms must be included in a generic potential and are out of control when the scalar field is of order of a few times the Planck mass [8]. For this reason, some inflationary model builders prefer to concentrate on models where the scalar field amplitude is small with respect to the Planck mass, as those discussed just below.

- Small field models ($\eta < 0 < \epsilon$)

  In this type of models, the scalar field is rolling away from an unstable maximum of the potential. This is a characteristic feature of spontaneous symmetry breaking. A typical potential is

  $$V(\phi) = \Lambda^4 \left[1 - \left(\frac{\phi}{\mu}\right)^p\right], \tag{III-7.71}$$

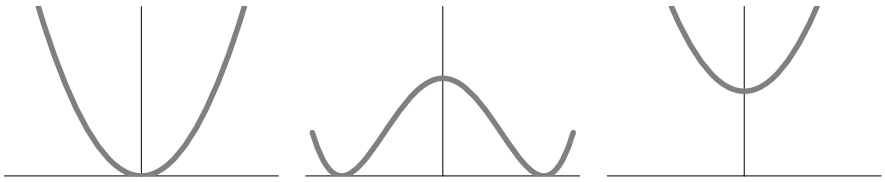

Figure 2. Schematic potential for t he three main categories of inflationary models: a. chaotic models b. symmetry breaking models; c. hybrid models

which can be interpreted as the lowest-order term in a Taylor expansion about the origin. Historically, this potential shape appeared in the so-called 'new inflation' scenario.

A particular feature of these models is that tensor modes are much more suppressed with respect to scalar modes than in the large-field models, as it will be shown later.

- Hybrid models ($0 < \epsilon < \eta$)

  This category of models, which appeared rather recently, relies on the presence of two scalar fields: one plays the traditional rôle of the inflaton, while the other is necessary to end inflation.

  As an illustration, let us consider the original model of hybrid inflation [6] based on the potential

$$V(\phi, \psi) = \frac{1}{2} m^2 \phi^2 + \frac{1}{2} \lambda' \psi^2 \phi^2 + \frac{1}{4} \lambda \left(M^2 - \psi^2 \right)^2 . \qquad \text{(III-7.72)}$$

  For values of the field ϕ larger than the critical value $\phi_c = \lambda M^2 / \lambda'$, the potential for ψ has its minimum at $\psi = 0$. This is the case during inflation. ψ is thus trapped in this minimum $\psi = 0$, so that the effective potential for the scalar field ϕ, which plays the rôle of the inflaton, is given by

$$V_{eff}(\phi) = V_0 + \frac{1}{2} m^2 \phi^2, \qquad \text{(III-7.73)}$$

  with $V_0 = \lambda M^4 / 4$. During the inflationary phase, the field ϕ slow-rolls until it reaches the critical value ϕ_c. The shape of the potential for ψ is then modified and new minima appear in $\psi = \pm M$. ψ will thus roll down into one of these new minima and, as a consequence, inflation will end.

  As far as inflation is concerned, hybrid inflation scenarios correspond effectively to single-field models with a potential character-

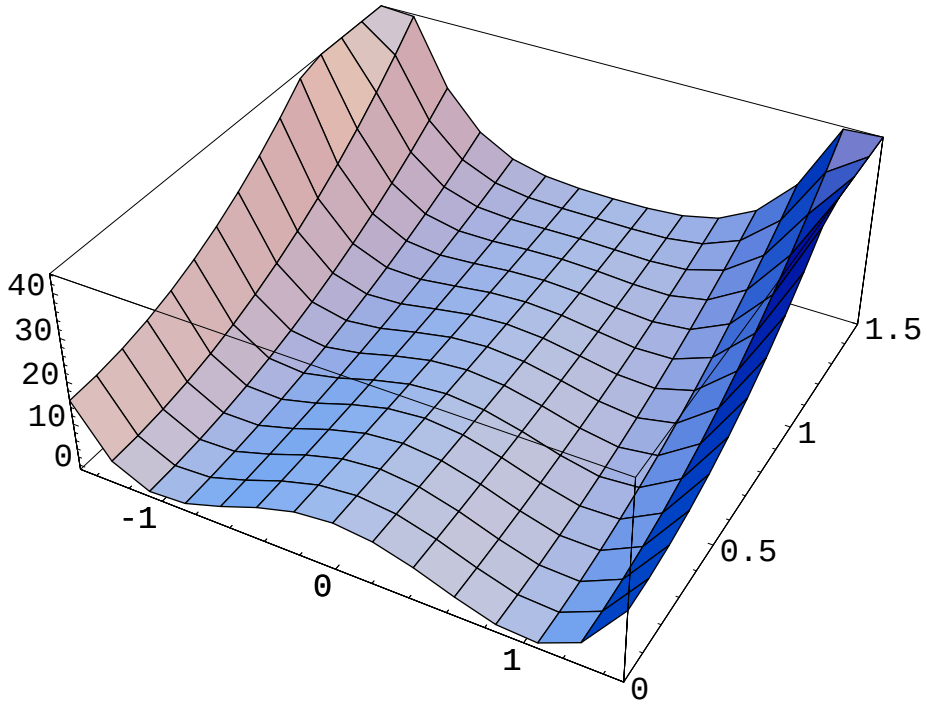

Figure 3. Typical potential $V(\phi, \psi)$ for hybrid inflation.

ized by $V''(\phi) > 0$ and $0 < \epsilon < \eta$. A typical potential is

$$V(\phi) = \Lambda^4 \left[1 + \left(\frac{\phi}{\mu}\right)^p\right]. \qquad \text{(III-7.74)}$$

Once more, this potential can be seen as the lowest order in a Taylor expansion about the origin.

In the case of hybrid models, the value ϕ_N of the scalar field as a function of the number of e-folds before the end of inflation is not determined by the above potential and, therefore, (ϕ_N/μ) can be considered as a freely adjustable parameter.

Many more details on inflationary models can be found in e.g. [7–9].

IV. The theory of cosmological perturbations

So far, we have concentrated our attention on strictly homogeneous and isotropic aspects of cosmology. Of course, this idealized version, although extremely useful, is not sufficient to account for real cosmology and it is now time to turn to the study of deviations from homogeneity and isotropy.

In cosmology, inhomogeneities grow because of the attractive nature of gravity, which implies that inhomogeneities were much smaller in the past. As a consequence, for most of their evolution, inhomogeneities can be treated as *linear perturbations*. The linear treatment ceases to be valid on small scales in our recent past, hence the difficulty to reconstruct the primordial inhomogeneities from large-scale structure, but it is quite adequate to describe the fluctuations of the CMB at the time of last scattering. This is the reason why the CMB is currently the best observational probe of primordial inhomogeneities. For more details on the relativistic theory of cosmological perturbations, which will be briefly introduced in this chapter, the reader is invited to read the standard reviews [10] in the literature.

From now on, we will be mostly working with the conformal time η, instead of the cosmic time t. The conformal time is defined as

$$\eta = \int \frac{dt}{a(t)} \qquad \text{(IV-0.75)}$$

so that the (spatially flat) FLRW metric takes the remarkably simple form

$$ds^2 = a^2(\eta)\left[-d\eta^2 + \delta_{ij}dx^i dx^j\right]. \qquad \text{(IV-0.76)}$$

IV-1 Perturbations of the geometry

Let us start with the linear perturbations of the geometry. The most general linear perturbation of the FLRW metric can be expressed as

$$ds^2 = a^2 \left\{ -(1+2A)d\eta^2 + 2B_i dx^i d\eta + (\delta_{ij} + h_{ij})\, dx^i dx^j \right\}, \quad \text{(IV-1.77)}$$

where we have considered only the spatially flat FLRW metric.

We have introduced a time plus space decomposition of the perturbations. The indices i, j stand for *spatial* indices and the perturbed quantities defined in (IV-1.77) can be seen as three-dimensional tensors, for which the indices can be lowered (or raised) by the spatial metric δ_{ij} (or its inverse).

It is very convenient to separate the perturbations into three categories, the so-called "scalar", "vector" and "tensor" modes. For example, a spatial vector field B^i can be decomposed uniquely into a longitudinal part and a transverse part,

$$B_i = \nabla_i B + \bar{B}_i, \quad \nabla_i \bar{B}^i = 0, \quad \text{(IV-1.78)}$$

where the longitudinal part is curl-free and can thus be expressed as a gradient, and the transverse part is divergenceless. One thus gets one "scalar" mode, B , and two "vector" modes $\bar{B}^i$ (the index i takes three values but the divergenceless condition implies that only two components are independent.

A similar procedure applies to the symmetric tensor h_{ij}, which can be decomposed as

$$h_{ij} = 2C\delta_{ij} + 2\nabla_i\nabla_j E + 2\nabla_{(i}E_{j)} + \overline{E}_{ij}, \quad \text{(IV-1.79)}$$

with $\overline{E}^{ij}$ transverse and traceless (TT), i.e. $\nabla_i \overline{E}^{ij} = 0$ (transverse) and $\overline{E}^{ij}\delta_{ij} = 0$ (traceless), and E_i transverse. The parentheses around the indices denote symmetrization, namely $2\nabla_{(i}E_{j)} = \nabla_i E_j + \nabla_j E_i$. We have thus defined two scalar modes, C and E, two vector modes, E_i, and two tensor modes, $\bar{E}_{ij}$.

In the following, we will be mainly interested in the metric with only *scalar* perturbations, since scalar modes are the most relevant for cosmology and they can be treated independently of the vector and tensor modes. In matrix notation, the perturbed metric will thus be of the form

$$g_{\mu\nu} = a^2 \begin{bmatrix} -(1+2A) & \nabla_i B \\ \nabla_j B & \{(1+2C)\delta_{ij} + 2\nabla_i\nabla_j E\} \end{bmatrix}. \quad \text{(IV-1.80)}$$

After the description of the perturbed geometry, we turn to the perturbations of the matter in the next subsection.

IV-2 Perturbations of the matter

Quite generally, the perturbed energy-momentum tensor can be written in the form

$$T^{\mu}_{\nu} = \begin{bmatrix} -(\rho+\delta\rho) & q_j \\ -q^i + (\rho+P)B^i & (P+\delta P)\delta^i_j + \pi^i_j \end{bmatrix}, \qquad \text{(IV-2.81)}$$

where q_i is the momentum density and π^i_j is the anisotropic stress tensor, which is traceless, i.e. $\pi^k_k = 0$. One can then decompose these tensors into scalar, vector and tensor components, as before for the metric components, so that

$$q_i = \nabla_i q + \bar{q}_i, \quad \nabla_i \bar{q}^i = 0, \qquad \text{(IV-2.82)}$$

and

$$\pi_{ij} = \nabla_i\nabla_j\pi - \frac{1}{3}\delta_{ij}\nabla_k\nabla^k\pi + 2\nabla_{(i}\pi_{j)} + \bar{\pi}_{ij}, \qquad \text{(IV-2.83)}$$

$$\nabla_k\pi^k = 0, \quad \nabla_k\bar{\pi}^{kl} = 0, \quad \bar{\pi}^k_k = 0. \qquad \text{(IV-2.84)}$$

Fluid. A widely used description for matter in cosmology is that of a fluid. Its homogeneous part is described by the energy-momentum tensor of a perfect fluid, as seen earlier, while its perturbed part can be expressed as

$$\delta T^0_0 = -\delta\rho, \qquad \text{(IV-2.85)}$$

$$\delta T^i_0 = -(\rho+p)\,v^i, \quad \delta T^0_i = (\rho+p)\,(v_i + B_i) \qquad \text{(IV-2.86)}$$

$$\delta T^i_j = \delta P\delta^i_j + \pi^i_j, \qquad \text{(IV-2.87)}$$

with $\pi^k_k = 0$ as before and where v^i is the three-dimensional fluid velocity defined by

$$\delta u^i = \frac{1}{a}v^i. \qquad \text{(IV-2.88)}$$

It is also possible to separate this perturbed energy-momentum tensor into scalar, vector and tensor parts by using decompositions for v_i and π_{ij} similar to (IV-2.82) and (IV-2.84).

Scalar field. Another type of matter, especially useful in the context of inflation, is a scalar field. The homogeneous description has already been given earlier and the perturbed expression for the energy-momentum tensor follows immediately from (III-1.41), taking into account the metric perturbations as well. One finds

$$a^2\delta T^0_0 = -a^2\delta\rho = -\phi'\delta\phi' - a^2V'\delta\phi + \phi'^2 A, \qquad \text{(IV-2.89)}$$

$$a^2\delta T_i^0 = a^2 q_i = -\phi'\partial_i\delta\phi, \qquad \text{(IV-2.90)}$$

$$a^2\delta T_j^i = -\delta_j^i\left(a^2V'\delta\phi + \phi'^2 A - \phi'\delta\phi'\right). \qquad \text{(IV-2.91)}$$

The last equation shows that, for a scalar field, there is no anisotropic stress in the energy-momentum tensor.

IV-3 Gauge transformations

It is worth noticing that there is a fundamental difference between the perturbations in general relativity and the perturbations in other field theories where the underlying spacetime is fixed. In the latter case, one can define the perturbation of a given field ϕ as

$$\delta\phi(p) = \phi(p) - \bar{\phi}(p), \qquad \text{(IV-3.92)}$$

where $\bar{\phi}$ is the unperturbed field and p is any point of the spacetime.

In the context of general relativity, spacetime is no longer a frozen background but must also be perturbed if matter is perturbed. As a consequence, the above definition does not make sense since the perturbed quantity ϕ lives in the perturbed spacetime $\mathcal{M}$, whereas the unperturbed quantity $\bar{\phi}$ lives in *another spacetime*: the unperturbed spacetime of reference, which we will denote $\bar{\mathcal{M}}$. In order to use a definition similar to (IV-3.92), one must introduce a one-to-one *identification*, ι, between the points of $\bar{\mathcal{M}}$ and those of $\mathcal{M}$. The perturbation of the field can then be defined as

$$\delta\phi(\bar{p}) = \phi(\iota(\bar{p})) - \bar{\phi}(\bar{p}), \qquad \text{(IV-3.93)}$$

where $\bar{p}$ is a point of $\bar{\mathcal{M}}$.

However, the identification ι is not uniquely defined, and therefore the definition of the perturbation depends on the particular choice for ι: two different identifications, ι_1 and ι_2 say, lead to two different definitions for the perturbations. This ambiguity can be related to the freedom of choice of the coordinate system. Indeed, if one is given a coordinate system in $\bar{\mathcal{M}}$, one can transport it into $\mathcal{M}$ via the identification. ι_1 and ι_2 thus define two different coordinate systems in $\mathcal{M}$, and in this respect, a *change of identification* can be seen as a *change of coordinates* in $\mathcal{M}$.

The metric perturbations, introduced in (IV-1.77), are modified in a coordinate transformation of the form

$$x^\alpha \to x^\alpha + \xi^\alpha, \qquad \xi^\alpha = (\xi^0, \xi^i). \qquad \text{(IV-3.94)}$$

It can be shown that the change of the metric components can be expressed as

$$\Delta\left(\delta g_{\mu\nu}\right) = -2D_{(\mu}\xi_{\nu)}. \qquad \text{(IV-3.95)}$$

where Δ represents the variation, due the coordinate transformation, at the same old and new coordinates (and thus at different physical points). The above variation can be decomposed into individual variations for the various components of the metric defined earlier. One finds

$$\begin{aligned} \Delta A &= -\xi^{0\prime} - \mathcal{H}\xi^0 && \text{(IV-3.96)} \\ \Delta B_i &= \nabla_i \xi^0 - \xi_i' && \text{(IV-3.97)} \\ \Delta h_{ij} &= -2\left(\nabla_{(i}\xi_{j)} - \mathcal{H}\xi^0\delta_{ij}\right). && \text{(IV-3.98)} \end{aligned}$$

The effect of a coordinate transformation can also be decomposed along the scalar, vector and tensor sectors introduced earlier. The generator ξ^α of the coordinate transformation can be written as

$$\xi^\alpha = (\xi^0, \nabla^i\xi + \overline{\xi}^i), \qquad \text{(IV-3.99)}$$

with $\overline{\xi}^i$ transverse. This shows explicitly that ξ^α contains two scalar components, ξ^0 and ξ, and two vector components, $\overline{\xi}^i$. The transformations (IV-3.96-IV-3.98) are then decomposed into :

$$\begin{aligned} A &\rightarrow A - \xi^{0\prime} - \mathcal{H}\xi^0 \\ B &\rightarrow B + \xi^0 - \xi' \\ C &\rightarrow C - \mathcal{H}\xi^0 \\ E &\rightarrow E - \xi \\ \overline{B}^i &\rightarrow \overline{B}^i - \overline{\xi}^{i\prime} \\ E^i &\rightarrow E^i - \overline{\xi}^i. \end{aligned} \qquad \text{(IV-3.100)}$$

The tensor perturbations remain unchanged since ξ^α does not contain any tensor component. The matter perturbations, either in the fluid description or in the scalar field description, follow similar transformation laws in a coordinate change.

In order to study the physically relevant modes and not spurious modes due to coordinate ambiguities, two strategies can a priori be envisaged. The first consists in working from the start in a specific gauge. A familiar choice in the literature on cosmological perturbations is the *longitudinal gauge* (also called conformal Newton gauge), which imposes

$$B_{\rm L} = 0, \quad E_{\rm L} = 0. \qquad \text{(IV-3.101)}$$

The second approach consists in defining *gauge-invariant variables*, i.e. variables that are left unchanged under a coordinate transformation. For the scalar metric perturbations, we start with four quantities (A, B, C and E) and we can use two gauge transformations (ξ^0 and ξ). This

implies that the scalar metric perturbations must be described by *two* independent gauge-invariant quantities. Two such quantities are

$$\Phi = A + (B - E')' + \mathcal{H}(B - E') \tag{IV-3.102}$$

and

$$\Psi = -C - \mathcal{H}(B - E'), \tag{IV-3.103}$$

as it can be checked by considering the explicit transformations in (IV-3.100). It turns out that, in the longitudinal gauge, the remaining scalar perturbations $A_{\rm L}$ and $C_{\rm L}$ are numerically equivalent to the gauge-invariant quantities just defined Φ and $-\Psi$.

In practice, one can combine the two strategies by doing explicit calculations in a given gauge and then by relating the quantities defined in this gauge to some gauge-invariant variables. It is then possible to translate the results in any other gauge. In the rest of these lectures, we will use the longitudinal gauge.

IV-4 The perturbed Einstein equations

After having defined the metric and the matter perturbations, we can now relate them via the perturbed Einstein equations. We will consider here explicitly *only the scalar sector*, which is the most complicated but also the most interesting for cosmological applications.

Starting from the perturbed metric (IV-1.80), one can compute the components of Einstein's tensor at linear order. In the *longitudinal gauge*, i.e. with $B_{\rm L} = E_{\rm L} = 0$, one finds

$$\left(\delta G^0_0\right)_{\rm L} = \frac{2}{a^2}\left[3\mathcal{H}^2 A_{\rm L} - 3\mathcal{H}C'_{\rm L} + \nabla^2 C_{\rm L}\right] \tag{IV-4.104}$$

$$\left(\delta G^0_i\right)_{\rm L} = \frac{2}{a^2}\nabla_i\left[-\mathcal{H}A_{\rm L} + C'_{\rm L}\right] \tag{IV-4.105}$$

$$\begin{aligned}\left(\delta G^i_j\right)_{\rm L} = {} & \frac{1}{a^2}\nabla^i\nabla_j\left(-C_{\rm L} - A_{\rm L}\right) + \frac{1}{a^2}\left[-2C''_{\rm L} - 4\mathcal{H}C'_{\rm L} + \nabla^2 C_{\rm L}\right. \\ & \left. +2\mathcal{H}A'_{\rm L} + \nabla^2 A_{\rm L} + 2\left(2\mathcal{H}' + \mathcal{H}^2\right)A_{\rm L}\right]\delta^i_j.\end{aligned} \tag{IV-4.106}$$

Combining with the perturbations of the energy-momentum tensor given in (IV-2.81), the perturbed Einstein equations yield, *in the longitudinal gauge*, the following relations: the energy constraint (from (IV-4.104))

$$3\mathcal{H}^2 A_{\rm L} - 3\mathcal{H}C'_{\rm L} + \nabla^2 C_{\rm L} = -4\pi G a^2 \delta\rho_{\rm L}, \tag{IV-4.107}$$

the momentum constraint (from (IV-4.105))

$$C'_{\rm L} - \mathcal{H}A_{\rm L} = 4\pi G\, a^2\, q_{\rm L}, \tag{IV-4.108}$$

the "anisotropy constraint" (from the traceless part of (IV-4.106))

$$-A_{\rm L} - C_{\rm L} = 8\pi G\, a^2\, \pi_{\rm L}, \qquad \text{(IV-4.109)}$$

and finally

$$C_{\rm L}'' + 2\mathcal{H}C_{\rm L}' - \mathcal{H}A_{\rm L}' - (2\mathcal{H}' + \mathcal{H}^2)A_{\rm L} - \frac{1}{3}\nabla^2(A_{\rm L} + C_{\rm L}) = -4\pi G\, \delta P_{\rm L}, \qquad \text{(IV-4.110)}$$

obtained from the trace of (IV-4.106).

The combination of the energy and momentum constraints gives the useful relation

$$\nabla^2\Psi = 4\pi G a^2\, (\delta\rho_{\rm L} - 3\mathcal{H}q_{\rm L}) \equiv 4\pi G a^2 \delta\rho_c, \qquad \text{(IV-4.111)}$$

where we have introduced the *comoving* energy density perturbation $\delta\rho_c$: this gauge-invariant quantity corresponds, according to its definition, to the energy density perturbation measured in comoving gauges characterized by $\delta T^0_i = q_i = 0$. We have also replaced $C_{\rm L}$ by $-\Psi$. Note that the above equation is quite similar to the Newtonian Poisson equation, but with quantities whose natural interpretation is given in *different* gauges.

IV-5 Equations for the matter

As mentioned earlier, a consequence of Einstein's equations is that the *total* energy-momentum tensor is covariantly conserved (see Eq. (II-0.9)). For a fluid, the conservation of the energy-momentum tensor leads to a continuity equation that generalizes the *continuity equation* of fluid mechanics, and a momentum conservation equation that generalizes the *Euler equation*. In the case of a single fluid, combinations of the perturbed Einstein equations obtained in the previous subsection lead necessarily to the perturbed continuity and Euler equations for the fluid. In the case of several non-interacting fluids, however, one must impose *separately* the covariant conservation of *each* energy-momentum tensor: this is not a consequence of Einstein's equations, which impose only the conservation of the *total* energy-momentum tensor.

In the perspective to deal with several cosmological fluids, it is therefore useful to write the perturbation equations, satisfied by a given fluid, that follow only from the conservation of the corresponding energy-momentum tensor, and independently of Einstein's equations.

The continuity equation can be obtained by perturbing $u^\mu D_\nu T^\nu_\mu = 0$. One finds, in *any gauge*,

$$\delta\rho' + 3\mathcal{H}\,(\delta\rho + \delta P) + (\rho + P)\left(3C' + \nabla^2 E' + \nabla^2 v\right) = 0. \qquad \text{(IV-5.112)}$$

Dividing by ρ, this can be reexpressed in terms of the density contrast $\delta = \delta\rho/\rho$:

$$\delta' + 3\mathcal{H}\left(\frac{\delta P}{\delta\rho} - w\right)\delta + (1+w)\left(\nabla^2 v + 3C' + \nabla^2 E'\right) = 0, \quad \text{(IV-5.113)}$$

where $w = p/\rho$ (w is not necessarily constant here). The perturbed Euler equation is derived from the spatial projection of $\delta(D_\nu T^\nu_\mu) = 0$. This gives

$$(v+B)' + \left(1 - 3c_s^2\right)\mathcal{H}\,(v+B) + A + \frac{\delta P}{\rho + P} + \frac{2}{3(\rho+P)}\nabla^2\pi = 0, \quad \text{(IV-5.114)}$$

where c_s is the sound speed, which is related to the time derivatives of the background energy density and pressure:

$$c_s^2 = \frac{p'}{\rho'}. \quad \text{(IV-5.115)}$$

There are as many systems of equations (IV-5.113-IV-5.114) as the number of fluids. If the fluids are interacting, one must add an interaction term on the right-hand side of the Euler equations.

Finally, let us stress that the fluid description is not always an adequate approximation for cosmological matter. A typical example is the photons during and after recombination: their mean free path becomes so large that they must be treated as a gas, which requires the use of the Boltzmann equation (see e.g. [11] for a presentation of the Boltzmann equation in the cosmological context).

IV-6 Initial conditions for standard cosmology

The notion of *initial conditions* depends in general on the context, since the initial conditions for a given period in the history of the universe can be seen as the final conditions of the previous phase. In cosmology, "initial conditions" usually refer to the state of the perturbations during the *radiation dominated era* (of standard cosmology) and on *wavelengths larger than the Hubble radius*.

Let us first address in details the question of initial conditions in the simple case of a single perfect fluid, radiation, with equation of state $p = \rho/3$ (which gives $c_s^2 = w = 1/3$). The four key equations are the continuity, Euler, Poisson and anisotropy equations, respectively Eqs (IV-5.113), (IV-5.114), (IV-4.111) and (IV-4.109). In terms of the Fourier components,

$$Q(\vec{k}) = \int \frac{d^3x}{(2\pi)^{3/2}} e^{-i\vec{k}.\vec{x}} Q(\vec{x}), \quad \text{(IV-6.116)}$$

and of the dimensionless quantity

$$x \equiv k\eta \tag{IV-6.117}$$

(during the radiation dominated era $\mathcal{H} = 1/\eta$), the four equations can be rewritten as

$$\frac{d\delta}{dx} - \frac{4}{3}\mathcal{V} + 4\frac{dC}{dx} = 0, \tag{IV-6.118}$$

$$\frac{d\mathcal{V}}{dx} + \frac{1}{4}\delta + A = 0, \tag{IV-6.119}$$

$$x^2 C = \frac{3}{2}\left(\delta - \frac{4}{x}\mathcal{V}\right) \tag{IV-6.120}$$

$$C = -A. \tag{IV-6.121}$$

We have introduced the quantity $\mathcal{V} \equiv kv$, which has the dimension of a velocity. Since we are interested in perturbations with wavelength larger than the Hubble radius, i.e. such that $x = k/\mathcal{H} \ll 1$, it is useful to consider a Taylor expansion for the various perturbations, for instance

$$\mathcal{V} = \mathcal{V}^{(0)} + x\mathcal{V}^{(1)} + \frac{x^2}{2}\mathcal{V}^{(2)} + \dots \tag{IV-6.122}$$

One then substitutes these Taylor expansions into the above system of equations. In particular, the Poisson equation (IV-6.120) gives

$$\mathcal{V}^{(0)} = 0, \tag{IV-6.123}$$

in order to avoid a divergence, as well as

$$\mathcal{V}^{(1)} = \frac{1}{4}\delta^{(0)}. \tag{IV-6.124}$$

The Euler equation (IV-6.119) then gives

$$A^{(0)} = -\frac{1}{2}\delta^{(0)}. \tag{IV-6.125}$$

The conclusion is that the initial conditions for each Fourier mode are determined by a single quantity, e.g. $\delta^{(0)}$, the other quantities being related via the above constraints.

In general, one must consider several cosmological fluids. Typically, the "initial" or "primordial" perturbations are defined deep in the radiation era but at temperatures low enough, i.e. after nucleosynthesis, so that the main cosmological components reduce to the usual photons, baryons, neutrinos and cold dark matter (CDM). The system (IV-6.118-IV-6.121) must thus be generalized to include a continuity equation and

a Euler equation for each fluid. The above various cosmological species can be characterized by their number density, n_X, and their energy density ρ_X. In a multi-fluid system, it is useful to distinguish *adiabatic* and *isocurvature* perturbations.

The *adiabatic mode* is defined as a perturbation affecting all the cosmological species such that the relative ratios in the number densities remain unperturbed, i.e. such that

$$\delta\left(n_X/n_Y\right) = 0. \tag{IV-6.126}$$

It is associated with a curvature perturbation, via Einstein's equations, since there is a global perturbation of the matter content. This is why the adiabatic perturbation is also called *curvature* perturbation. In terms of the energy density contrasts, the adiabatic perturbation is characterized by the relations

$$\frac{1}{4}\delta_\gamma = \frac{1}{4}\delta_\nu = \frac{1}{3}\delta_b = \frac{1}{3}\delta_c, \tag{IV-6.127}$$

They follow directly from the prescription (IV-6.126), each coefficient depending on the equation of state of the particuler species.

Since there are several cosmological species, it is also possible to perturb the matter components without perturbing the geometry. This corresponds to *isocurvature* perturbations, characterized by variations in the particle number ratios but with vanishing curvature perturbation. The variation in the relative particle number densities between two species can be quantified by the so-called *entropy perturbation*

$$S_{A,B} \equiv \frac{\delta n_A}{n_A} - \frac{\delta n_B}{n_B}. \tag{IV-6.128}$$

When the equation of state for a given species is such that $w \equiv p/\rho =$ Const, then one can reexpress the entropy perturbation in terms of the density contrast, in the form

$$S_{A,B} \equiv \frac{\delta_A}{1+w_A} - \frac{\delta_B}{1+w_B}. \tag{IV-6.129}$$

It is convenient to choose a species of reference, for instance the photons, and to define the entropy perturbations of the other species relative to it:

$$S_b \equiv \delta_b - \tfrac{3}{4}\delta_\gamma, \tag{IV-6.130}$$

$$S_c \equiv \delta_c - \tfrac{3}{4}\delta_\gamma, \tag{IV-6.131}$$

$$S_\nu \equiv \tfrac{3}{4}\delta_\nu - \tfrac{3}{4}\delta_\gamma, \tag{IV-6.132}$$

thus define respectively the *baryon isocurvature mode*, the *CDM isocurvature mode*, the *neutrino isocurvature mode*. In terms of the entropy perturbations, the adiabatic mode is obviously characterized by $S_b = S_c = S_\nu = 0$.

In summary, we can decompose a general perturbation, described by four density contrasts, into one adiabatic mode and three isocurvature mode. In fact, the problem is slightly more complicated because the evolution of the initial velocity fields. For a single fluid, we have seen that the velocity field is not an independent initial condition but depends on the density contrast so that there is no divergence backwards in time. In the case of the four species mentioned above, there remains however one arbitrary relative velocity between the species, which gives an additional mode, usually named the *neutrino isocurvature velocity* perturbation.

The CMB is a powerful way to study isocurvature perturbations because (primordial) adiabatic and isocurvature perturbations produce very distinctive features in the CMB anisotropies [12]. Whereas an adiabatic initial perturbation generates a cosine oscillatory mode in the photon-baryon fluid, leading to an acoustic peak at $\ell \simeq 220$ (for a flat universe), a pure isocurvature initial perturbation generates a sine oscillatory mode resulting in a first peak at $\ell \simeq 330$. The unambiguous observation of the first peak at $\ell \simeq 220$ has eliminated the possibility of a dominant isocurvature perturbation. The recent observation by WMAP of the CMB polarization has also confirmed that the initial perturbation is mainly an adiabatic mode. But this does not exclude the presence of a subdominant isocurvature contribution, which could be detected in future high-precision experiments such as Planck.

IV-7 Super-Hubble evolution

In the case of *adiabatic perturbations*, there is only one (scalar) dynamical degree of freedom. One can thus choose either an energy density perturbation or a metric perturbation, to study the dynamics, the other quantities being determined by the constraints.

If one considers the metric perturbation $\Psi = \Phi$ (assuming $\pi = 0$), one can combine Einstein's equations (IV-4.110), with $\delta p = c_s^2 \delta\rho$ (for adiabatic perturbations), and (IV-4.107) to obtain a second-order differential equation in terms of Ψ only:

$$\Psi'' + 3\mathcal{H}(1 + c_s^2)\Psi' + \left[2\mathcal{H}' + (1 + 3c_s^2)\mathcal{H}\right]\Psi - k^2 c_s^2 \Psi = 0. \quad \text{(IV-7.133)}$$

Using the background Friedmann equations, the sound speed can be reexpressed in terms of the scale factor and its derivatives. For scales larger than the sonic horizon, i.e. such that $kc_s \ll \mathcal{H}$, the above equation

can be integrated explicitly and yields

$$\Psi = \frac{\mathcal{H}}{a^2}\left[\alpha \int d\eta \frac{a^2\left(\mathcal{H}' - \mathcal{H}^2\right)}{\mathcal{H}^2} + \beta\right], \tag{IV-7.134}$$

where α and β are two integration constants.

For a scale factor evolving like $a \propto t^p$, one gets

$$\Psi = -\frac{\alpha}{p+1} + \beta\, p\, t^{-p-1}. \tag{IV-7.135}$$

They are two modes: a constant mode and a decaying mode. Note that, in the previous subsection on the initial conditions, we eliminated the decaying mode to avoid the divergence when going backwards in time.

In a transition between two cosmological phases characterized respectively by the scale factors $a \propto t^{p_1}$ and $a \propto t^{p_2}$, one can easily finds the relation between the asymptotic behaviours of Ψ (i.e. after the decaying mode becomes negligible) by using the constancy of α. This gives

$$\Psi_2 = \frac{p_2+1}{p_1+1}\Psi_1. \tag{IV-7.136}$$

This is valid only asymptotically. In the case of a sharp transition, Ψ must be continuous at the transition and the above relation will apply only after some relaxation time. For a transition radiation/non-relativistic matter, one finds

$$\Psi_{mat} = \frac{9}{10}\Psi_{rad}. \tag{IV-7.137}$$

In practice and for more general cases, it turns out that it is much more convenient to follow the evolution of cosmological perturbations by resorting to quantities that are *conserved on super-Hubble scales*. A familiar example of such a quantity is the *curvature perturbation on uniform-energy-density hypersurfaces*, which can be expressed in any gauge as

$$\zeta = C - \mathcal{H}\frac{\delta\rho}{\rho'}. \tag{IV-7.138}$$

This is a gauge-invariant quantity by definition. The conservation equation (IV-5.113) can then be rewritten as

$$\zeta' = -\frac{\mathcal{H}}{\rho + P}\delta P_{nad} - \frac{1}{3}\nabla^2(E' + v), \tag{IV-7.139}$$

where δP_{nad} is the non-adiabatic part of the pressure perturbation, defined by

$$\delta P_{nad} = \delta P - c_s^2 \delta\rho. \tag{IV-7.140}$$

The expression (IV-7.139) shows that ζ is conserved *on super-Hubble scales* in the case of *adiabatic perturbations*.

Another convenient quantity, which is sometimes used in the literature instead of ζ, is the *curvature perturbation on comoving hypersurfaces*, which can be written in any gauge as

$$-\mathcal{R} = C + \frac{\mathcal{H}}{\rho + P} q. \tag{IV-7.141}$$

It is easy to relate the two quantities ζ and $\mathcal{R}$. Substituting e.g. $\delta\rho = \delta\rho_c + 3\mathcal{H}q$, which follows from the definition (see (IV-4.111)) of the comoving energy density perturbation, into (IV-7.138), one finds

$$\zeta = -\mathcal{R} + \frac{\delta\rho_c}{\rho + P}. \tag{IV-7.142}$$

Using Einstein's equations, in particular (IV-4.111), this can be rewritten as

$$\zeta = -\mathcal{R} - \frac{2\rho}{3(\rho + P)} \left(\frac{k}{aH}\right)^2 \Psi. \tag{IV-7.143}$$

The latter expression shows that ζ and $\mathcal{R}$ coincide in the super-Hubble limit $k \ll aH$.

The quantity $\mathcal{R}$ can also be expressed in terms of the two Bardeen potentials Φ and Ψ. Using the momentum constraint (IV-4.108) and the Friedmann equations, one finds

$$\mathcal{R} = \Psi - \frac{H}{\dot{H}} \left(\dot{\Psi} + H\Phi\right). \tag{IV-7.144}$$

In a cosmological phase dominated by a fluid with no anisotropic stress, so that $\Phi = \Psi$, and with an equation of state $P = w\rho$ with w constant, we have already seen that Ψ is constant with time. Since the scale factor evolves like $a \propto t^p$ with $p = 2/3(1 + w)$, the relation (IV-7.144) between $\mathcal{R}$ and Ψ reduces to

$$\mathcal{R} = \frac{5 + 3w}{3(1 + w)} \Psi. \tag{IV-7.145}$$

In the radiation era, $\mathcal{R} = (3/2)\Psi$, whereas in the matter era, $\mathcal{R} = (5/3)\Psi$, and since $\mathcal{R}$ is conserved, one recovers the conclusion given in Eq. (IV-7.137).

During inflation, $w \simeq -1$ and

$$w + 1 = \frac{\dot{\phi}^2}{\rho} \simeq -\frac{2}{3} \frac{\dot{H}}{H^2} \tag{IV-7.146}$$

so that

$$\mathcal{R} \simeq -\frac{H^2}{\dot{H}}\Psi_{inf}. \qquad \text{(IV-7.147)}$$

For a scalar field, the perturbed equation of motion reads

$$\ddot{\delta\sigma} + 3H\dot{\delta\sigma} + \left(\frac{k^2}{a^2} + V''\right)\delta\sigma = \dot{\sigma}\left(\dot{\Phi} + 3\dot{\Psi}\right) - 2V'\Phi. \qquad \text{(IV-7.148)}$$

V. Quantum fluctuations and "birth" of cosmological perturbations

In the previous section, we have discussed the *classical* evolution of the cosmological perturbations. In the *classical* context, the initial conditions, defined deep in the radiation era, are a priori arbitrary. What is remarkable with inflation is that the accelerated expansion can convert *initial vacuum quantum fluctuations* into "macroscopic" cosmological perturbations (see [13] for the seminal works). In this sense, inflation provides us with "natural" initial conditions, which turn out to be the initial conditions that agree with the present observations.

V-1 Massless scalar field in de Sitter

As a warming-up, it is instructive to discuss the case of a massless scalar field in a so-called de Sitter universe, or a FLRW spacetime with exponential expansion, $a \propto \exp(Ht)$. In conformal time, the scale factor is given by

$$a(\eta) = -\frac{1}{H\eta}. \qquad \text{(V-1.149)}$$

The conformal time is here negative (so that the scale factor is positive) and goes from $-\infty$ to 0. The action for a massless scalar field in this geometry is given by

$$S = \int d^4x\sqrt{-g}\left(-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi\right) = \int d\eta\ d^3x\ a^4\left[\frac{1}{2a^2}\phi'^2 - \frac{1}{2a^2}\vec{\nabla}\phi^2\right], \qquad \text{(V-1.150)}$$

where we have substituted in the action the cosmological metric (IV-0.76). Note that, whereas we still allow for spatial variations of the scalar field, i.e. inhomogeneities, we will assume here, somewhat inconsistently, that the geometry is completely fixed as homogeneous. We will deal later with the question of the metric perturbations.

It is possible to write the above action with a canonical kinetic term via the change of variable

$$u = a\phi. \qquad \text{(V-1.151)}$$

After an integration by parts, the action (V-1.150) can be rewritten as

$$S = \frac{1}{2}\int d\eta\; d^3x\; \left[u'^2 - \vec{\nabla}u^{\,2} + \frac{a''}{a}u^2\right]. \qquad \text{(V-1.152)}$$

The first two terms are familiar since they are the same as in the action for a free massless scalar field in Minkowski spacetime. The fact that our scalar field here lives in de Sitter spacetime rather than Minkowski has been reexpressed as a *time-dependent effective mass*

$$m_{eff}^2 = -\frac{a''}{a} = -\frac{2}{\eta^2}. \qquad \text{(V-1.153)}$$

Our next step will be to quantize the scalar field u by using the standard procedure of quantum field theory. One first turns u into a quantum field denoted $\hat{u}$, which we expand in Fourier space as

$$\hat{u}(\eta,\vec{x}) = \frac{1}{(2\pi)^{3/2}}\int d^3k \left\{\hat{a}_{\vec{k}}u_k(\eta)e^{i\vec{k}.\vec{x}} + \hat{a}^\dagger_{\vec{k}}u_k^*(\eta)e^{-i\vec{k}.\vec{x}}\right\}, \qquad \text{(V-1.154)}$$

where the $\hat{a}^\dagger$ and $\hat{a}$ are creation and annihilation operators, which satisfy the usual commutation rules

$$\left[\hat{a}_{\vec{k}},\hat{a}_{\vec{k}'}\right] = \left[\hat{a}^\dagger_{\vec{k}},\hat{a}^\dagger_{\vec{k}'}\right] = 0, \quad \left[\hat{a}_{\vec{k}},\hat{a}^\dagger_{\vec{k}'}\right] = \delta(\vec{k}-\vec{k}'). \qquad \text{(V-1.155)}$$

The function $u_k(\eta)$ is a complex time-dependent function that must satisfy the *classical* equation of motion in Fourier space, namely

$$u_k'' + \left(k^2 - \frac{a''}{a}\right)u_k = 0, \qquad \text{(V-1.156)}$$

which is simply the equation of motion for an oscillator with a time-dependent mass. In the case of a massless scalar field in Minkowski spacetime, this effective mass is zero ($a''/a = 0$) and one usually takes $u_k = (\hbar/2k)^{1/2}e^{-ik\eta}$ (the choice for the normalization factor will be clear below). In the case of de Sitter, one can solve explicitly the above equation with $a''/a = 2/\eta^2$ and the general solution is given by

$$u_k = \alpha e^{-ik\eta}\left(1 - \frac{i}{k\eta}\right) + \beta e^{ik\eta}\left(1 + \frac{i}{k\eta}\right). \qquad \text{(V-1.157)}$$

Canonical quantization consists in imposing the following commutation rules on the η =constant hypersurfaces:

$$\left[\hat{u}(\eta,\vec{x}),\hat{u}(\eta,\vec{x}')\right] = \left[\hat{\pi}_u(\eta,\vec{x}),\hat{\pi}_u(\eta,\vec{x}')\right] = 0 \qquad \text{(V-1.158)}$$

and

$$\left[\hat{u}(\eta,\vec{x}),\hat{\pi}_u(\eta,\vec{x}')\right] = i\hbar\delta(\vec{x}-\vec{x}'), \tag{V-1.159}$$

where $\pi_u \equiv \delta S/\delta u'$ is the conjugate momentum of u. In the present case, $\pi_u = u'$ since the kinetic term is canonical.

Subtituting the expansion (V-1.154) in the commutator (V-1.159), and using the commutation rules for the creation and annihilation operators (V-1.155), one obtains the relation

$$u_k {u'_k}^* - u_k^* u'_k = i\hbar, \tag{V-1.160}$$

which determines the normalization of the Wronskian.

The choice of a specific function $u_k(\eta)$ corresponds to a particular prescription for the physical vacuum $|0\rangle$, defined by

$$\hat{a}_{\vec{k}}|0\rangle = 0. \tag{V-1.161}$$

A different choice for $u_k(\eta)$ is associated to a different decomposition into creation and annihilitation modes and thus to a different vacuum.

Let us now note that the wavelength associated with a given mode k can always be found *within* the Hubble radius provided one goes sufficiently far backwards in time, since the comoving Hubble radius is shrinking during inflation. In other words, for $|\eta|$ sufficiently big, one gets $k|\eta| \gg 1$. Moreover, for a wavelength smaller than the Hubble radius, one can neglect the influence of the curvature of spacetime and the mode behaves as in a Minkowski spacetime, as can also be checked explicitly with the equation of motion (V-1.156) (the effective mass is negligible for $k|\eta| \gg 1$). Therefore, the most natural physical prescription is to take the particular solution that corresponds to the usual Minkowski vacuum, i.e. $u_k \sim \exp(-ik\eta)$, in the limit $k|\eta| \gg 1$. In view of (V-1.157), this corresponds to the choice

$$u_k = \sqrt{\frac{\hbar}{2k}} e^{-ik\eta}\left(1-\frac{i}{k\eta}\right), \tag{V-1.162}$$

where the coefficient has been determined by the normalisation condition (V-1.160). This choice, in the jargon of quantum field theory on curved spacetimes, corresponds to the *Bunch-Davies vacuum*.

Finally, one can compute the *correlation function* for the scalar field ϕ in the vacuum state defined above. When Fourier transformed, the correlation function defines the *power spectrum* $\mathcal{P}_\phi(k)$:

$$\langle 0|\hat{\phi}(\vec{x}_1)\hat{\phi}(\vec{x}_2)|0\rangle = \int d^3k\; e^{i\vec{k}.(\vec{x}_1-\vec{x}_2)}\frac{\mathcal{P}_\phi(k)}{4\pi k^3}. \tag{V-1.163}$$

Note that the homogeneity and isotropy of the quantum field is used implicitly in the definition of the power spectrum, which is "diagonal" in Fourier space (homogeneity) and depends only on the norm of $\vec{k}$ (isotropy). In our case, we find

$$2\pi^2 k^{-3}\mathcal{P}_\phi = \frac{|u_k|^2}{a^2}, \qquad \text{(V-1.164)}$$

which gives in the limit when the wavelength is *larger than the Hubble radius*, i.e. $k|\eta| \ll 1$,

$$\mathcal{P}_\phi(k) \simeq \hbar \left(\frac{H}{2\pi}\right)^2 \qquad (k \ll aH) \qquad \text{(V-1.165)}$$

Note that, in the opposite limit, i.e. for wavelengths smaller than the Hubble radius ($k|\eta| \gg 1$), one recovers the usual result for fluctuations in Minkowski vacuum, $\mathcal{P}_\phi(k) = \hbar(k/2\pi a)^2$.

We have used a quantum description of the scalar field. But the cosmological perturbations are usually described by *classical random fields*. Roughly speaking, the transition between the quantum and classical (although stochastic) descriptions makes sense when the perturbations exit the Hubble radius. Indeed each of the terms in the Wronskian (V-1.160) is roughly of the order $\hbar/2(k\eta)^3$ in the super-Hubble limit and the non-commutativity can then be neglected. In this sense, one can see the exit outside the Hubble radius as a quantum-classical transition, although much refinement would be needed to make this statement more precise.

V-2 Quantum fluctuations with metric perturbations

Let us now move to the more realistic case of a perturbed inflaton field living in a perturbed cosmological geometry. The situation is more complicated than in the previous problem, because Einstein's equations imply that scalar field fluctuations must necessarily coexist with *metric fluctuations*. A correct treatment, either classical or quantum, must thus involve both the scalar field perturbations and metric perturbations.

In order to quantize this coupled system, the easiest procedure consists in identifying the *true degrees of freedom*, the other variables being then derived from them via constraint equations. As we saw in the classical analysis of cosmological perturbations, there exists only one scalar degree of freedom in the case of a single scalar field, which we must now identify.

The starting point is the action of the coupled system scalar field plus gravity expanded up to second order in the linear perturbations. Formally this can be written as

$$S[\bar{\phi} + \delta\phi, g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}] \;=\; S^{(0)}[\bar{\phi}, \bar{g}_{\mu\nu}] + S^{(1)}[\delta\phi, h_{\mu\nu}; \bar{\phi}, \bar{g}_{\mu\nu}]$$

$$+S^{(2)}[\delta\phi, h_{\mu\nu}; \bar{\phi}, \bar{g}_{\mu\nu}], \qquad \text{(V-2.166)}$$

where the first term $S^{(0)}$ contains only the homogeneous part, $S^{(1)}$ contains all terms linear in the perturbations (with coefficients depending on the homogeneous variables), and finally $S^{(2)}$ contains the terms quadratic in the linear perturbations. When one substitutes the FLRW equations of motion in $S^{(1)}$ (after integration by parts), one finds that $S^{(1)}$ vanishes, which is not very surprising since this is how one gets the homogeneous equations of motion, via the Euler-Lagrange equations, from the variation of the action.

The term $S^{(2)}$ is the piece we are interested in: the corresponding Euler-Lagrange equations give the equations of motion for the linear perturbations, which we have already obtained; but more importantly, this term enables us to quantize the linear perturbations and to find the correct normalization.

If one restricts oneself to the scalar sector, the quadratic part of the action depends on the four metric perturbations A, B, C, E, as well as on the scalar field perturbation $\delta\phi$. After some cumbersome manipulations, by using the FLRW equations of motion, one can show that the second-order action for scalar perturbations can be rewritten in terms of a single variable [14]

$$v = a\left(\delta\phi - \frac{\phi'}{\mathcal{H}}C\right), \qquad \text{(V-2.167)}$$

which is a linear combination mixing scalar field *and* metric perturbations. The variable v represents the true dynamical degree of freedom of the system, and one can check immediately that it is indeed a gauge-invariant variable.

In fact, v is proportional to the comoving curvature perturbation defined in (IV.5) and which, in the case of a single scalar field, takes the form

$$\mathcal{R} = -C + \frac{\mathcal{H}}{\phi'}\delta\phi. \qquad \text{(V-2.168)}$$

Note also that, if one can check a posteriori that v is the variable describing the true degree of freedom by expressing the action in terms of v only (modulo, of course, a multiplicative factor depending only on homogeneous quantities: the v defined here is such that it gives a canonical kinetic term in the action), one can identify v in a systematic way by resorting to Hamiltonian techniques, in particular the Hamilton-Jacobi equation [15].

With the variable v, the quadratic action takes the extremely simple form

$$S_v = \frac{1}{2}\int d\eta\ d^3x\ \left[v'^2 + \partial_i v \partial^i v + \frac{z''}{z}v^2\right], \tag{V-2.169}$$

with

$$z = a\frac{\phi'}{\mathcal{H}}. \tag{V-2.170}$$

This action is thus analogous to that of a scalar field in Minkowski spacetime with a time-dependent mass. One is thus back in a situation similar to the previous subsection, with the notable difference that the effective time-dependent mass is now z''/z, instead of a''/a.

The quantity we will be eventually interested in is the comoving curvature perturbation $\mathcal{R}$, which is related to the canonical variable v by the relation

$$v = -z\mathcal{R}. \tag{V-2.171}$$

Since, by analogy with (V-1.164), the power spectrum for v is given by

$$2\pi^2 k^{-3}\mathcal{P}_v(k) = |v_k|^2, \tag{V-2.172}$$

the corresponding power spectrum for $\mathcal{R}$ is found to be

$$2\pi^2 k^{-3}\mathcal{P}_{\mathcal{R}}(k) = \frac{|v_k|^2}{z^2}. \tag{V-2.173}$$

In the case of an inflationary phase in the *slow-roll* approximation, the evolution of ϕ and of H is much slower than that of the scale factor a. Consequently, one gets approximately

$$\frac{z''}{z} \simeq \frac{a''}{a}, \qquad (\text{slow} - \text{roll}) \tag{V-2.174}$$

and all results of the previous section obtained for u apply directly to our variable v in the slow-roll approximation. This implies that the properly normalized function corresponding to the Bunch-Davies vacuum is approximately given by

$$v_k \simeq \sqrt{\frac{\hbar}{2k}}e^{-ik\eta}\left(1 - \frac{i}{k\eta}\right). \tag{V-2.175}$$

In the super-Hubble limit $k|\eta| \ll 1$ the function v_k behaves like

$$v_k \simeq -\sqrt{\frac{\hbar}{2k}}\,\frac{i}{k\eta} \simeq i\sqrt{\frac{\hbar}{2k}}\,\frac{aH}{k}, \tag{V-2.176}$$

where we have used $a \simeq -1/(H\eta)$. Consequently, on scales larger than the Hubble radius, the power spectrum for $\mathcal{R}$ is found, combining (V-2.173), (V-2.170) and (V-2.175), to be given by

$$\mathcal{P}_{\mathcal{R}} \simeq \frac{\hbar}{4\pi^2}\left(\frac{H^4}{\dot{\phi}^2}\right)_{k=aH}, \tag{V-2.177}$$

where we have reintroduced the cosmic time instead of the conformal time. This is the famous result for the spectrum of scalar cosmological perturbations generated from vacuum fluctuations during a slow-roll inflation phase. Note that during slow-roll inflation, the Hubble parameter and the scalar field velocity slowly evolve: for a given scale, the above amplitude of the perturbations is determined by the value of H and $\dot{\phi}$ when the scale exited the Hubble radius. Because of this effect, the obtained spectrum is not strictly scale-invariant.

It is also instructive to recover the above result by a more intuitive derivation. One can think of the metric perturbations in the radiation era as resulting from the time difference for the end of inflation at different spatial points (separated by distances larger than the Hubble radius), the shift for the end of inflation being a consequence of the scalar field fluctuations $\delta\phi \sim H/2\pi$. Indeed,

$$\Psi_{\rm rad} \sim \frac{\delta a}{a} \sim H\delta t, \tag{V-2.178}$$

and the time shift is related to the scalar field fluctuations by $\delta t \sim \delta\phi/\dot{\phi}$, which implies

$$\Psi_{\rm rad} \sim \frac{H^2}{\dot{\phi}}, \tag{V-2.179}$$

which agrees with the above result since during the radiation era $\mathcal{R} = (3/2)\Psi$ (see Eq. (IV-7.145)). It is also worth noticing that, during inflation, in the case of the slow-roll approximation, the term involving C in the linear combination (V-2.167) defining v is negligible with respect to the term involving $\delta\phi$. One can therefore "ignore" the rôle of the metric perturbations *during inflation* in the computation of the quantum fluctuations and consider only the scalar field perturbations. But this simplification is valid only in the context of slow-roll approximation. It is not valid in the general case, as can be verified for inflation with a power-law scale factor.

V-3 Gravitational waves

We have focused so far our attention on scalar perturbations, which are the most important in cosmology. Tensor perturbations, or primordial gravitational waves, if ever detected in the future, would be a

remarkable probe of the early universe. In the inflationary scenario, like scalar perturbations, primordial gravitational waves are generated from vacuum quantum fluctuations. Let us now explain briefly this mechanism.

The action expanded at second order in the perturbations contains a tensor part, which given by

$$S_g^{(2)} = \frac{1}{64\pi G}\int d\eta\, d^3x\, a^2 \eta^{\mu\nu}\partial_\mu \bar{E}_j^i \partial_\nu \bar{E}_i^j, \tag{V-3.180}$$

where $\eta^{\mu\nu}$ denotes the Minkoswki metric. Apart from the tensorial nature of E_j^i, this action is quite similar to that of a scalar field in a FLRW universe (V-1.150), up to a renormalization factor $1/\sqrt{32\pi G}$. The decomposition

$$a\bar{E}_j^i = \sum_{\lambda=+,\times}\int \frac{d^3k}{(2\pi)^{3/2}} v_{k,\lambda}(\eta)\epsilon_j^i(\vec{k};\lambda)e^{i\vec{k}.\vec{x}} \tag{V-3.181}$$

where the $\epsilon_j^i(\vec{k};\lambda)$ are the polarization tensors, shows that the gravitational waves are essentially equivalent to two massless scalar fields (for each polarization) $\phi_\lambda = m_P\bar{E}_\lambda/2$.

The total power spectrum is thus immediately deduced from (V-1.164):

$$\mathcal{P}_g = 2\times\frac{4}{m_P^2}\times\hbar\left(\frac{H}{2\pi}\right)^2, \tag{V-3.182}$$

where the first factor comes from the two polarizations, the second from the renormalization with respect to a canonical scalar field, the last term being the power spectrum for a scalar field derived earlier. In summary, the tensor power spectrum is

$$\mathcal{P}_g = \frac{2\hbar}{\pi^2}\left(\frac{H}{m_P}\right)^2_{k=aH}, \tag{V-3.183}$$

where the label recalls that the Hubble parameter, which can be slowly evolving during inflation, must be evaluated when the relevant scale exited the Hubble radius during inflation.

V-4 Power spectra

Let us rewrite the scalar and tensor power spectra, respectively given in (V-2.177) and (V-3.183), in terms of the scalar field potential only. This can be done by using the slow-roll equations (III-2.49-III-2.50).

One finds for the scalar spectrum

$$\mathcal{P}_{\mathcal{R}} = \frac{1}{12\pi^2}\left(\frac{V^3}{m_P^6 V'^2}\right)_{k=aH} \tag{V-4.184}$$

with subscript meaning that the term on the right hand side must be evaluated at *Hubble radius exit* for the scale of interest. The scalar spectrum can also be written in terms of the first slow-roll parameter defined in (III-2.52), in which case it reads

$$\mathcal{P}_{\mathcal{R}} = \frac{1}{24\pi^2}\left(\frac{V}{m_P^4 \epsilon_V}\right)_{k=aH}. \tag{V-4.185}$$

From the observations of the CMB fluctuations,

$$\mathcal{P}_{\mathcal{R}}^{1/2} = \frac{1}{2\sqrt{6}\pi}\left(\frac{V^{1/2}}{m_P^2 \epsilon_V^{1/2}}\right) \simeq 5 \times 10^{-5}. \tag{V-4.186}$$

If ϵ_V is order 1, as in chaotic models, one can evaluate the typical energy scale during inflation as

$$V^{1/4} \sim 10^{-3} m_P \sim 10^{15}\mathrm{GeV}. \tag{V-4.187}$$

The tensor power spectrum, in terms of the scalar field potential, is given by

$$\mathcal{P}_g = \frac{2}{3\pi^2}\left(\frac{V}{m_P^4}\right)_{k=aH}. \tag{V-4.188}$$

The ratio of the tensor and scalar amplitudes is proportional to the slow-roll parameter ϵ_V:

$$r \equiv \frac{\mathcal{P}_g}{\mathcal{P}_{\mathcal{R}}} = 16\epsilon_V. \tag{V-4.189}$$

The scalar and tensor spectra are almost scale invariant but not quite since the scalar field evolves slowly during the inflationary phase. In order to evaluate quantitatively this variation, it is convenient to introduce a scalar *spectral index* as well as a tensor one, defined respectively by

$$n_S(k) - 1 = \frac{d\ln \mathcal{P}_{\mathcal{R}}(k)}{d\ln k}, \quad n_T(k) = \frac{d\ln \mathcal{P}_g(k)}{d\ln k}. \tag{V-4.190}$$

One can express the spectral indices in terms of the slow-roll parameters. For this purpose, let us note that, in the slow-roll approximation, $d\ln k = d\ln(aH) \simeq d\ln a$, so that

$$\frac{d\phi}{d\ln a} = \frac{\dot\phi}{H} \simeq -\frac{V'}{3H^2} \simeq -m_P^2\frac{V'}{V}, \tag{V-4.191}$$

where the slow-roll equations (III-2.49-III-2.50) have been used. Therefore, one gets

$$n_s(k) - 1 = 2\eta_V - 6\epsilon_V, \tag{V-4.192}$$

where ϵ_V and η_V are the two slow-roll parameters defined in (III-2.52) and (III-2.54). Similarly, one finds for the tensor spectral index

$$n_T(k) = -2\epsilon_V. \tag{V-4.193}$$

Comparing with Eq. (V-4.189), this yields the relation

$$r = -8n_T, \tag{V-4.194}$$

the so-called *consistency relation* which relates purely *observable* quantities. This means that if one was able to observe the primordial gravitational waves and measure the amplitude and spectral index of their spectrum, a rather formidable task, then one would be able to test directly the paradigm of single field slow-roll inflation.

Finally, let us mention the possibility to get information on inflation from the measurement of the running of the spectral index. Introducing the second-order slow-roll parameter

$$\xi_V = m_P^4 \frac{V'V'''}{V^2}, \tag{V-4.195}$$

the running is given by

$$\frac{dn_s}{d\ln k} = -24\epsilon_V^2 + 16\epsilon_V\eta_V - 2\xi_V. \tag{V-4.196}$$

As one can see, the amplitude of the variation depends on the slow-roll parameters and thus on the models of inflation.

V-5 Conclusions

To conclude, let us mention the existence of more sophisticated models of inflation, such as models where several scalar fields contribute to inflation. In contrast with the single inflaton case, which can generate only adiabatic primordial fluctuations, because all types of matter are decay product of the same inflaton, *multi-inflaton models* can generate both adiabatic and isocurvature perturbations, which can even be correlated [16].

Another recent direction of research is the possibility to disconnect the fluctuations of the inflaton, the field that drives inflation, from the observed cosmological perturbations, which could have been generated from the quantum fluctuations of another scalar field [17].

From the theoretical point of view an important challenge remains to identify viable and natural candidates for the inflaton field(s) in the framework of high energy physics, with the hope that future observations of the cosmological perturbations will be precise enough to discriminate between various candidates and thus give us a clue about which physics really drove inflation.

Alternative scenarios to inflation can also been envisaged, as long as they can predict unambiguously primordial fluctuations compatible with the present observations. In this respect, one must emphasize that the cosmological perturbations represent today essentially the only observational window that gives access to the very high energy physics, hence the importance for any early universe model to be able to give firm predictions for the primordial fluctuations it generates.

Acknowledgements:

I would like to thank the organizers of this Cargese school for inviting me to lecture at this very nice village in Corsica. I also wish to thank the other lecturers and participants for their remarks and questions, as well as Gérard Smadja and Filippo Vernizzi for their careful reading of the draft.

References

[1] R. Wald, *General Relativity* (Chicago, 1984); S.M. Carroll, *Spacetime and Geometry*, (Addison Wesley, San Francisco, 2004).

[2] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[3] A. A. Starobinsky, Phys. Lett. B **91**, 99 (1980).

[4] A. D. Linde, Phys. Lett. B **108**, 389 (1982); A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[5] A. D. Linde, Phys. Lett. B **129**, 177 (1983).

[6] A. D. Linde, Phys. Rev. D **49**, 748 (1994) [arXiv:astro-ph/9307002].

[7] A. Linde, *Particle physics and Inflationary Cosmology*, (Harwood, Chur, 1990).

[8] D.H. Lyth and A. Riotto, Phys. Rept. **314**, 1 (1998) [hep-ph/9807278].

[9] A.R. Liddle and D.H. Lyth, *Cosmological inflation and large-scale structure*, Cambridge University Press (2000).

[10] H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. **78**, 1 (1984); V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Phys. Rept. **215**, 203 (1992).

[11] C. P. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995) [arXiv:astro-ph/9506072].

[12] G. Smoot, these proceedings.

[13] V. F. Mukhanov and G. V. Chibisov, JETP Lett. **33**, 532 (1981) [Pisma Zh. Eksp. Teor. Fiz. **33**, 549 (1981)]; A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982); A. A. Starobinsky, Phys. Lett. B **117**, 175 (1982); S. W. Hawking, Phys. Lett. B **115**, 295 (1982).

[14] V. F. Mukhanov, Phys. Lett. B **218**, 17 (1989).

[15] D. Langlois, Class. Quant. Grav. **11**, 389 (1994).

[16] D. Langlois, Phys. Rev. D **59**, 123512 (1999) [arXiv:astro-ph/9906080].

[17] D. H. Lyth and D. Wands, Phys. Lett. B **524**, 5 (2002) [arXiv:hep-ph/0110002] ; G. Dvali, A. Gruzinov and M. Zaldarriaga, Phys. Rev. D **69**, 023505 (2004) [arXiv:astro-ph/0303591].

COSMIC MICROWAVE BACKGROUND COSMOLOGY

New Economy Model Universe

George F. Smoot
LBNL & Department of Physics
366 Le Conte Hall
University of California
Berkeley, CA 94720
gfsmoot@lbl.gov

Abstract We review the CMB spectrum and the upper limits on the effects which would distort the Planck black-body distribution, such as Compton scattering, reionisation, plasma effects. The experimental analysis of the angular anisotropies is discussed, as well as the physics pertinent to the different angular momentum regions. The connection between cosmology and the observed inhomogeneities and the determination of the cosmological parameters are summarised, as well as the underlying assumptions. The physical origin of the polarisation of the CMB is explained, and the first experimental results are recalled.

Keywords: CMB, Cosmology, Sachs-Wolf Plateau, Acoustic peaks, Damping tail, Polarisation, Foregrounds

I. The CMB Spectrum

The energy content in radiation from beyond our Galaxy is dominated by the Cosmic Microwave Background (CMB), discovered in 1965 by Penzias and Wilson[1] The spectrum of the CMB (shown in Figure 1) is well described by a blackbody function with $T = 2.73$K. By the late 1970s it was becoming clear that the spectrum was thermal.

The most accurate measurements of the CMB spectrum to date have come from the Far InfraRed Absolute Spectrophotometer (FIRAS) on the COsmic Background Explorer (COBE) ([2]). FIRAS is a differential spectrometer that measured the difference between the sky and an internal reference source that was very nearly a blackbody. An external calibrator was periodically moved into the sky horn to replace the sky

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 279–310.

signal with a very, very good blackbody. The FIRAS observations show that the rms deviation from a blackbody is only 50 parts per million of the peak I_ν of the blackbody ([3]) and a recalibration of the thermometers on the external calibrator yield a blackbody temperature of 2.725 ± 0.001 K ([4]). Upper limits on any deviations from of the CMB from a blackbody place strong constraints on energy transfer between the CMB and matter at redshifts less than 2×10^6.

A non-interacting Planckian distribution of temperature $T_{\rm i}$ at redshift $z_{\rm i}$ transforms with the universal expansion to another Planckian distribution at redshift $z_{\rm f}$ with temperature $T_{\rm f}/(1+z_{\rm f}) = T_{\rm i}/(1+z_{\rm i})$. Hence the thermal spectrum, once established (e.g. at the nucleosynthesis Epoch or earlier), is preserved by the expansion, in spite of the fact that photons decoupled from matter at early times.

This spectral form is one of the main pillars of the hot Big Bang model for the early Universe. The lack of any observed deviations from a blackbody spectrum constrains physical processes over the history of the universe (at redshifts $z < 10^8$). All viable present cosmological models predict however a very nearly Planckian spectrum so are not stringently limited.

The observed cosmic microwave background (CMB) radiation provides strong evidence for the hot big bang. The success of primordial nucleosynthesis calculations requires a cosmic background radiation (CBR) characterized by a temperature $kT \sim 1\,\mathrm{MeV}$ at a redshift of $z \simeq 10^9$. In their pioneering work, Gamow, Alpher, and Herman[5] realized this and predicted the existence of a faint residual relic, primordial radiation, with a present temperature of a few degrees. The observed CMB is interpreted as the current manifestation of the required CBR.

Atomic fine structure line observations along the lines of sight to distant quasars constrain the temperature out to $z \approx 2 - 3$, giving direct support for the CMB being hotter at early times[7]. Because there are about 10^9 photons per nucleon, the transition from the ionized primordial plasma to neutral atoms at $z \sim 1000$ does not significantly alter the CBR spectrum[6]. CMB temperature variations observed at this epoch provide further support for the hot big bang, as well as for the presence of primordial density perturbations which grew into today's cosmological structure through gravitational instability.

The CMB is very nearly isotropic, but a dipole anisotropy of amplitude 3.346 mK is believed to be due to the motion of the Solar System barycenter at 368 ± 2 km/sec relative to the CMB center of momentum frame and thus likely the center of mass of the observable universe.

Higher order intrinsic inhomogeneities in the Universe existed at the time of last scattering. While the photons have traveled freely since the

last scattering, 380,000 years after the initial moments of Big Bang, the spectrum of the CMB was established very early and has not changed significantly after one to two months into the Big Bang. On the other hand, the inhomogeneities traced by the CMB photons have been in place since the inflationary epoch, 10^{-35} sec after the Big Bang.

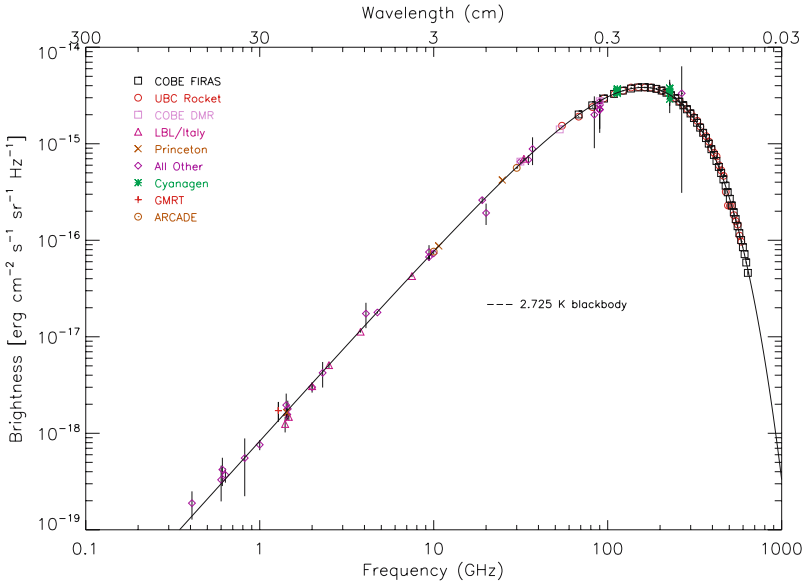

Figure 1. Precise measurements of the CMB spectrum. The line represents a 2.73 K blackbody, which describes the spectrum very well, especially around the peak of intensity. The spectrum is less well constrained at 10 cm and longer wavelengths.

I-1 The CMB frequency spectrum

The remarkable precision with which the CMB spectrum is fitted by a Planckian distribution provides limits on possible energy releases in the early Universe, at roughly the fractional level of 10^{-4} of the CBR energy, for redshifts $\sim 10^7$ (corresponding to epochs ~ 1 year). The following three important classes of theoretical spectral distortions (see figure CMBdis generally correspond to energy releases at different epochs. They result from the CBR photon interactions with a hot electron gas at temperature T_e.

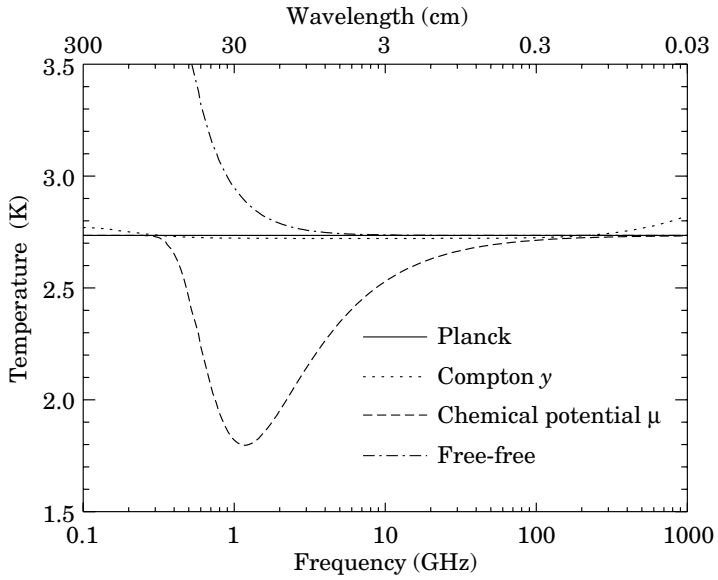

Figure 2. The shapes of expected, but so far unobserved, CMB distortions, resulting from energy-releasing processes at different epochs.

I-2 Compton distortion (Late energy release ($z \sim 10^5$)

Compton scattering ($\gamma e \rightarrow \gamma' e'$) of the CBR photons by a hot electron gas creates spectral distortions by transferring energy from the electrons to the photons. Compton scattering cannot produce a Planckian spectrum for $y \sim 1$, where

$$y = \int_0^z \frac{kT_e(z') - kT_\gamma(z')}{m_e c^2}\, \sigma_T\, n_e(z')\, c\, \frac{dt}{dz'}\, dz',$$

is the integral of the number of interactions, $\sigma_T\ n_e(z)\ c\ dt$, times the mean-fractional photon-energy change per collision[72]. For $T_e \gg T_\gamma$ y is also proportional to the integral of the electron pressure $n_e k T_e$ along the line of sight. For standard thermal histories $y < 1$ for epochs later than $z \simeq 10^5$.

The resulting CMB distortion is a temperature decrement

$$\Delta T_{\rm RJ} = -2y\, T_\gamma$$

in the Rayleigh-Jeans ($x \equiv h\nu/kT \ll 1$) portion of the spectrum, and a rise in temperature in the Wien ($h\nu/kT \gg 1$) region, *i.e.* photons are

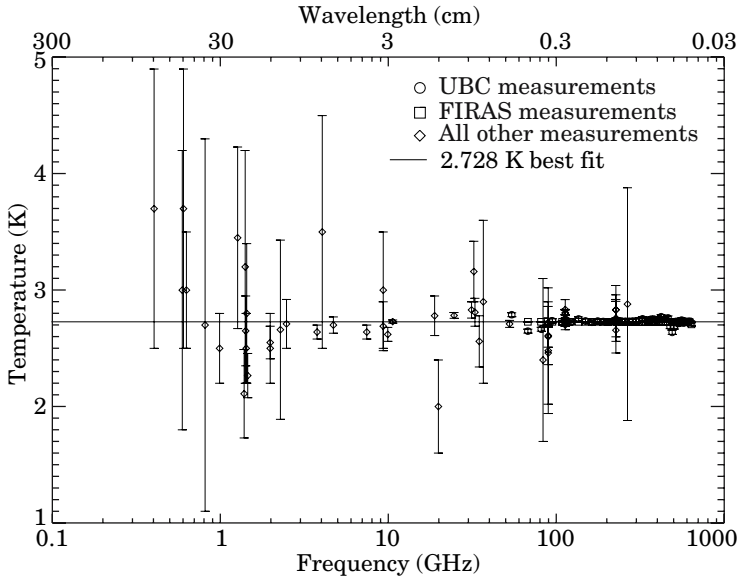

Figure 3. Observed thermodynamic temperature as a function frequency.

shifted from low to high frequencies. The magnitude of the distortion is related to the total energy transfer[72] ΔE by

$$\Delta E/E_{\rm CBR} = e^{4y} - 1 \simeq 4y.$$

A prime candidate for producing a Comptonized spectrum is a hot intergalactic medium. A hot ($T_e > 10^5$ K) medium in clusters of galaxies can and does produce a partially Comptonized spectrum as seen through the cluster, known as the Sunyaev-Zel'dovich effect[8]. Based upon X-ray data and the WMAP estimate on the re-ionization optical depth, the predicted large angular scale total combined effect of the hot intracluster medium and the ionized intergalactic medium should produce $y \sim 10^{-6}$ [9]. Detection of the S-Z effect through clusters demonstrates that the CMB is universal and can be used to estimate the Hubble constant, and counts of such clusters as a function of redshift hold the promise of constraining the equation of state of the Dark Energy.

I-3 Bose-Einstein or chemical potential distortion (Early energy release ($z \sim 10^5$–10^7)

After many Compton scatterings ($y >> 1$), the photons and electrons will reach statistical (not thermodynamic) equilibrium, because Compton scattering conserves photon number. This equilibrium is described

by the Bose-Einstein distribution with non-zero chemical potential:

$$n = \frac{1}{e^{x+\mu_0} - 1},$$

where $x \equiv h\nu/kT$ and $\mu_0 \simeq 1.4\ \Delta E/E_{\rm CBR}$, with μ_0 being the dimensionless chemical potential that is required to conserve photon number. The collisions of electrons with nuclei in the plasma produce free-free (thermal bremsstrahlung) radiation: $eZ \rightarrow e'Z'\gamma$. Free-free emission thermalizes the spectrum to the plasma temperature at long (> centimeter) wavelengths and Compton scattering begins to shift these photons upward.

The equilibrium Bose-Einstein distribution results from the oldest non-equilibrium processes ($10^5 < z < 10^7$), such as the decay of relic particles or primordial inhomogeneities. Note that free-free emission (thermal bremsstrahlung) and radiative-Compton scattering effectively erase any distortions[10] to a Planckian spectrum for epochs earlier than $z \sim 10^7$.

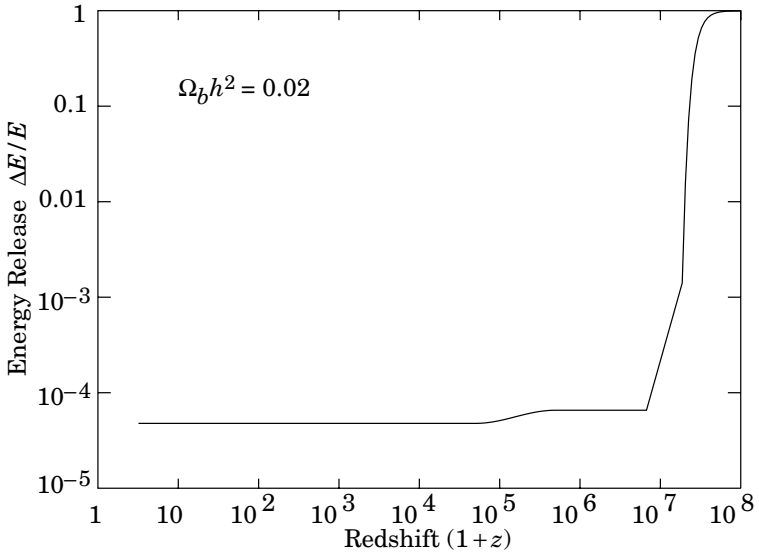

Figure 4. Upper Limits (95% CL) on fractional energy ($\Delta E/E_{\rm CBR}$) releases from processes at different epochs as set by resulting lack of CMB spectral distortions. These can be translated into constraints on the mass, lifetime and photon branching ratio of unstable relic particles, with some additional dependence on cosmological parameters such as Ω_B [?]. This figure shows that the observed CMB spectrum is a relic from $z \sim 10^7$.

I-4 Free-free distortion (Very late energy release ($z \ll 10^3$)

Free-free emission can create rather than erase spectral distortion in the late Universe, for recent reionization ($z < 10^3$) and from a warm intergalactic medium. The distortion arises because of the lack of Comptonization at recent epochs. The effect on the present-day CMB spectrum is described by

$$\Delta T_{\rm ff} = T_\gamma \, Y_{\rm ff}/x^2,$$

where T_γ is the undistorted photon temperature, $x \equiv h\nu/kT$ is the dimensionless frequency, and $Y_{\rm ff}/x^2$ is the optical depth to free-free emission:

$$Y_{\rm ff} = \int_0^z \frac{T_e(z') - T_\gamma(z')}{T_e(z')} \frac{8\pi e^6 h^2 n_e^2 \, g}{3 m_e (kT_\gamma)^3 \sqrt{6\pi \, m_e \, kT_e}} \frac{dt}{dz'} dz'.$$

Here h is Planck's constant, n_e is the electron density and g is the Gaunt factor[11].

I-5 Effect of Reionization of the Universe:

It is well-established that the Universe is nearly fully reionized by a redshift of $z \sim 6$ by the Gunn-Peterson test[12]. The WMAP first year data indicate that the re-ionization of the universe began earlier and thus that the Universe was partially ionized back to a redshift of about 20 with an effective optical depth of $\tau = 0.17 \pm 0.04$)[17].

Typically, we might anticipate at during the time of ionization, the typical temperature of the electrons will be around 10^4 K. This would leave us with a typical Compton y-parameter of $y \sim 3 \times 10^{-7}$, which is near that expected to be produced by the integrated ionized galactic cluster medium, yielding our estimate of $y \sim 10^{-6}$ for the net of these.

The expected effect in terms of the free-free distortion parameter is $Y_{ff} \sim 10^{-5}$. Both of these effects are somewhat below the current level of detection via spectral measurement.

I-6 Spectrum summary

The CMB spectrum is consistent with a blackbody distribution over more than three decades of frequency around the peak. The best-fit to the COBE FIRAS data yields $T_\gamma = 2.725 \pm 0.002$ K (95% CL)[4]. The

following table is a summary of all CMB spectrum measurements:

$$\begin{aligned}
T_\gamma &= 2.725 \pm 0.002K \ (95\% \text{ CL})\,; \\
n_\gamma &= (2\zeta(3)/\pi^2)\, T_\gamma^3 \simeq 411\,\mathrm{cm}^{-3}\,; \\
\rho_\gamma &= (\pi^2/15)\, T_\gamma^4 \simeq 4.64 \times 10^{-34}\,\mathrm{g\,cm}^{-3} \simeq 0.260\,\mathrm{eV\,cm}^{-3}\,; \\
|y| &< 1.2 \times 10^{-5} \qquad (95\%\ CL)\,; \\
|\mu_0| &< 9 \times 10^{-5} \qquad (95\%\ CL)\,; \\
|Y_{\mathrm{ff}}| &< 1.9 \times 10^{-5} \qquad (95\%\ CL)\,.
\end{aligned}$$

These limits[13] correspond to constraints[13][14] on energetic processes

$\Delta E/E_{\mathrm{CBR}} < \ 2 \times 10^{-4}$ occurring between redshifts 10^3 and 5×10^6 (see Figure 4).

We notice that the frequency spectrum is extremely well-described by a Planckian and that the lack of anticipated distortions allow us to set cosmological limits on processes dissipating energy that eventually effects the CMB, either directly or through its interactions with the heated medium. This provides us with a well-defined tool for investigating and imaging the early universe.

II. Description of CMB Anisotropies

Another observable quantity inherent in the CMB is the variation in temperature (or intensity) from one part of the microwave sky to another [15]. Since the first detection of these anisotropies by the *COBE* satellite [16], there has been intense activity to map the sky at increasing levels of sensitivity and angular resolution. A series of ground- and balloon-based measurements has recently been joined by the first results from NASA's Wilkinson Microwave Anisotropy Probe (*WMAP*)[17]. These observations have led to a stunning confirmation of the 'Standard Model of Cosmology.' In combination with other astrophysical data, the CMB anisotropy measurements place quite precise constraints on a number of cosmological parameters, and have launched us into an era of precision cosmology.

Observations show that the CMB contains anisotropies at roughly the 10^{-5} level, over a wide range of angular scales. These anisotropies are usually expressed by using a spherical harmonic expansion of the CMB sky:

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi).$$

The vast majority of the cosmological information is contained in the temperature 2 point function, i.e., the variance as a function of separation angle θ. Equivalently, the power per unit Legendre polynominal number $\ln \ell$ is $\ell \sum_m |a_{\ell m}|^2 / 4\pi = \ell(2\ell + 1)C_\ell / 4\pi$.

II-1 The Monopole

The CMB has a mean temperature of $T_\gamma = 2.725 \pm 0.001\,\mathrm{K}$ (1σ)[4], which can be considered as the monopole component of CMB maps, a_{00}. Since all mapping experiments involve difference measurements, they are insensitive to this average level. Monopole measurements can only be made with absolute temperature devices, such as the FIRAS instrument on the *COBE* satellite[4]. Such measurements of the spectrum are consistent with a blackbody distribution over more than three decades in frequency. A blackbody of the measured temperature corresponds to $n_\gamma = (2\zeta(3)/\pi^2)\,T_\gamma^3 \simeq 411\,\mathrm{cm}^{-3}$ and $\rho_\gamma = (\pi^2/15)\,T_\gamma^4 \simeq 4.64 \times 10^{-34}\,\mathrm{g\,cm}^{-3} \simeq 0.260\,\mathrm{eV\,cm}^{-3}$.

II-2 The CMB Dipole Anisotropy

The largest anisotropy is the $\ell = 1$ (dipole) first spherical harmonic, with amplitude $3.346 \pm 0.017\,\mathrm{mK}$[17]. The dipole is interpreted to be the result of the Doppler shift caused by the solar system motion relative to the nearly isotropic blackbody field, as confirmed by measurements of the velocity field of local galaxies[18]. The motion of an observer with velocity $\beta = v/c$ relative to an isotropic Planckian radiation field of temperature T_0 produces a Doppler-shifted temperature pattern

$$\begin{aligned} T(\theta) &= T_0(1-\beta^2)^{1/2}/(1-\beta\cos\theta) \\ &= T_0\,\left(1+\beta\cos\theta+(\beta^2/2)\cos 2\theta + O(\beta 3)\right). \end{aligned}$$

At every point in the sky, the spectrum is essentially blackbody, but the spectrum of the dipole is the differential of a blackbody spectrum, as confirmed by [19]. Assuming this to be true, then a combination of well-calibrated anisotropy measurements traces out the differential of the blackbody spectrum and can be used to find its temperature. This has been done by the COBE DMR using the dipole anisotropy and is the sources of the observations labeled COBE DMR on the spectral plots.

The dipole anisotropy pattern is believed due to the motion of the observer relative to the rest of the Universe or at least the center of momentum frame for the CMB. Shortly after the Cosmic Microwave Background (CMB) was discovered, the first indications of anisotropy in the CMB were reported ([20]). Additional observations [21] and by [22] also indicated variations in the observed sky temperature. The definitive observations of the dipole ([23]) came in the late 1970's showed a very definite cosine pattern as expected for a Doppler effect, and placed an upper limit on any further variations in T_{CMB} and contamination by Galactic foregrounds. Further improvements in the measurement of the dipole anisotropy were made by the Differential Microwave Radiometers

(DMR) experiment on COBE ([24] and by the Wilkinson Microwave Anisotropy Probe ([25]).

The implied velocity[13][17] for the solar system barycenter is $v = 368 \pm 2\,\mathrm{km\,s^{-1}}$, assuming a value $T_0 = T_\gamma$, towards $(\ell, b) = (263.85^\circ \pm 0.10^\circ, 48.25^\circ \pm 0.04^\circ)$. Such a solar system velocity implies a velocity for the Galaxy and the Local Group of galaxies relative to the CMB. The derived value is $v_{\mathrm{LG}} = 627 \pm 22\,\mathrm{km\,s^{-1}}$ toward $(\ell, b) = (276^\circ \pm 3^\circ, 30^\circ \pm 3^\circ)$, where most of the error comes from uncertainty in the velocity of the solar system relative to the Local Group.

The dipole is a frame dependent quantity, and one can thus determine the 'absolute rest frame' of the Universe as that in which the CMB dipole would be zero. Our velocity relative to the Local Group, as well as the velocity of the Earth around the Sun, and any velocity of the receiver relative to the Earth, is normally removed for the purposes of CMB anisotropy study.

III. Theory

The first theoretical predictions of $\Delta T/T = 10^{-2}$ ([26]) and $\Delta T/T = 10^{-3.5}$ ([27]) were superseded by predictions based on cold dark matter ([28], [29]). These CDM predictions were consistent with the small anisotropy seen by COBE and furthermore predicted a series of peaks at a particular angular scale due to acoustic oscillations in the baryon/photon fluid prior to recombination. The position of these peaks and other peaks in the angular power spectrum of the CMB anisotropy depends on a combination of the density parameter Ω_m and the vacuum energy density Ω_V, so this peak provides a means to determine the density of the Universe ([30]). A tentative detection of the first and largest peak at the position predicted for a flat Universe was made by 1994 ([31]). The peak was localized to $\ell_{pk} = 229 \pm 8.5$ ([32]) by the beginning of 2000. Later the BOOMERanG group claimed to have made a dramatic improvement in this datum to $\ell_{pk} = 197 \pm 6$ ([33]). This smaller value for ℓ_{pk} favored a moderately closed model for the Universe. However, at the same time the MAXIMA group found the first peak to be located at $\ell_{pk} = 220 \pm 15$ [34] In the next year BOOMERanG recalibrated and improved pointing [35]-[36], MAXIMA reported new results [37], DASI reported its observations [38] resulting in improved estimates of the first peak location and the clear establishment of the second and third peaks. Those observations combined with Supernovae Ia observations of and accelerating Universe [39] resulted in a standard ΛCDM model of cosmology. More results came in 2003 including *WMAP* ([40]) which gives the first peak location $\ell_{pk} = 220.1 \pm 0.8$ and higher peaks observed by ACBAR [41] all

of which are consistent with a flat ΛCDM model and provide improved cosmological constraints on the parameters of that model.

III-1 Higher Order Multipoles

Excess variance in CMB maps at higher multipoles ($\ell \geq 2$) is interpreted as being the result of perturbations in the energy density of the early Universe, manifesting themselves at the epoch of the last scattering of the CMB photons. In the hot Big Bang picture, this happens at a redshift $z \simeq 1100$, with little dependence on the details of the model. The process by which the hydrogen and helium nuclei can hold onto their electrons is usually referred to as recombination[43]. Before this epoch, the CMB photons are tightly coupled to the baryons, while afterwards they can freely stream towards us.

Theoretical models generally predict that the $a_{\ell m}$ modes are Gaussian random fields, and all tests are consistent with this simplifying assumption [44]. With this assumption, and if there is no preferred axis, then it is the variance of the temperature field which carries the cosmological information, rather than the values of the individual $a_{\ell m}$s; in other words the power spectrum in ℓ fully characterizes the anisotropies. The power at each ℓ is $(2\ell + 1)C_\ell/(4\pi)$, where $C_\ell \equiv \left\langle |a_{\ell m}|^2 \right\rangle$, and a statistically isotropic sky means that all ms are equivalent. We use our estimators of the C_ℓs to constrain their expectation values, which are the quantities predicted by a theoretical model. For an idealized full-sky observation, the variance of each measured C_ℓ (the variance of the variance) is $[2/(2\ell+1)]C_\ell^2$. This sampling uncertainty (known as cosmic variance) comes about because each C_ℓ is $\chi 2$ distributed with $(2\ell + 1)$ degrees of freedom for our observable volume of the Universe. For partial sky coverage, $f_{\rm sky}$, this variance is increased by $1/f_{\rm sky}$ and the modes become partially correlated.

It is important to understand that theories predict the expectation value of the power spectrum, whereas our sky is a single realization. Hence the 'cosmic variance' is an unavoidable source of uncertainty when constraining models; it dominates the scatter at lower ℓs, while the effects of instrumental noise and resolution dominate at higher ℓs.

III-2 Angular Resolution and Binning

There is no one-to-one conversion between the angle subtended by a particular wavevector projected on the sky and multipole ℓ. However, a single spherical harmonic $Y_{\ell m}$ corresponds to angular variations of $\theta \sim \pi/\ell$. CMB maps contain anisotropy information from the size of the map (or in practice some fraction of that size) down to the beam-size of

the instrument, σ. One can think of the effect of a Gaussian beam as rolling off the power spectrum with the function $\mathrm{e}^{-\ell(\ell+1)\sigma 2}$.

For less than full sky coverage, the ℓ modes are correlated. Hence, experimental results are usually quoted as a series of 'band powers', defined as estimators of $\ell(\ell+1)C_\ell/2\pi$ over different ranges of ℓ. because of the strong foreground signals in the Galactic Plane, even 'all-sky' surveys, such as *COBE* and *WMAP* involve a cut sky. The amount of binning required to obtain uncorrelated estimates of power also depends on the map size.

IV. Cosmological Parameters

The current 'Standard Model' of cosmology contains around 10 free parameters (some versions are up to about 15 but 10 describes most of what is currently known). The basic framework is the Friedmann-Robertson-Walker metric (i.e., a universe that is approximately homogeneous and isotropic on large scales), with density perturbations laid down at early times and evolving into today's structures. These perturbations can be either 'adiabatic' (meaning that there is no change to the entropy per particle for each species, *i.e.*, $\delta\rho/\rho$ for matter is $(3/4)\delta\rho/\rho$ for radiation) or 'isocurvature' (meaning that, for example, matter perturbations compensate radiation perturbations so that the total energy density remains unperturbed, *i.e.*, $\delta\rho$ for matter is $-\delta\rho$ for radiation). These different modes give rise to distinct phases during growth, and the adiabatic scenario is strongly preferred by the data. Models that generate mainly isocurvature type perturbations (such as most topological defect scenarios) are no longer considered to be viable.

Within the adiabatic family of models, there, is in principle, a free function describing how the comoving curvature perturbations, $\mathcal{R}$, vary with scale. In inflationary models, the Taylor series expansion of $\ln\mathcal{R}(\ln k)$ has terms of steadily decreasing size. For the simplest models, there are thus 2 parameters describing the initial conditions for density perturbations: the amplitude and slope of the power spectrum, $\left\langle|\mathcal{R}|^2\right\rangle \propto k^n$. This can be explicitly defined, for example, through:

$$\Delta^2_{\mathcal{R}} \equiv (k^3/2\pi^2)\left\langle|\mathcal{R}|^2\right\rangle,$$

and using $A^2 \equiv \Delta^2_{\mathcal{R}}(k_0)$ with $k_0 = 0.05\,\mathrm{Mpc}^{-1}$. There are many other equally valid definitions of the amplitude parameter, and we caution that the relationships between some of them can be cosmology dependent. In 'slow roll' inflationary models this normalization is proportional to the combination $V^3/(V')^2$, for the inflationary potential $V(\phi)$. The slope n also involves V'', and so the combination of A and n can, in principle, constrain potentials.

Inflationary models can generate tensor (gravity wave) modes as well as scalar (density perturbation) modes. This fact introduces another parameter measuring the amplitude of a possible tensor component, or equivalently the ratio of the tensor to scalar contributions. The tensor amplitude $A_{\rm T} \propto V$, and thus one expects a larger gravity wave contribution in models where inflation happens at higher energies. The tensor power spectrum also has a slope, often denoted $n_{\rm T}$, but since this seems likely to be extremely hard to measure, it is sufficient for now to focus only on the amplitude of the gravity wave component. It is most common to define the tensor contribution through r, the ratio of tensor to scalar perturbation spectra at large scales (say $k = 0.002\,{\rm Mpc}^{-1}$). There are other definitions in terms of the ratio of contributions to C_2, for example. Different inflationary potentials will lead to different predictions, e.g. for $\lambda\phi^4$ inflation, $r = 0.32$, while other models can have arbitrarily small values of r. In any case, whatever the specific definition, and whether they come from inflation or something else, the 'initial conditions' give rise to a minimum of 3 parameters: A, n and r.

The background cosmology requires an expansion parameter (the Hubble Constant, H_0, often represented through $H_0 = 100\,h\,{\rm km\,s^{-1}Mpc^{-1}}$) and several parameters to describe the matter and energy content of the Universe. These are usually given in terms of the critical density, *i.e.* for species 'x , $\Omega_{\rm x} = \rho_{\rm x}/\rho_{\rm crit}$, where $\rho_{\rm crit} = 3H_0^2/8\pi G$. Since physical densities $\rho_{\rm x} \propto \Omega_{\rm x}h^2 \equiv \omega_{\rm x}$ are what govern the physics of the CMB anisotropies, it is these ωs that are best constrained by CMB data. In particular CMB observations constrain $\Omega_{\rm B}h^2$ for baryons and $\Omega_{\rm M}h^2$ for baryons plus Cold Dark Matter.

The contribution of a cosmological constant Λ (or other form of Dark Energy) is usually included through a parameter which quantifies the curvature, $\Omega_{\rm K} \equiv 1 - \Omega_{\rm tot}$, where $\Omega_{\rm tot} = \Omega_{\rm M} + \Omega_\Lambda$. The radiation content, while in principle a free parameter, is precisely enough determined through the measurement of T_γ.

The main effect of astrophysical processes on the C_ℓs comes through reionization. The Universe became reionized at some redshift long after recombination, affecting the CMB through the integrated Thomson scattering optical depth:

$$\tau = \int_0^{z_{\rm i}} \sigma_{\rm T} n_{\rm e}(z) \frac{dt}{dz} dz,$$

where $\sigma_{\rm T}$ is the Thomson cross-section, $n_{\rm e}(z)$ is the number density of free electrons (which depends on astrophysics) and dt/dz is fixed by the background cosmology. In principle, τ can be determined from the small scale power spectrum together with the physics of structure formation

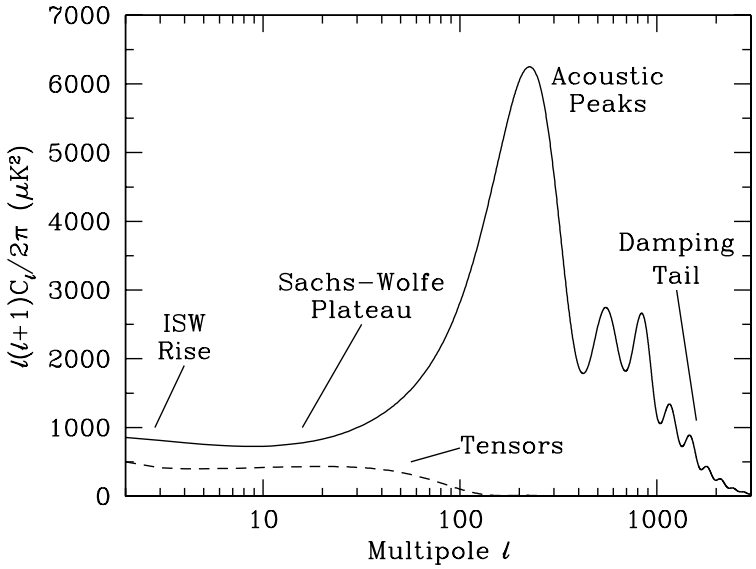

Figure 5. Plot of the theoretical CMB anisotropy power spectrum, using a standard ΛCDM model from `CMBFAST`. The x-axis is logarithmic here. The regions are labeled as in the text: the ISW Rise; Sachs-Wolfe Plateau; Acoustic Peaks; and Damping Tail. Also shown is the shape of the tensor (gravity wave) contribution, with an arbitrary normalization.

and feedback processes. However, this is a sufficiently complicated calculation that τ needs to be considered as a free parameter.

Thus we have 8 basic cosmological parameters: A, n, r, h, $\Omega_{\rm B}h2$, $\Omega_{\rm M}h2$, $\Omega_{\rm tot}$, and τ. One can add additional parameters to this list, particularly when using the CMB in combination with other data sets. The next most relevant ones might be: $\Omega_\nu h2$, the massive neutrino contribution; w ($\equiv p/\rho$), the equation of state parameter for the Dark Energy; and $dn/d\ln k$, measuring deviations from a constant spectral index. To these 11 one could of course add further parameters describing additional physics, such as details of the reionization process, features in the initial power spectrum, a sub-dominant contribution of isocurvature modes, *etc.*

As well as these underlying parameters, there are other quantities that can be derived from them. Such quantities include the actual Ωs of the various components (e.g. $\Omega_{\rm M}$), the variance of density perturbations at particular scales (e.g.- σ_8), the age of the Universe today (t_0), the age of the Universe at recombination, reionization, *etc.*

V. Physics of Anisotropies

The cosmological parameters affect the anisotropies through the well understood physics of the evolution of linear perturbations within a background FRW cosmology. There are very effective, fast, and publicly-available software codes for computing the CMB anisotropy, polarization, and matter power spectra, e.g., `CMBFAST`[45] and `CAMB`[46]. `CMBFAST` is the most extensively used code; it has been tested over a wide range of cosmological parameters and is considered to be accurate to better than the 1% level[47].

A description of the physics underlying the C_ls can be separated into 3 main regions, as shown in

V-1 The Sachs-Wolfe plateau: $l \lesssim 100$

The horizon scale (or more precisely, the angle subtended by the Hubble radius) at last scattering corresponds to $\ell \simeq 100$. Anisotropies at larger scales have not evolved significantly, and hence directly reflect the 'initial conditions.' The combination of gravitational redshift and intrinsic temperature fluctuations leads to $\delta T/T \simeq (1/3)\delta\phi/c2$, where $\delta\phi$ is the perturbation to the gravitational potential. This is usually referred to as the 'Sachs-Wolfe' effect[26].

Assuming that a nearly scale-invariant spectrum of density perturbations was laid down at early times (*i.e.* , $n \simeq 1$, meaning equal power per decade in k), then $\ell(\ell+1)C_\ell \simeq$ constant at low ℓs. This effect is hard to see unless the multipole axis is plotted logarithmically (as in

Time variation in the potentials (*i.e.* , time-dependent metric perturbations) leads to an upturn in the C_ls in the lowest several multipoles; any deviation from a total equation of state $w = 0$ has such an effect. So the dominance of the Dark Energy at low redshift makes the lowest ℓs rise above the plateau. This is sometimes called the 'integrated Sachs-Wolfe effect' (or ISW Rise), since it comes from the line integral of $\dot{\phi}$. It has been confirmed through correlations between the large-angle anisotropies and large-scale structure[48]. Specific models can also give additional contributions at low l (e.g., perturbations in the Dark Energy component itself[49]) but typically these are buried in the cosmic variance.

In principle, the mechanism that produces primordial perturbations would generate scalar, vector, and tensor modes. However, the vector (vorticity) modes decay with the expansion of the Universe. Tensors also decay when they enter the horizon, and so they contribute only to angular scales above about 1° (see Hence some fraction of the low ℓ signal could be due to a gravity wave contribution, although small amounts of tensors are essentially impossible to discriminate from other effects that might raise the level of the plateau. However the tensors *can* be distinguished using polarization information (section 6).

V-2 The acoustic peaks: $100 \lesssim l \lesssim 1000$

On sub-degree scales, the rich structure in the anisotropy spectrum is the consequence of gravity-driven acoustic oscillations occurring before the atoms in the universe became neutral. Perturbations inside the horizon at last scattering have been able to evolve causally and produce anisotropy at the last scattering epoch which reflects that evolution. The frozen-in phases of these sound waves imprint a dependence on the cosmological parameters, which gives CMB anisotropies their great constraining power.

The underlying physics can be understood as follows. When the proton-electron plasma was tightly coupled to the photons, these components behaved as a single 'photon-baryon fluid', with the photons providing most of the pressure and the baryons the inertia. Perturbations in the gravitational potential, dominated by the dark matter component, are steadily evolving. They drive oscillations in the photon-baryon fluid, with photon pressure providing the restoring force. The perturbations are quite small, $O(10^{-5})$, and so evolve linearly. That means each Fourier mode evolves independently and is described by a driven harmonic oscillator, with frequency determined by the sound speed in the fluid. Thus, there is an oscillation of the fluid density, with velocity $\pi/2$ out of phase and having amplitude reduced by the sound speed.

After the Universe recombined the baryons and radiation decoupled, and the radiation could travel freely towards us. At that point the phases of the oscillations were frozen-in, and projected on the sky as a harmonic series of peaks. The main peak is the mode that went through 1/4 of a period, reaching maximal compression. The even peaks are maximal *under*-densities, which are generally of smaller amplitude because the rebound has to fight against the baryon inertia. The troughs, which do not extend to zero power, are partially filled because they are at the velocity maxima.

An additional effect comes from geometrical projection. The scale associated with the peaks is the sound horizon at last scattering, which can be confidently calculated as a physical length scale. This scale is projected onto the sky, leading to an angular scale that depends on the background cosmology. Hence the angular position of the peaks is a sensitive probe of the spatial curvature of the Universe (*i.e.* , $\Omega_{\rm tot}$), $\ell_{peak} \sim 220/\sqrt{\Omega_{tot}}$ with the peaks lying at higher ℓ in open universes and lower ℓ in closed geometry. A the current level of CMB observation precision one must take into account second order effects which could in principle shift the location of the acoustic peaks by a few per cent. The most important of these include the relative amount dark energy and dark matter.

One last effect arises from reionization at redshift $z_{\rm i}$. A fraction of photons will be isotropically scattered at $z < z_{\rm i}$, partially erasing the anisotropies at angular scales smaller than those subtended by the Hubble radius at $z_{\rm i}$. This corresponds typically to ℓs above about a few 10s, depending on the specific reionization model. The acoustic peaks are therefore reduced by a factor $e^{-2\tau}$ relative to the plateau.

These acoustic peaks were a clear theoretical prediction going back to about 1970[51]. Their empirical existence started to become clear around 1994[31], and the emergence, over the following decade, of a coherent series of acoustic peaks and troughs is a triumph of modern cosmology. One can think of these peaks as a snapshot of stochastic standing waves. And, since the physics governing them is simple, then one can see how they encode information about the cosmological parameters.

V-3 The damping tail: $l \gtrsim 1000$

The recombination process is not instantaneous, giving a thickness to the last scattering surface. This leads to a damping of the anisotropies at the highest ℓs, corresponding to scales smaller than that subtended by this thickness. One can also think of the photon-baryon fluid as having imperfect coupling, so that there is diffusion between the two components, and the oscillations have amplitudes that decrease with time. These effects lead to a damping of the C_ls, sometimes called Silk damping[28], which cuts off the anisotropies at multipoles above about 2000.

An extra effect at high ℓs comes from gravitational lensing, caused mainly by non-linear structures at low redshift. The C_ls are convolved with a smoothing function in a calculable way, partially flattening the peaks, generating a power-law tail at the highest multipoles, and complicating the polarization signal[54]. This is an example of a 'secondary

effect', *i.e.* , the processing of anisotropies due to relatively nearby structures. Galaxies and clusters of galaxies give several such effects, but all are expected to be of low amplitude and are typically only important for the highest ℓs.

VI. Current Anisotropy Data

There has been a steady improvement in the quality of CMB data that has led to the development of the present-day cosmological model. Probably the most robust constraints currently available come from the combination of the *WMAP* first year data[17] with smaller scale results from the CBI[55] and ACBAR[41] experiments. We plot these power spectrum estimates in Other recent experiments, such as ARCHEOPS [56], BOOMERANG[57], DASI[42], MAXIMA[58] and VSA[52] also give powerful constraints, which are quite consistent with what we describe below. There have been some comparisons among data-sets[59], which indicate very good agreement, both in maps and in derived power spectra (up to systematic uncertainties in the overall calibration for some experiments). This makes it clear that systematic effects are largely under control. However, a fully self-consistent joint analysis of all the current data sets has not been attempted, one of the reasons being that it requires a careful treatment of the overlapping sky coverage.

Figure 6 shows band-powers from the first year *WMAP* data[80], together with CBI and ACBAR data at higher ℓ. The points are in very good agreement with a 'ΛCDM' type model, as described in the previous section, with several of the peaks and troughs quite apparent. For details of how these estimates were arrived at, the strength of any correlations between band-powers and other information required to properly interpret them, turn to the original papers[17][55][41].

VII. CMB Polarization

In Penzias and Wilson's first paper the CMB was shown to be unpolarized to the 10% level. Polarization of the CMB was shown to be < 300 μK ([60]). COBE put a limit of < 15 μK on the polarization anisotropy. The linear polarization of the CMB was first detected by DASIPOL ([61]), and the cross-correlation of the temperature and polarization anisotropies was confirmed by *WMAP* ([62]).

The detected polarization level is an order of magnitude lower than the anisotropy. The observed polarization is caused by electron scattering during the late stages of recombination on small angular scales and after reionization on large angular scales. The magnitude of the polarization on small angular scales depends on the anisotropy being in place at

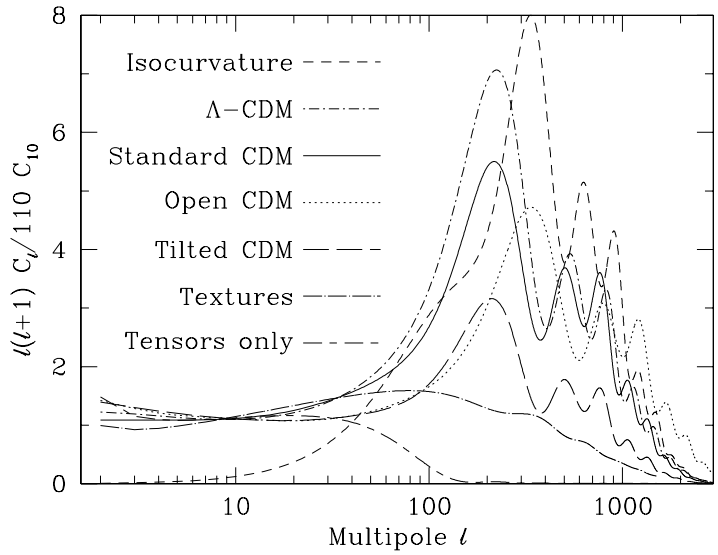

Figure 6. Theoretical Curves for C_ℓ vs ℓ for a number of cosmological models.

recombination, as is the case for primordial adiabatic perturbations but not for topological defects; the electron scattering cross-section; and the recombination coefficient of hydrogen. The detection of this polarization is a very strong confirmation of the standard model for CMB anisotropy.

Because polarization is a vector field, two distinct modes or patterns can arise ([63], [64]): the gradient of a scalar field (the "E" mode) or the curl of a vector field (the "B" mode). Electron scattering only produces the E mode. Electron scattering gives a polarization pattern that is correlated with the temperature anisotropy, so the E modes can be detected by cross-correlating the polarization with the temperature. The B modes cannot be detected this way, and the predicted level of the B modes is at least another order of magnitude below the E modes, or two orders of magnitude below the temperature anisotropy.

Since Thomson scattering of an anisotropic radiation field also generates linear polarization, the CMB is predicted to be polarized at the roughly 5% level[65]. Polarization is a spin 2 field on the sky, and the

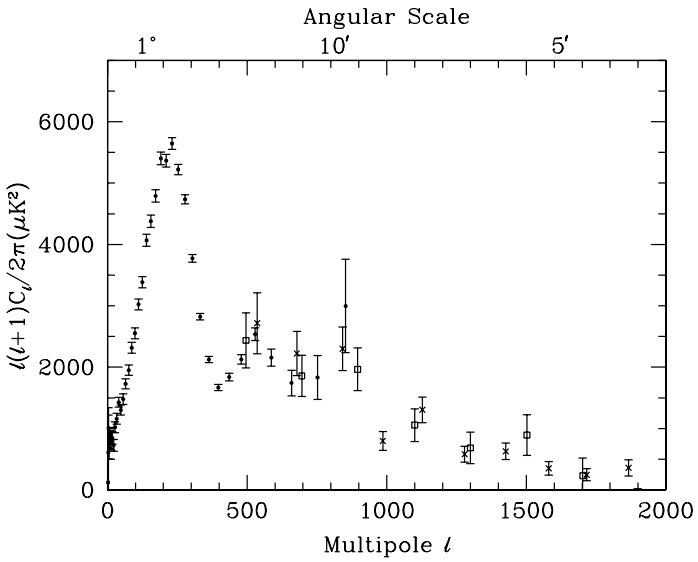

Figure 7. Band-power estimates from the *WMAP*, CBI, and ACBAR experiments. The *WMAP* data are the points, while squares are CBI and crosses ACBAR. We have shown only CBI and ACBAR data relevant for $l > 500$, and both experiments also probe to higher ℓ than shown. This plot represents only a fraction of experimental results, with several other data-sets being of similar quality. The multipole axis here is linear, so the Sachs-Wolfe plateau is hard to see. The acoustic peaks and damping region are very clearly observed, with no need for a theoretical curve to guide the eye.

algebra of the modes in ℓ-space is strongly analogous to spin-orbit coupling in quantum mechanics[66]. The linear polarization pattern can be decomposed in a number of ways, with two quantities required for each pixel in a map, often given as the Q and U Stokes parameters. However, the most intuitive and physical decomposition is a geometrical one, splitting the polarization pattern into a part that comes from a divergence (often referred to as the 'E-mode') and a part with a curl (called the 'B-mode')[67]. More explicitly, the modes are defined in terms of second derivatives of the polarization amplitude, with the Hessian for the E-modes having principle axes in the same sense as the polarization, while the B-mode pattern can be thought of simply as a 45° rotation

of the E-mode pattern. Globally one sees that the E-modes have $(-1)^{\ell}$ parity (like the spherical harmonics), while the B-modes have $(-1)^{\ell+1}$ parity.

The existence of this linear polarization allows for 6 different cross power spectra to be determined from data that measure the full temperature and polarization anisotropy information. Parity considerations make 2 of these zero, and we are left with 4 potential observables: $C_\ell^{\rm TT}$, $C_\ell^{\rm TE}$, $C_\ell^{\rm EE}$, and $C_\ell^{\rm BB}$. Since scalar perturbations have no handedness, the B-mode power spectrum can only be generated by vectors or tensors. Hence, in the context of inflationary models, the determination of a non-zero B-mode signal is a way to measure the gravity wave contribution (and thus potentially derived the energy scale of inflation), even if it is rather weak. However, one must first eliminate the foreground contributions and other systematic effects down to very low levels.

The oscillating photon-baryon fluid also results in a series of acoustic peaks in the polarization power spectra. The main 'EE' power spectrum has peaks that are out of phase with those in the 'TT' spectrum, because the polarization anisotropies are sourced by the fluid velocity. The correlated component of the polarization and temperature patterns comes from correlations between density and velocity perturbations on the last scattering surface, which can be both positive and negative. There is no polarization 'Sachs-Wolfe' effect, and hence no large-angle plateau. However, scattering during a recent period of reionization can create a polarization 'bump' at large angular scales.

The strongest upper limits on polarization are at the roughly $10\,\mu$K level from the POLAR[69] experiment at large angular scales and the PIQUE[70] and COMPASS[71] experiments at smaller scales. The first measurement of a polarization signal came in 2002 from the DASI experiment[61], which provided a convincing detection, confirming the general paradigm, but of low enough significance that it lends little constraint to models. As well as the E-mode signal, DASI also made a statistical detection of the TE correlation.

More recently the *WMAP* experiment was able to measure the TE cross-correlation power spectrum with high precision[62]. The results are shown in Figure 8. along with some estimates from the DASI experiment. The detected shape of the cross-correlation power spectrum provides supporting evidence of the adiabatic nature of the perturbations, as well as directly constraining the thickness of the last scattering surface. Since the polarization anisotropies are generated in this scattering surface, the existence of correlations at angles above about a degree demonstrate that there were super-Hubble fluctuations at the recombination epoch.

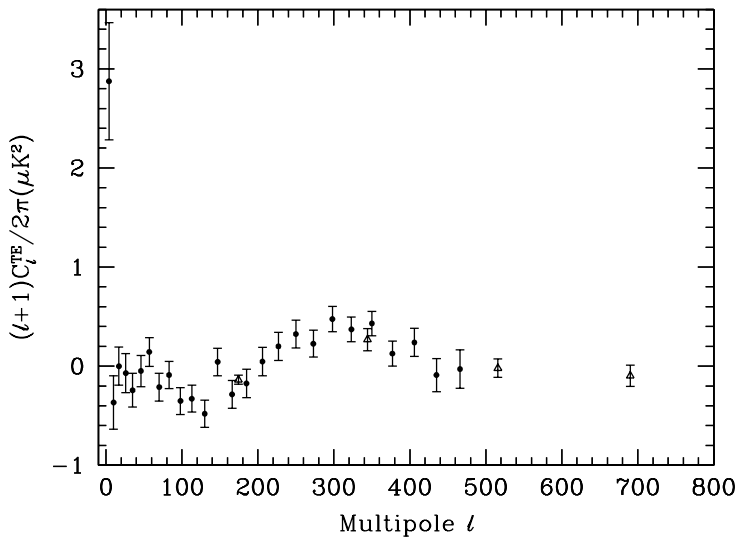

Figure 8. Cross power spectrum of the temperature anisotropies and E-mode polarization signal from *WMAP* (points), together with some estimates from DASI (triangles) which extend to higher ℓ. Note that the DASI bands are much wider in ℓ than those of *WMAP*. Also note that the y-axis is not multiplied by the additional ℓ, which helps to show both the large and small angular scale features.

Perhaps the most intriguing result from the polarization measurements is at the largest angular scales ($l < 10$), where there is an excess signal compared to that expected from the temperature power spectrum alone. This is precisely the signal expected from an early period of reionization, arising from Doppler shifts during the partial scattering at $z < z_{\rm i}$. It seems to indicate that the first stars (presumably the source of the ionizing radiation) formed around $z = 20$.

VIII. Complications

There are a number of issues which complicate the interpretation of CMB anisotropy data, some of which we sketch out below.

VIII-1 Foregrounds

The microwave sky contains significant emission from our Galaxy and from extragalactic sources. Fortunately, the frequency dependence of these various sources are in general substantially different than the CMB anisotropy signals. The combination of Galactic synchrotron, bremsstrahlung and dust emission reaches a minimum at a wavelength of roughly 3 mm (or about 100 GHz). As one moves to greater angular resolution, the minimum moves to slightly higher frequencies, but becomes more sensitive to unresolved (point-like) sources.

At frequencies around 100 GHz and for portions of the sky away from the Galactic Plane the foregrounds are typically 1 to 10% of the CMB anisotropies. By making observations at multiple frequencies, it is relatively straightforward to separate the various components and determine the CMB signal to the few per cent level. For greater sensitivity it is necessary to improve the separation techniques by adding spatial information and statistical properties of the foregrounds compared to the CMB.

The foregrounds for CMB polarization are expected to follow a similar pattern, but are less well studied, and are intrinsically more complicated. Whether it is possible to achieve sufficient separation to detect B-mode CMB polarization is still an open question. However, for the time being, foreground contamination is not a major issue for CMB experiments.

VIII-2 Secondary Anisotropies

With increasingly precise measurements of the primary anisotropies, there is growing theoretical and experimental interest in 'secondary anisotropies.' Effects which happen at $z \ll 1000$ become more important as experiments push to higher angular resolution and sensitivity.

These secondary effects include gravitational lensing, patchy reionization and the Sunyaev-Zel'dovich (SZ) effect[72]. This is Compton scattering ($\gamma e \rightarrow \gamma' e'$) of the CMB photons by a hot electron gas, which creates spectral distortions by transferring energy from the electrons to the photons. The effect is particularly important for clusters of galaxies, through which one observes a partially Comptonized spectrum, resulting in a decrement at radio wavelengths and an increment in the submillimeter. This can be used to find and study individual clusters and to obtain estimates of the Hubble constant. There is also the potential to constrain the equation of state of the Dark Energy through counts of clusters as a function of redshift[73].

VIII-3 Higher-order Statistics

Although most of the CMB anisotropy information is contained in the power spectra, there will also be weak signals present in higher-order statistics. These statistics will measure primordial non-Gaussianity in the perturbations, as well as non-linear growth of the fluctuations on small scales and other secondary effects (plus residual foreground contamination). Although there are an infinite variety of ways in which the CMB could be non-Gaussian, there is a generic form to consider for the initial conditions, where a quadratic contribution to the curvature perturbations is parameterized through a dimensionless number $f_{\rm NL}$. This weakly non-linear component can be constrained through measurements of the bispectrum or Minkowski functionals for example, and the result from *WMAP* is $-58 < f_{\rm NL} < 134$ (95% confidence region)[44].

IX. Constraints on Cosmologies

The most important outcome of the newer experimental results is that the standard cosmological paradigm is in good shape. A large amount of high precision data on the power spectrum is adequately fit with fewer than 10 free parameters. The framework is that of Friedmann-Robertson-Walker models, which have nearly flat geometry, containing Dark Matter and Dark Energy, and with adiabatic perturbations having close to scale invariant initial conditions.

Within this framework, bounds can be placed on the values of the cosmological parameters. Of course, much more stringent constraints can be placed on models which cover a restricted number of parameters, e.g. assuming that $\Omega_{\rm tot} = 1$, $n = 1$ or $r = 0$. More generally, the constraints depend upon the adopted priors, even if they are implicit, for example by restricting the parameter freedom or the ranges of parameters (particularly where likelihoods peak near the boundaries), or by using different choices of other data in combination with the CMB. When the data become even more precise, these considerations will become less important, but for now we caution that restrictions on model space and choice of priors need to be kept in mind when adopting specific parameter values and uncertainties.

There are some combinations of parameters that fit the CMB anisotropies almost equivalently. For example, there is a nearly exact geometric degeneracy, where any combination of $\Omega_{\rm M}$ and Ω_Λ that gives the same angular diameter distance to last scattering will give nearly identical C_ℓs. There are also other near degeneracies among the parameters. Such degeneracies can be broken when using the CMB data in combination with other cosmological data sets. Particularly useful are complementary

constraints from galaxy clustering, the abundance of galaxy clusters, weak gravitational lensing measurements, Type Ia supernova distances and the distribution of Lyman α forest clouds.

The combination of *WMAP*, CBI and ACBAR, together with weak priors (on h and $\Omega_{\rm B}h2$ for example), and within the context of a 6 parameter family of models (which fixes $\Omega_{\rm tot} = 1$), yields the following results[74]: $A = 2.7(\pm 0.3) \times 10^{-9}$, $n = 0.97 \pm 0.03$, $h = 0.73 \pm 0.05$, $\Omega_{\rm B}h2 = 0.023 \pm 0.001$, $\Omega_{\rm M}h2 = 0.13 \pm 0.01$ and $\tau = 0.17 \pm 0.07$. Note that for h, the CMB data alone provide only a very weak constraint, unless spatial flatness or some other cosmological data are used. For $\Omega_{\rm B}h2$ the precise value depends sensitively on how much freedom is allowed in the shape of the primordial power spectrum. The best constraint on $\Omega_{\rm tot}$ is 1.02 ± 0.02. This comes from including priors from h and supernova data. Slightly different, but consistent results come from using different data combinations.

The 95% confidence upper limit on r is 0.53 (including some extra constraint from galaxy clustering). This limit is stronger if we restrict ourselves to $n < 1$ and weaker if we allow $dn/d\ln k \neq 0$.

There are also constraints on parameters over and above the basic 8 that we have described. But for such constraints it is necessary to include additional data in order to break the degeneracies. For example the addition of the Dark Energy equation of state, w adds the partial degeneracy of being able to fit a ridge in (w, h) space, extending to low values of both parameters. This degeneracy is broken when the CMB is used in combination with independent H_0 limits, for example [75], giving $w < -0.5$ at 95% confidence. Tighter limits can be placed using restrictive model-spaces and/or additional data.

For the optical depth τ, the error bar is large enough that apparently quite different results can come from other combinations of data. The constraint from the combined *WMAP* $C_\ell^{\rm TT}$ and $C_\ell^{\rm TE}$ data is $\tau = 0.17 \pm 0.04$, which corresponds (within reasonable models) to a reionization redshift $9 < z_{\rm i} < 30$ (95% CL)[62]. This is a little higher than some theoretical predictions and some suggestions from studies of absorption in high-z quasar spectra[76]. The excitement here is that we have direct information from CMB polarization which can be combined with other astrophysical measurements to understand when the first stars formed and brought about the end of the cosmic dark ages.

X. Particle Physics Constraints

CMB data are beginning to put limits on parameters which are directly relevant for particle physics models. For example there is a limit

on the neutrino contribution $\Omega_\nu h2 < 0.0076$ (95% confidence) from a combination of *WMAP* and galaxy clustering data from the 2dFGRS project[77]. This directly implies a limit on neutrino mass, assuming the usual number density of fermions which decoupled when they were relativistic.

A combination of the *WMAP* data with other data-sets gives some hint of a running spectral index, *i.e.* , $dn/d\ln k \neq 0$[74]. Although this is still far from resolved[78], things will certainly improve as new data come in. A convincing measurement of a non-zero running of the index would be quite constraining for inflationary models [79].

One other hint of new physics lies in the fact that the quadrupole and some of the other low ℓ modes seem anomalously low compared with the best-fit ΛCDM model[80]. This is what might be expected in a universe which has a large scale cut-off to the power spectrum, or is topologically non-trivial. However, because of cosmic variance, possible foregrounds *etc.*, the significance of this feature is still a matter of debate[81].

In addition it is also possible to put limits on other pieces of physics [82], for example the neutrino chemical potentials, time variation of the fine-structure constant, or physics beyond general relativity. Further particle physics constraints will follow as the anisotropy measurements increase in precision.

Careful measurement of the CMB power spectra and non-Gaussianity can in principle put constraints on high energy physics, including ideas of string theory, extra dimensions, colliding branes, *etc.* At the moment any calculation of predictions appears to be far from definitive. However, there is a great deal of activity on implications of string theory for the early Universe, and hence a very real chance that there might be observational implications for specific scenarios.

XI. Fundamental Lessons

More important than the precise values of parameters is what we have learned about the general features which describe our observable Universe. Beyond the basic hot Big Bang picture, the CMB has taught us that:

- The Universe recombined at $z \simeq 1100$ and started to become ionized again at $z \simeq 10$–30.
- The geometry of the Universe is close to flat.
- Both Dark Matter and Dark Energy are required.
- Gravitational instability is sufficient to grow all of the observed large structures in the Universe.

- Topological defects were not important for structure formation.
- There are 'synchronized' super-Hubble modes generated in the early Universe.
- The initial perturbations were adiabatic in nature.
- The perturbations had close to Gaussian (*i.e.* , maximally random) initial conditions.

It is very tempting to make an analogy between the status of the cosmological 'Standard Model' and that of particle physics. In cosmology there are about 10 free parameters, each of which is becoming well determined, and with a great deal of consistency between different measurements. However, none of these parameters can be calculated from a fundamental theory, and so hints of the bigger picture, 'physics beyond the Standard Model' are being searched for with ever more challenging experiments.

Despite this analogy, there are some basic differences. For one thing, many of the cosmological parameters change with cosmic epoch, and so the measured values are simply the ones determined today, and hence they are not 'constants', like particle masses for example (although they *are* deterministic, so that if one knows their values at one epoch, they can be calculated at another). Moreover, the number of parameters is not as fixed as it is in the particle physics Standard Model; different researchers will not necessarily agree on what the free parameters are, and new ones can be added as the quality of the data improves. In addition parameters like τ, which come from astrophysics, are in principle calculable from known physical processes, although this is currently impractical. On top of all this, other parameters might be 'stochastic' in that they may be fixed only in our observable patch of the Universe.

In a more general sense the cosmological 'Standard Model' is much further from the underlying 'fundamental theory' which will provide the values of the parameters from first principles. On the other hand, any genuinely complete 'theory of everything' must include an explanation for the values of these cosmological parameters as well as the parameters of the Standard Model.

XII. Future Directions

With all the observational progress in the CMB and the tying down of cosmological parameters, what can we anticipate for the future? Of course there will be a steady improvement in the precision and confidence with which we can determine the appropriate cosmological model and

its parameters. We can anticipate that the evolution from one year to four years of *WMAP* data will bring improvements from the increased statistical accuracy and from the more detailed treatment of calibration and systematic effects. Ground-based experiments operating at the smaller angular scales will also improve over the next few years, providing significantly tighter constraints on the damping tail. In addition, the next CMB satellite mission, *Planck*, is scheduled for launch in 2007, and there are even more ambitious projects currently being discussed.

Despite the increasing improvement in the results, it is also true that the addition of the latest experiments has not significantly changed the cosmological model (apart from a suggestion of higher reionization redshift perhaps). It is therefore appropriate to ask: what should we expect to come from *Planck* and from other more grandiose future experiments, including the proposed *Inflation Probe* or *CMBPol*? *Planck* certainly has the the advantage of high sensitivity and a full sky survey. A detailed measurement of the third acoustic peak provides a good determination of the matter density; this can only be done by measurements which are accurate relative to the first two peaks (which themselves constrained the curvature and the baryon density). A detailed measurement of the damping tail region will also significantly improve the determination of n and any running of the slope. *Planck* should also be capable of measuring $C_\ell^{\rm EE}$ quite well, providing both a strong check on the Standard Model and extra constraints that will improve parameter estimation.

A set of cosmological parameters are now known to roughly 10% accuracy, and that may seem sufficient for many people. However, we should certainly demand more of measurements which describe *the entire observable Universe*! Hence a lot of activity in the coming years will continue to focus on determining those parameters with increasing precision. This necessarily includes testing for consistency among different predictions of the Standard Model, and searching for signals which might require additional physics.

A second area of focus will be the smaller scale anisotropies and 'secondary effects.' There is a great deal of information about structure formation at $z \ll 1000$ encoded in the CMB sky. This may involve higher-order statistics as well as spectral signatures. Such investigations can also provide constraints on the Dark Energy equation of state, for example. *Planck*, as well as experiments aimed at the highest ℓs, should be able to make a lot of progress in this arena.

A third direction is increasingly sensitive searches for specific signatures of physics at the highest energies. The most promising of these may be the primordial gravitational wave signals in $C_\ell^{\rm BB}$, which could be a probe of the $\sim 10^{16}$ GeV energy range. Whether the amplitude

of the effect coming from inflation will be detectable is unclear, but the prize makes the effort worthwhile.

Anisotropies in the CMB have proven to be the premier probe of cosmology and the early Universe. Theoretically the CMB involves well-understood physics in the linear regime, and is under very good calculational control. A substantial and improving set of observational data now exists. Systematics appear to be well understood and not a limiting factor. And so for the next few years we can expect an increasing amount of cosmological information to be gleaned from CMB anisotropies, with the prospect also of some genuine surprises.

References

[1] Penzias A.A., R. Wilson R. Astrophys. J.**142** (1965) 419.

[2] Boggess N. W., *et al.*, Astrophys. J.**397** (1992) 420.

[3] Fixsen, D. J., Cheng, E. S., Gales, J. M., Mather, J. C., Shafer, R. A., and Wright, E. L. Astrophys. J.**47** (1996) 576.

[4] Mather J.C.,et al. , Astrophys. J.**512** (1999) 511.

[5] Alpher R.A., Herman R.C., Physics Today **41**, No. 8, (1988) 24.

[6] Peebles P.J.E., "Principles of Physical Cosmology," Princeton U. Press (1993) 168.

[7] Roth K.C., Bauer J.M., Astrophys. J.**515** (1999) L57.

[8] Birkinshaw M., Phys. Rep., **310** (1999) 98.

[9] da Silva A.C., Barbosa D., Liddle A.R., Thomas P.A., Monthly Not. Royal Astron. Soc., **317** (2000) 37. astro-ph/9907224

[10] Danese L., De Zotti G.F., Astr. & Ap.**107** (1982) 39.

[11] Bartlett J.G., Stebbins A. Astrophys. J., **371** (1991) 8.

[12] Gunn, J. E., Peterson B. E. Astrophys. J.**142** (1965) 1633.

[13] D.J. Fixsen et al. , Astrophys. J.**473** (1996) 576.

[14] M. Bersanelli et al. , Astrophys. J.**424** (1994) 517.

[15] M. White, D. Scott, and J. Silk, Ann. Rev. Astron. & Astrophys. **32** (1994) 329.

[16] G. F. Smoot et al. Astrophys. J.**396** (1992) L1.

[17] C.L. Bennett et al. , Astrophys. J., Supp.**148** (2003) 1. (astro-ph/0302207).

[18] S. Courteau, J.A. Willick, M.A. Strauss, D. Schlegel, and M. Postman, Astrophys. J.**544** (2000) 636.

[19] D.J. Fixsen et al. , Astrophys. J.**420** (1994) 445.

[20] Conklin, E. K. Nature**222** (1969) 971.

[21] Henry, P. S. Nature**231** (1971) 516.

[22] Corey, B. E. and Wilkinson, D. T. Bull. Am. Astr. Soc.**8** (1976) 351.

[23] Smoot, G. F., Goernstein, M. V., and Muller, R. A. Phys. Rev. Lett.**39** (1977) 898.

[24] C.L. Bennett et al. , Astrophys. J.**464** (1996) L1. astro-ph/9601067

[25] Bennett C. L., *et al.*, Astrophys. J., Supp.**148** (2003) 1.

[26] Sachs, R. K. and Wolfe, A. M. Astrophys. J.**147** (1967) 73.

[27] Silk, J. Astrophys. J.**151** (1968) 459.

[28] P.J.E. Peebles Astrophys. J., Lett.**263** (1982) L1.

[29] Bond, J. R. and Efstathiou, G. Mon. Not. R. A. S.**226** (1987) 655.

[30] G. Jungman, G., Kamionkowski, M., Kosowsky, A., and Spergel, D. N. Phys. Rev. Lett.**76** (1996) 1007.

[31] D. Scott, J. Silk, and M. White, Science, **268** (1995) 829.

[32] L. Knox and L. Page Phys. Rev. Lett.**85** (2000) 1366.

[33] P. de Bernardis, *et al.*, Nature**404** (2000) 955.

[34] S. Hanany,*et al.*, Astrophys. J., Supp.**545** (2000) L5.

[35] J.E. Ruhl et al. , Astrophysical J., in press, astro-ph/0212229

[36] P. de Bernardis et al. , Symposium **216 :** MAPS of the Cosmos, Sydney, July 2003, APS Conference Series

[37] Lee A.T., *et al.*, Astrophys. J.**561** (2001) L1.

[38] J. Kovac et al. , Nature **420** (2002) 772.

[39] S. Perlmutter et al. , Astrophys. J.**517** (1999) 565.

[40] L. Page,*et al.*, Astrophys. J., Supp.**148** (2003) 233.

[41] C.L. Kuo et al. , Astrophysical J., in press, astro-ph/0212289

[42] E.M. Leitch et al. , Astrophys. J.**568** (2002) 28.

[43] S. Seager, D.D. Sasselov, and D. Scott, Astrophys. J., Supp.**128** (2000) 207.

[44] E. Komatsu et al. , Astrophys. J., Supp., **148** (2003) 119. (astro-ph/0302223)

[45] U. Seljak and M. Zaldarriaga, Astrophys. J.**469** (1996) 437.

[46] U. Seljak, Astrophys. J., Supp.**129** (2000) 431.

[47] U. Seljak, N. Sugiyama, M. White, M. Zaldarriaga, Physical Review D, in press, astro-ph/0306052

[48] M.R. Nolta et al. , Astrophysical J., in press, astro-ph/0305097

[49] W. Hu and D. J. Eisenstein, prd **59** (1999) 083509.

[50] W. Hu, D.J. Eisenstein, M. Tegmark, M. White, Phys. Rev. D**59** (1999) 023512.

[51] P.J.E. Peebles and J.T. Yu, Astrophys. J.**162** (1970) 815.

[52] P.F. Scott et al. , Monthly Not. Royal Astron. Soc., **341** (2003) 1076.

[53] J. Silk, Astrophys. J.**151** (1968) 459.

[54] M. Zaldarriaga and U. Seljak, Phys. Rev. D**58** (1998) 023003.

[55] T.J. Pearson et al. , Astrophys. J.**591** (2003) 556.

[56] A. Benoit et al. , A&A, **399** (2003) L19.

[57] J.E. Ruhl et al. , Astrophysical J., in press, astro-ph/0212229

[58] A.T. Lee et al. , Astrophys. J., **561** (2001) L1.

[59] M.E. Abroe et al. , Astrophysical J., in press, astro-ph/0308355

[60] Lubin, P. M. and Smoot, G. F. (1981). Astrophys. J.**245** (1981) 1-17.

[61] J. Kovac et al. , Nature, **420** (2002) 772.

[62] A. Kogut et al. , Astrophys. J., Supp.**148** (2003) 161. astro-ph/0302213

[63] Kamionkowski, M., Kosowsky, A., and Stebbins, A. Phys. Rev. D**55** (1997) 7368.

[64] Seljak, U. and Zaldarriaga, M. Phys. Rev. Lett.**78** (1997) 2054.

[65] W. Hu, M. White, New Astron. **2** (1997) 323.

[66] W. Hu, M. White, Phys. Rev. D**56** (1997) 596.

[67] M. Zaldarriaga and U. Seljak, Phys. Rev. D**55** (1997) 1830.

[68] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. D**55** (1997) 7368.

[69] B.G. Keating et al. , Astrophysical J., in press, astro-ph/0107013

[70] M.M. Hedman et al. , Astrophys. J.**548** (2001) L111.

[71] P.C. Farese et al. , Astrophysical J., in press, astro-ph/0308309

[72] Sunyaev R.A., Zel'dovich Ya.B., Ann. Rev. of Astr. & Ap., **18** (1980) 537.

[73] J.E. Carlstrom, G.P. Holder, and E.D. Reese, Ann. Rev. Astron. & Astrophys. **40** (2002) 643.

[74] D.N. Spergel et al. , Astrophys. J., Supp.**148**(2003) 175. astro-ph/0302209

[75] W.L. Freedman et al. , Astrophys. J.**553** (2001) 47.

[76] X. Fan et al. , Astrophys. J.**123**(2002) 1247.

[77] M. Colless et al. , Monthly Not. Royal Astron. Soc. **328** (2001) 1039.

[78] U. Seljak, P. McDonald, and A. Makarov, Monthly Not. Royal Astron. Soc. **342** (2003) L79. astro-ph/0302571

[79] H.V. Peiris et al. , Astrophys. J., Supp.**148** (2003) 213. astro-ph/0302225

[80] G. Hinshaw et al. , Astrophys. J., Supp.**148** (2003) 135. astro-ph/0302217

[81] A. de Oliveira-Costa et al. , Phys. Rev. D, in press astro-ph/0307282

[82] M. Kamionkowski, A. Korowsky, Ann. Rev. Nucl. Part. Science, **49** (1999) 77.

COSMOLOGY WITH SUPERNOVAE 1A

Smadja Gérard
IPNL,Université Cl. Bernard Lyon 1
F69622 Villeurbanne Cedex
g.smadja@ipnl.in2p3.fr

Abstract The basic relations between the luminosity distance, the matter density Ω_M, and the cosmological constant Ω_Λ are derived. The universal character of the luminosity of SN1a is described, and the experimental status of the determinations of the Hubble constant, Ω_M, and Ω_Λ are recalled. The large supernovae surveys foreseen in the near future, and their expected performances are reviewed. Some aspects of the supernovae explosions are briefly summarised

Keywords: Supernovae, Friedmann's equation, luminosity distance, Hubble expansion, cosmological constant, matter density, dust, evolution

Introduction

We shall assume that the mean luminosity of Type Ia supernovae does not change with redshift: they are standard. There is no indication at present for any such variation, although the bounds are weak. Should future data contradict this assumption at some level, the present values of the densities Ω_M, Ω_Λ should be appropriately corrected, but the analysis discussed here would not be drastically altered.

The observed luminosity of supernovae is in this context a direct indicator of their distance (using a modified inverse squared law), and the redshift of the host galaxy lines gives its recession velocity. The relation between distance and redshift reflects the history of the Hubble 'constant' between the time of emission of the SN light and its observation today. In a homogeneous and isotropic universe, this variation with time (or redshift) can be explicitly obtained from the Friedman equation which we shall recall. Since the first quantitative results were published

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 311–327.

by [6] and [7], the indications for a finite cosmological constant have been strengthening.

I. Distances in curved space

I-1 2D Geometry

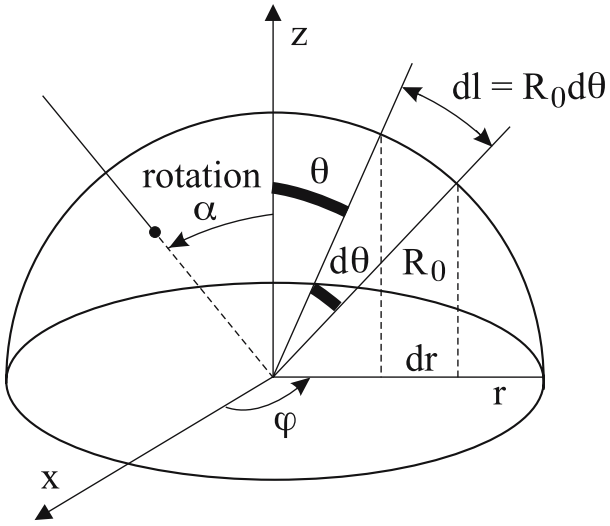

Figure 1. Distances on the 2D surface of a sphere

The 2D surface of a sphere of radius R_0 gives an intuitive introduction to the relation between the projected cordinates (r, ϕ) and distances. The geodesic line element is

$$dl = \frac{dr}{\sin\theta} = \frac{dr}{\sqrt{1 - r^2/R_0^2}}$$

The translation operation does not preserve the surface, and must be replaced by a rotation. This will be directly transposed to the 3D surface of a 4D sphere, which might describe the universe at large scales (if curved).

I-2 3D curved space : changing the origin

The homogeneity of space at large scales imposes a constant radius.

$$x_1^2 + x_2^2 + x_3^2 + kx_4^2 = kR_0^2$$

with $k = -1, 0, 1$ depending on the sign of the curvature (the constant R_0 is related to but not equal to the 3D curvature). Let us stress that the auxiliary 4th coordinate is NOT time. A change of origin cannot be described by a translation, but rather by 3D or 4D 'rotations' which leave the line element

$$dl^2 = dx_1^2 + dx_2^2 + dx_3^2 + kdx_4^2 \tag{I-2.1}$$

invariant. The usual spherical coordinates can be introduced if $k > 0$.

$$\begin{aligned} x_1 &= R_0 \sin\chi \sin\theta \cos\phi \\ x_2 &= R_0 \sin\chi \sin\theta \sin\phi \\ x_3 &= R_0 \sin\chi \cos\theta \\ x_4 &= R_0 \cos\chi \end{aligned}$$

The line element can be expressed as

$$dl^2 = a(t)^2 R_0^2 \left[r_a(d\theta^2 + \sin^2\theta (d\phi)^2) + \frac{(dr_a)^2}{1 - kr_a^2} \right] \tag{I-2.2}$$

with

$$r_a = \frac{r}{R_0} = \sin\chi \tag{I-2.3}$$

The scale factor $a(t)R_0$ converts (comoving) dimensionless coordinates to distances at time t. A change of origin should be described by a rotation involving the 4th auxiliary coordinate. Assume $k = 1$

$$\begin{aligned} x_4^* &= x_4 \cos\chi + x_3 \sin\chi = R_0 \cos\chi^* \\ x_3^* &= x_3 \cos\chi - x_4 \sin\chi = R_0 \sin\chi^* \cos\theta^* \\ x_2^* &= x_2 = \sin\chi^* \sin\theta^* = \sin\chi \sin\theta \end{aligned}$$

If $k = -1$ hyperbolic lines should be substituted with $cos \to ch$, $sin \to sh$, and no minus sign in the formulae. The observed angle of a distant object is obtained from

$$\theta^* = \pi/2, \qquad \sin\theta = R_0 \sin\chi^* / R_0 \sin\chi$$

and by analogy with the usual relation, an angular distance is defined as $\sin\theta = L_A/d_A$, where $L_A = a(t_e)R_0 \sin\chi^*$ is the diameter of the object (galaxy) at time $t = t_e$, the emission time, and $d_A = a(t_e)R_0 \sin\chi$, its angular distance.

II. Luminosity and distance

II-1 The energy flux

We have already introduced the spherical coordinate transformations associated with a change of origin. The telescope has a diameter $D_0 = a(t_0)r^* = a(t_0)R_0 \sin\chi^*$ with a local origin on the earth, and its area is $\Delta S = \pi D^2/4$. The scale factor $a(t_0)$ is the universal scale at the time of observation t_0, today (in contrast with the case of the angular distance), and is usually set to 1.

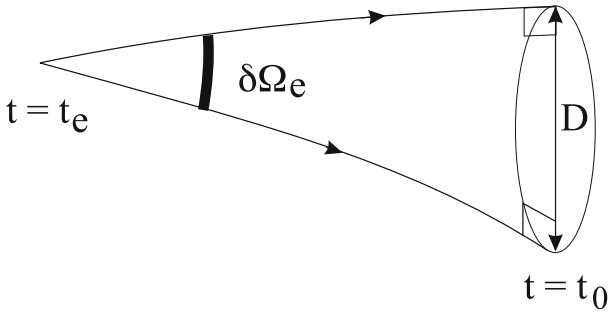

Figure 2. Solid angle at emission and Telescope diameter

The number of photons is conserved from emission to observation:

$$\frac{dN_\gamma}{dt_0 dS} dt_0 \Delta S = \frac{dN_\gamma}{dt_e d\Omega_e} dt_e \left(\frac{\Delta S}{4\pi a^2(t_0) r^2(z_e)} \right)$$

The flux is reduced by two factors equal to $(1+z)^{-1}$: the time dilatation dt_e/dt_0 and the redshift $a(t_0)/a(t_e)$.

$$\frac{dW_0}{dt_0 dS} dt_0 \Delta S = \frac{1}{(1+z)^2} \frac{dW_e}{dt_e d\Omega_e} dt_e \left(\frac{\Delta S}{4\pi a^2(t_0) r^2(z_e)} \right)$$

Defining the (rest frame) luminosity L as

$$L = \frac{dW_e}{dt_e d\Omega_e}$$

$$\frac{dW_0}{dt_0 dS} \Delta S = L \frac{\Delta S}{4\pi d_L^2} \tag{II-1.4}$$

with

$$d_L = a(t_0)(1+z_e) r(z_e) = (1+z_e) r(z_e) \tag{II-1.5}$$

the luminosity distance, while $r(z_e) = R_0 \sin\chi_e$, the (comoving) coordinate at redshift z_e.

II-2 From redshift to distance

It is seen in figure 3 that along the ray trajectory,

$$dl = cdt = a(t)\frac{R_0 dr_a}{1-k(r_a)^2} = R_0 a(t) d\chi$$

(the variable $r_a = r/R_0$ is the same as in equation (1.3) of subsection 1.2). One obtains

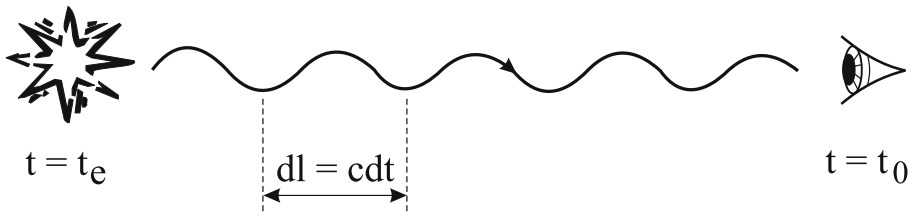

Figure 3. Propagation of a light ray from the SNIa to an observer

$$R_0 d\chi = \frac{cdt}{a(t)} = c\frac{da}{a(da/dt)}$$

$$R_0\chi = c\int_{a(t_e)}^{a(t_0)} \frac{da}{a(da/dt)} = c\int_{a(t_e)}^{a(t_0)} \frac{da}{a^2(da/adt)}$$

The ratio $\dot{a}/a = H(z)$ is the Hubble 'constant' at redshift z, and we can switch from the scale variable $a(t)$ to the redshift z with the relation

$$a(z) = \frac{a_0}{1+z} = \frac{a(z=0)}{1+z} \quad \text{the Doppler formula}$$

so that $dz = -da/a^2$.

The final result is

$$R_0\chi_e = c\int_{z_0}^{z_e} \frac{dz}{H(z)} \tag{II-2.6}$$

The function $H(z)$ has been derived from Friedmann's equation in the lecture of professor D. Langlois:

$$H(z)^2 = H_0^2\left(\Omega_M(z) + \Omega_R(z) + \Omega_\Lambda + \frac{\Omega_k}{a(z)^2}\right) = H_0^2 h^2(z)$$

with $\Omega_k = 1 - \Omega_M(z=0) - \Omega_R(z=0) - \Omega_\Lambda$. The reduced energy densities are known functions of the redshift z, $\Omega_i(z) = \Omega_i(0)(1+z)^{3(1+w_i)}$ where

w_i is the ratio between the partial pressure p_i and the energy density ρ_i ($w_1 = 0$ for the non relativistic matter, $w_2 = 1/3$ for neutrinos and photons, and $w_3 = -1$ for the cosmological constant). The luminosity distance d_L is obtained from equations (1.3),(2.1),(2.2) with different expressions according to the value of k.

$$\begin{aligned} d_L &= (1+z_e)r(z_e) = (1+z_e)R_0 sin\chi_e \quad \text{k = 1} \\ d_L &= (1+z_e)r(z_e) = (1+z_e)R_0 sh\chi_e \quad \text{k = -1} \\ d_L &= (1+z_e)r(z_e) = (1+z_e)R_0\chi_e \quad \text{k = 0} \end{aligned}$$

- d_L is a function of the history of the expansion rate ('Hubble constant') along the ray trajectory.
- $a(z=0) = 1$ has been assumed, else, $a(0)R_0$ should be substituted to R_0 in the previous equations.
- if $k > 0$ there is a maximal distance $a(0)R_0$ in the universe (it does not seem to be the case).

One can define a function $sinh\chi$, with values equal respectively to $\sin\chi$, $sh\chi$, χ, depending on the value of the curvature coefficient $k = 1, -1, 0$.

The spherical (or hyperbolic) coordinate χ is evaluated from the integral

$$I(z_e) = \frac{c}{a_0 H_0}\int_{z_0}^{z_e}\frac{dz}{h(z)} \tag{II-2.7}$$

The comoving coordinate $r(z_e)$ can then be expressed as a function of $I(z)$:

$$r(z_e) = R_0 sinh\left(\frac{I(z_e)}{a_0 R_0 H_0}\right) \tag{II-2.8}$$

$$d_L(z_e) = (1+z_e)r(z_e) \tag{II-2.9}$$

The integral $I(z)$ is a function of the parameters Ω_M, Ω_R, Ω_Λ via the Friedman equation. The radiation density is however negligible in the present universe, as supernovae can only be observed (and produced!) only up to $z \sim 2$. The variation of the luminosity of SNIae as a function of the redshift z is then directly and simply related to the energy density and the geometry of the universe on cosmological scales.

III. The Experimental method

III-1 The Search for Supernovae

Supernovae are found by comparing observed telescope fields with reference fields. After correcting for the variation of the optical properties of the atmosphere (the seeing, which varies between 0.4" and 1.2" depending on the site and the date), a subtraction allows to select variable objects, such as supernovae, cepheids, quasars, Active Galactic Nuclei. An example is given in figure 7. The light curve and the spectral properties contribute to the identification. The first generation of surveys accumulated typically 200 supernovae over a period of five years, spread on a wide variety of instruments. A large angular domain is clearly the key to an efficient discovery rate at moderate values of z. Below $z = 0.8$, a few supernovae can be found each night in a field of a few square degrees.

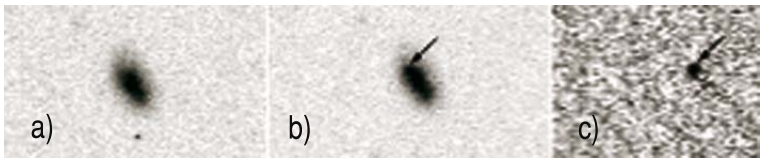

Figure 4. The detection of SNIae by subtraction from [1]

III-2 Luminosity dispersion of SNIae and the time scale

The explosion leading to SNIae is characterised by a fixed mass scale, the Chandrasekhar mass, of about 1.4 solar mass, and the typical time scales in the SN rest frame are about 15 days for the rise time, and 20 days for the decline. Thanks to a scaling law to be explained later between the luminosity and the time evolution of supernovae, this spread can be impressively reduced when the light curve of the supernova is taken into account. The time scale for the light curve can be parametrised either by ΔM_{15}, the drop in the measured (blue filter) magnitude between the peak, and its value 15 days after maximum, or by the stretch parameter s, defined by the ratio of time scales between the observed light curve shape and a reference template. A correction is then applied to the observed magnitude:

$$m_B = m_{obs} + 0.6(s - 1)$$

It is seen in figure 5 that the luminosity dispersion is reduced to 20 % once the stretch (or ΔM_{15}) corrections have been applied. This can be

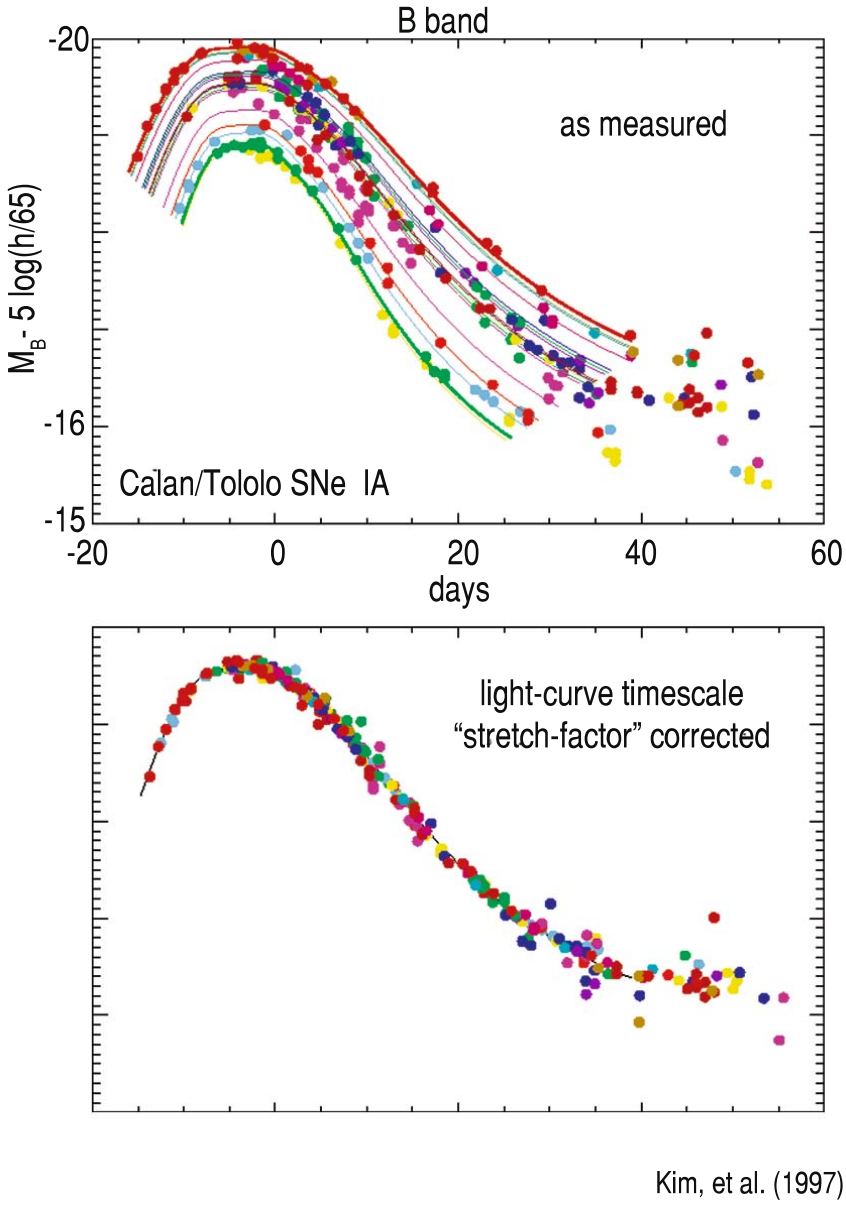

Figure 5. Distribution of the peak luminosity without and with stretch correction

understood semi-quantitatively if the peak luminosity is directly correlated to the mass of ^{56}Ni produced in the explosion. According to [8],[9] the opacity of the final ^{56}Ni increases the diffusion time of the photons, and shifts the peak date to later values. It is quite remarkable that as stressed in [10], the luminosity and the stretch factor are also strongly correlated to spectral features such as the ratios of neighbouring pairs of Si and Ca lines in figure 6.

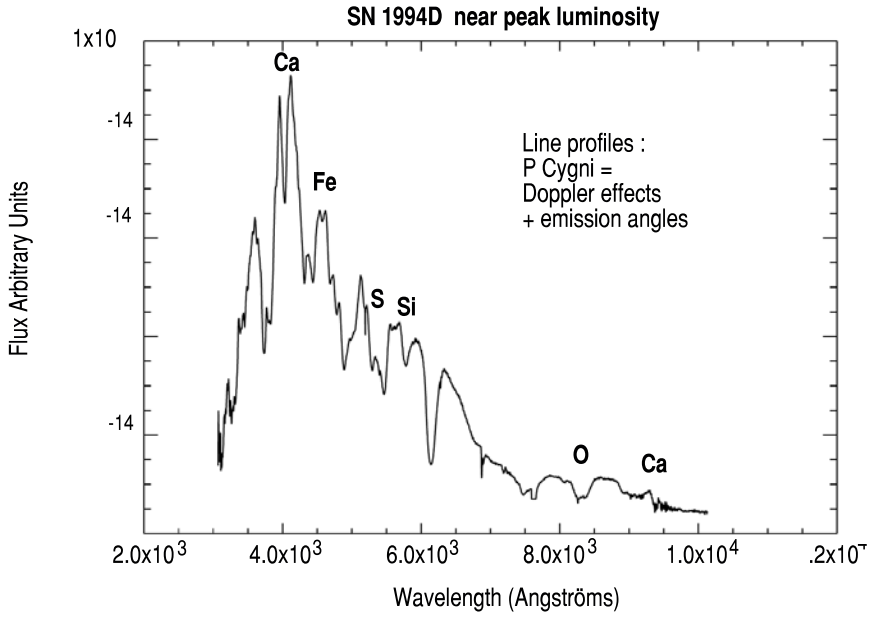

Figure 6. The main features of the SNIa spectrum

III-3 The Colour correction

It is apparent in figure 7 from [2] that there is a relation between the colour of the SN, as defined by the magnitude difference between two filters (B-V) or (V-I) at maximum, and the time-scale parameter ΔM_{15}, which measures the luminosity. Most of this correlation is usually attributed to the galactic extinction in the host galaxy and in ours, without a compelling case. A significant part may be an intrinsic luminosity-colour correlation. The extinction correction does reduce the dispersion

of SN luminosities. In addition, were it not applied, systematic effects from different extinctions in nearby and distant SNIae might arise.

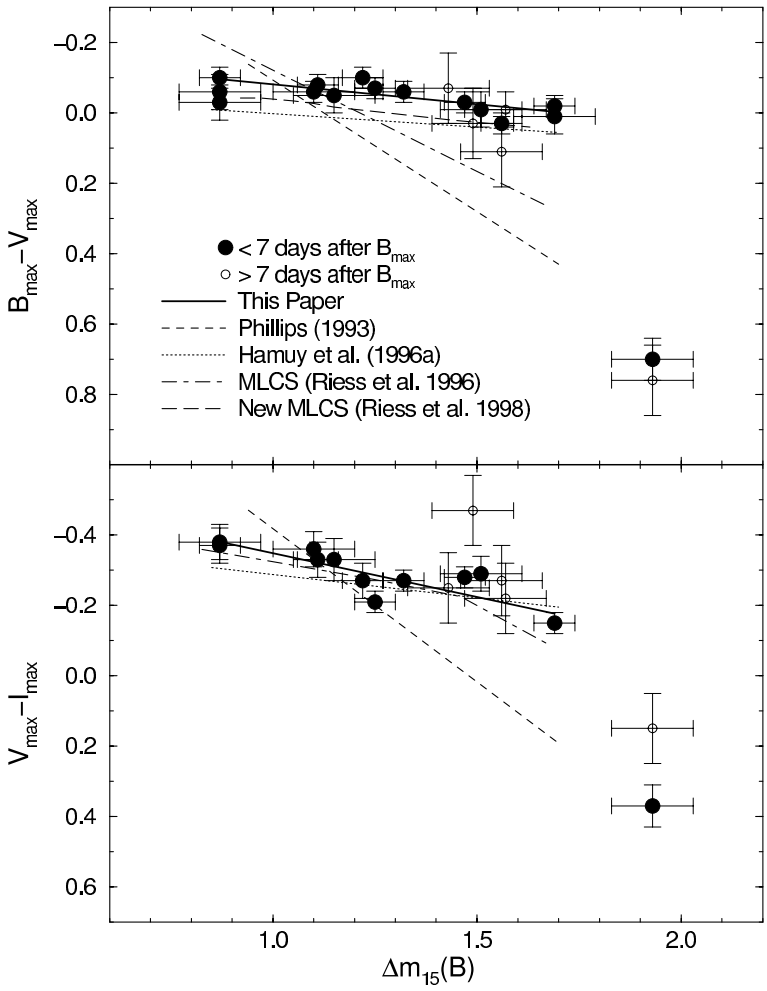

Figure 7. The correlation between colour and ΔM_{15}

IV. The cosmological parameters

IV-1 The determination of the Hubble constant

The determination of the Hubble constant from the luminosity of Type Ia supernovae requires the knowledge of their absolute magnitude. The distance of the host galaxy must then be measured directly without

using its redshift. The standard tools are Cepheids, surface brightness fluctuations, or the Tully-Fisher relation for the rotation. This restricts the distance of the galaxy to our neighbourhood, and the statistics of supernovae in these galaxies is still quite small today. A recent review of the Cepheid calibration of the peak brightness can be found in [3],[4], where the absolute blue magnitude of SNIae is averaged from a sample of 9 supernovae to be $M_B = -19.35 \pm 0.24$ after application of the time scale(ΔM_{15}) and extinction corrections. Assuming this fixed luminosity, the value of the Hubble constant is then derived from the apparent magnitudes in a larger sample with $z < 0.2$ to be $H_0(t_0) = 59.7 \pm 6.3$ km/s/Mpc. It is indeed found in [11] that the dispersion around the $1/z^2$ straight line (for $z < 0.1$) is reduced to about 13 % after the time scale and colour corrections. Although the supernovae can be observed at larger distances than Cepheids, and have a smaller luminosity dispersion, the benefit they give in the H_0 determination is still dependent today on the knowledge of the distances of a very small sample of SN Ia.

IV-2 Dust and evolution

The observed supernovae are less luminous than they would be in an empty universe with $\Omega_\Lambda = \Omega_M = 0$, and by an amount of 30 to 40%. Some unknown galactic or intergalactic dust might generate a similar attenuation. Although the present data do not rule out this contribution, the origin and nature of such a dust would need explanations. The present data in figure 8 of [12] hints that the z dependence might be turning around at $z > 1$, as predicted by cosmology, while the dust interpretation would favour an ever stronger decrease of the luminosity at large z, as seen in figure 9.

IV-3 The matter and energy densities

Once the SNIae have been 'standardized' by the time scale and colour correction, a fit of the luminosities as a function of redshift, using equations (2.1) and (2.5) allows the determination of the parameters Ω_M and Ω_Λ. A large lever arm in z is needed, and the experiments combine different surveys, with different systematics. The probability contours found by the combination of the data sets available in 2003 are shown in figures 10 and 11 from [12], which includes 200 Supernovae: $\Omega_M = 0.60 \pm 0.55$, $\Omega_\Lambda = 1.3 \pm 0.7$, and an $w = -1.1 \pm 0.3$. This result is compatible (to within 1 standard deviation) with the critical density measured by the 3K radiation (see the lecture of G. Smoot), with $\Omega_M + \Omega_\Lambda = 1$ (flatness).

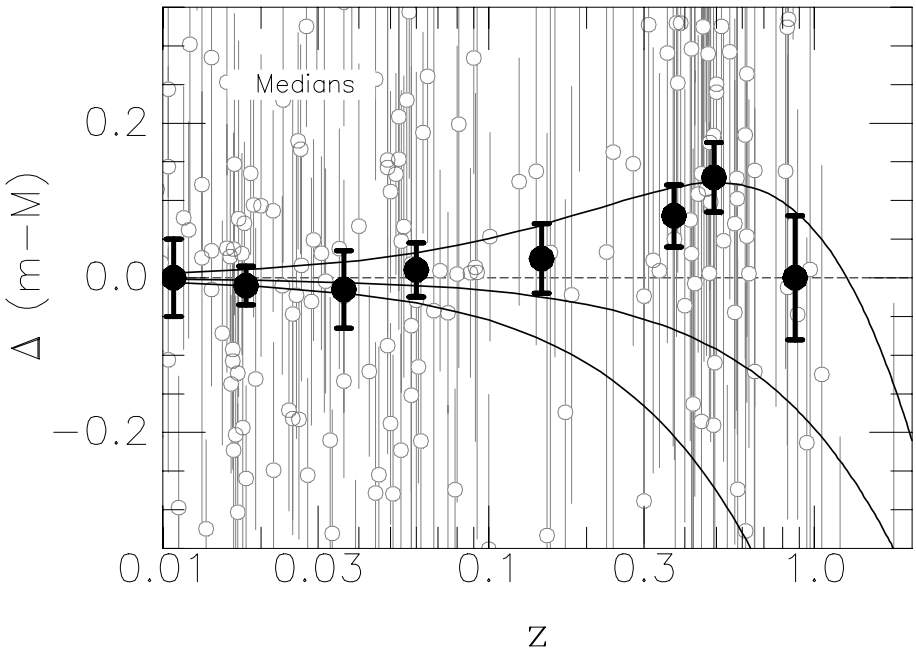

Figure 8. Comparison with an empty universe (from [12])

When the constraint of flatness is included, the contours are modified as indicated in the small ellipse in figure 10 and 11: the corresponding values are $\Omega_M = 0.27 \pm 0.05$, and $\Omega_\Lambda = 0.73 \pm 0.22$

The remaining uncertainty concerns a possible evolution of supernovae from early times ($z = 1$) to the present era. There are neither indications for such an evolution, nor strong limits. It is hoped that the future experiments, with large statistics at low and high z, and good spectroscopy will give quantitative estimates.

V. Future programmes

V-1 Short term

Up to now, the two main surveys by the HZT and SCP collaborations have been painstakingly gathered over a wide set of telescopes, distributed all over the world (WHT, VLT, CFHT, Keck,etc...in the US), with scarce but high quality data from HST. The detection beyond $z = 0.3$ up to $z = 1$ can be performed with instruments of 4m diameter (CFHT,WHT), while the spectral analysis needed for redshifts and

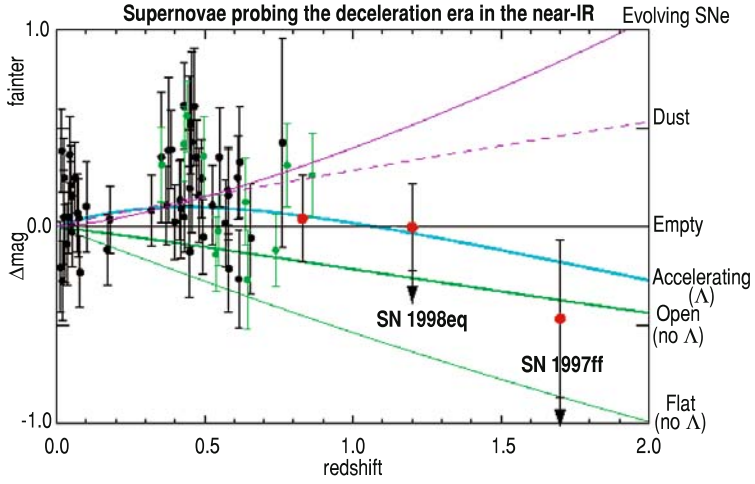

Figure 9. Comparison of different models with a flat (empty) universe

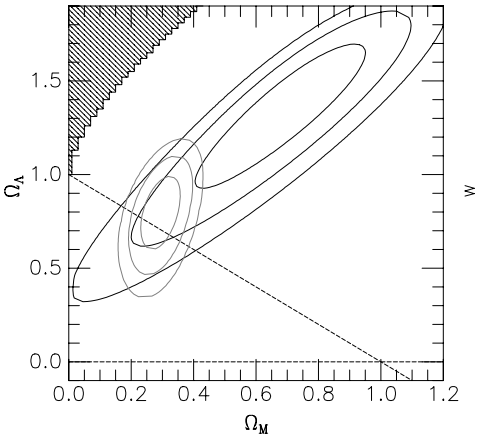

Figure 10. Contour plots for Ω_Λ and Ω_M

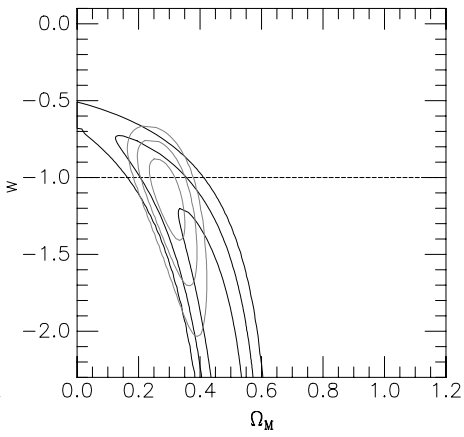

Figure 11. Contour plots for Ω_Λ and w

classification requires typically (for $z > 0.5$) a whole night on an 8m telescope (VLT, Keck).

Several dedicated surveys have now been initiated, the SNLS (Supernovae Legacy Survey) at CFHT and the GOODS ([13]) (Great Observatories Origin Deep Surveys) combining several large space instruments (Chandra, XMM, SIRTF, Hubble). The SNLS survey should provide over 5 years 700 SNIae with $z > 0.3$. It is shown in figure 12 how the SNLS survey, combined with a nearby supernovae sample obtained by SNIFS(Supernovae Integral Field Spectrometer)will reduce the statistical errors on $(\Omega_M, \Omega_\Lambda)$ to $(\pm 0.06, \pm 0.02)$ [14].

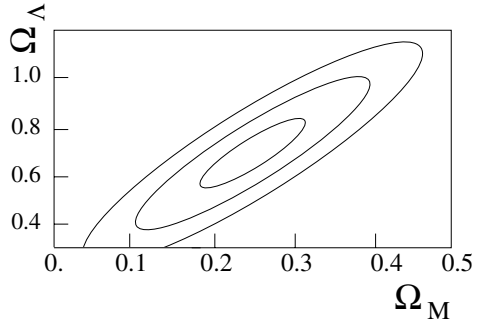

Figure 12. CFHLS contours for Ω_M,Ω_Λ

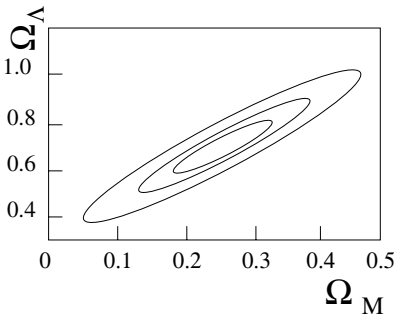

Figure 13. Combining SNLS(distant) and SNIFS(nearby) Supernovae

As has been discussed in the lecture of D. Langlois, the cosmological constant is only an approximate description of a scalar field with an equation of state $p = w/\rho$, where $w \sim -1$. All inflation models predict a variable ratio p/ρ as a function of z for this universal fluid, and the measurement of w and its variation would bring a crucial constraint to the elucidation of the properties of the field. A true fixed value of $w = -1$ has been assumed in figure 13, and the error is seen to be ± 0.1, decreasing to 0.03 if Ω_M is determined from other experiments to an accuracy of 0.03.

The slope parameter $w1$ with $w(z) = w_0 + zw_1$ will remain however very poorly known from the combination of the nearby (SNIFS)and distant (CFHT-LS) samples: space projects are needed to control the systematical errors to the accuracy needed.

V-2 Long term space projects

The ground observations are subject to large atmospheric corrections, which differ intrinsically in the blue (nearby SN) and the IR ranges (remote SN). They will unavoidably limit the quality of high statistics samples collected from the ground, and a dedicated effort for a large sample of SN collected in space is justified.

The Supernova Acceleration Probe (SNAP) has received support from NASA and DOE in the frame of the Joint Dark Energy Mission(JDEM), to be launched within the next ten years. The satellite proposed would be equipped with a 2m telescope. The focal plane consists of 500×10^6 pixels, covering a field of view of about 1 square degree. It would give the instrument an outstanding capability to study:

- supernovae up to $z = 1.7$
- weak lensing
- galactic evolution and clusters

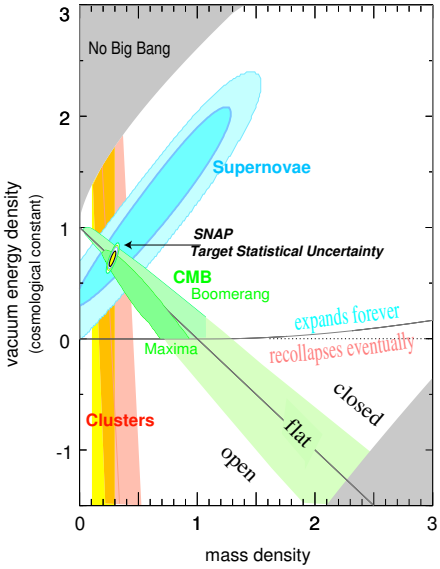

Figure 14. SNAP contours for Ω_M,Ω_Λ from [15]

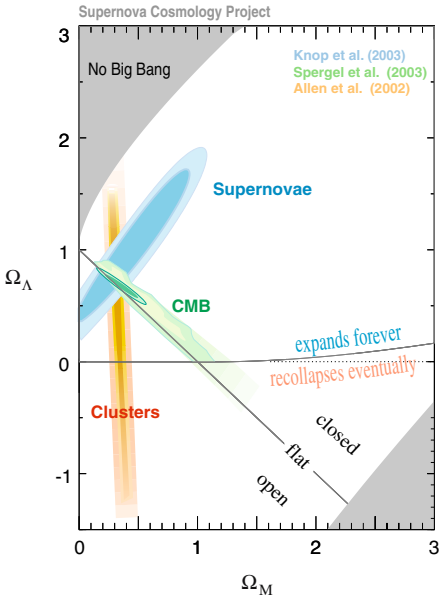

Figure 15. probability contours for Ω_M,Ω_Λ from CMB [16], galaxy clusters [17], SNIae [18]

The probability contours for Ω_M, Ω_Λ as known today are shown in figure 14, and can be compared with the the design accuracy reached by the SNAP/JDEM project in figure 15. Extreme care is needed to provide a significant step in the accuracy on the parameter w and its variation with z, as all calibrations should reach or exceed the 1% accuracy level. On the other hand, the only other technique with a sensitivity to the

equation of state parameter w is the measurement weak lensing, which is also at the edge of experimental analysis.

VI. Acknowledgments

We would like to thank the physicists of the Supernovae Cosmology Project and the High Z Team, who have strongly stimulated the interest into the Supernovae studies worldwide. The french groups are particularly indebted to the SCP collaboration. Ed Baron has straightened several misconceptions concerning the explosion of SNIa. Most of this work was initiated thanks to an invitation of Svetlana Ivanovna, at the university of Dubna (Russia).

References

[1] D. Hardin et al., "Type Ia supernovae rate at $z \sim 0.1$" Astronomy and Astrophysics **362** (2000) 419

[2] M. M. Phillips et al., "The reddening-free decline rate versus luminosity relationship for Type 1A Supernovae" Astrophysical Journal **118**(1999) 1766

[3] A. Saha et al., "Cepheid calibration of the peak brightness of SNIa" Astrophysical Journal **551**(2001) 973

[4] G. A. Tammann et al., "Cepheids,Supernovae,H_0, and the age of the universe" A New Era in Cosmology Conference (2001) (ASP Proceedings)

[5] S. Perlmutter et al., "Measurements of the Cosmological Parameters Ω and Λ from the first 7 Supernovae at $z > 0.35$" Astrophysical Journal **476** (1977) L63

[6] S. Perlmutter et al., "Measurement of Ω and Λ from 42 high red-shift supernovae", Astrophysical Journal **517**(1999) 1

[7] A.G. Riess et al., "Observational evidence from Supernovae for an accelerating universe and a cosmological constant" Astronomical Journal **116**(1998)1009

[8] D. Arnett, "Universality in SNIae and the Phillips Relation", astro-ph/9908169

[9] D. Arnett, Astrophysical Journal **253** (1982) 785

[10] P. Nugent et al., "Evidence for a spectroscopic sequence among type 1a supernovae", Astrophysical Journal **455** (1995) L147

[11] M. M. Phillips et al. "The reddening free decline rate versus luminosity relationship for Type 1a Supernovae" Astrophysical Journal **118**(1999)1766

[12] J. Tonry et al. "Cosmological results from High-z Supernovae" Astrophysical Journal **594**(2003)1

[13] http://www.stsci.edu.ftp/science/goods

[14] http://supernovae.in2p3.fr/ astier/loi.ps

[15] http://snap.lbl.gov (SNAP Science,Mission, and Simulation, presentation by A. Kim)

[16] D. N. Spergel et al. (WMAP), "First Year of WMAP observations, determination of the cosmological parameters" astro-ph/0302209 (June 2003), accepted in Astrophysical Journal

[17] S. W. Allen et al. , Monthly Notices of the Royal Astronomical Society **342** (2003)287

[18] R. A. Knop et al., "New Constraints on Ω_M, Ω_Λ, and w from an independent set of eleven high-redshift supernovae observed with HST" astro-ph/0309368 accepted in Astrophysical Journal

DARK MATTER AND GALAXY FORMATION

Joseph Silk
Astrophysics, Denys Wilkinson Building,
Keble Road,Oxford OX1 3RH, UK
silk@astro.ox.ac.uk

Abstract I describe recent challenges in dark matter. I review the budgets for baryonic and nonbaryonic dark matter. Problems with cold dark matter in the context of galaxy formation are summarized, and possible solutions are presented. I conclude with a description of the prospects for observing cold dark matter.

Keywords: Baryonic matter, black holdes, dark matter, galaxy formation, star formation

I. Challenges of dark matter

We are confronted by a paradox. Baryonic dark matter (BDM) exists and contributes to Ω_b. There are examples of baryonic dark matter, but we cannot reliably calculate the BDM mass fraction. On the other hand, cold dark matter (CDM) is motivated by theory and explains much of the large-scale structure of the universe. CDM dominates Ω_m. We can calculate the relic CDM mass fraction, but no CDM candidate particles are known to exist.

Current observations have attained considerable precision, as a result of surveys over much of the sky. These consist of cosmic microwave background maps, such as WMAP, and galaxy redshift surveys such as 2DF and SDSS. The combination of CMB and LSS experiments WMAP + CBI/ACBAR + 2DF, supplemented with the Lyman alpha forest analysis that probes the density correlations in the intergalactic medium, provides a measure of the homogeneity of the universe and the flatness of space: $\Omega_{total} = 1.02 \pm 0.02$. We infer that the age of the universe satisfies $t_0 = 13.7 \pm 0.2$Gyr. There is known to be dominance of dark matter $\Omega_m = 0.27 \pm 0.02$ relative to baryonic matter. The baryon density amounts to $\Omega_b = 0.044 \pm 0.004$, and is about 15% of the dark matter. Dark energy with $\Omega_\Lambda = 0.70 \pm 0.03$ dominates over dark matter, but only contributes to the unclustered component of the density. The pre-

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 329–347.

ceding results are sensitive to the adopted priors. The uncertainties can be increased by an order of magnitude if more radical priors are adopted, such as inclusion of an admixture of primordial adiabatic and isocurvature density fluctuations [1].

Galaxy clustering also provides a powerful probe of the distribution of dark matter. One constrains the dark matter component by several independent techniques. These include virial theorem estimates with optical data on galaxy redshifts, x-ray emission and hydrostatic support of the hot intracluster gas, and gravitational lensing both of background galaxies and of cluster members. Strong lensing of remote galaxies via the formation of giant arcs probes the cluster core, and weak lensing of distortions to cluster galaxies via gravitationally-induced shear maps the outer region. The Sunyaev-Zeldovich effect supplements these studies by measuring the gas pressure.

II. Global baryon inventory

There is a significant but subdominant mass in dark baryons. Spheroid stars amount to 10% of the baryons or 0.004 in terms of Ω_b. Disk stars contribute 5% or 0.002 in Ω_b. Intracluster gas amounts to 5% or $\Omega_b = 0.002$. The Lyman alpha forest (at $z \sim 0$) contains [2] $29\pm4\%$ of the baryons or $\Omega_b = 0.008$. This is all we actually observe in any quantifiable amount. In addition, intermediate temperature intergalactic gas, the so-called warm/hot intergalactic medium (WHIM) has been detected, at a temperature of $10^5 - 10^6$K. It is estimated from simulations (at $z \sim 0$) to amount to 30% of the local baryons or $\Omega_b = 0.012$, with however a large uncertainty. Indeed the WHIM simulations do not resolve the Jeans mass at the resolution limit, and and the existence of WHIM is purely a theoretical inference, at least in so far as its quantitative fraction is concerned.

The total detected baryonic contribution in the universe, including the hypothesised WHIM, sums to $\Omega_b = 0.028 \pm 0.005$. The corresponding baryon fraction Ω_b/Ω_m is 0.10 ± 0.02. This is to be compared with primordial nucleosynthesis at $z \sim 10^9$: $\Omega_b = 0.04\pm0.004$. In addition the CMB peak heights at $z \sim 1000$ yield a similar value: $\Omega_b = 0.044 \pm 0.003$. Finally, Lyman alpha forest modelling at $z \sim 3$ suggests that $\Omega_b \approx 0.04$. There is also the indirect measurement of baryon fraction from the intracluster gas fraction of 15%. From this, combined with Ω_m, we also find that $\Omega_b \approx 0.04$.

I conclude that approximately $25 \pm 15\%$ of the baryons could be dark. If so, and this is far from a robust conclusion, one can pose the following question: where are they? Intriguingly, the possible shortfall is com-

parable to the mass observed in stars ($\sim \Omega_*$). Could there be a mass in *dark* stars comparable to that in visible stars? Or could early star formation and death have resulted in the ejection of a comparable mass of baryons from the galaxy and its halo? Were the situation to rest here, there would be little reason to take the issue of dark baryons any further. However there is strong local evidence that there is a baryonic deficiency.

III. Confirmation via detailed census of MWG/M31

Observations of our galaxy and also of M31 give direct local measures of the baryon fraction. The virial mass is $M_{virial} \approx 10^{12}\mathrm{M}_\odot$, whereas the mostly stellar baryon mass is $M_b \approx 6 \times 10^{10}\mathrm{M}_\odot$. This yields a baryon fraction of $\sim 6\%$, comparable to the global baryon mass fraction *measured* in stars and gas at $z \sim 0$. A similar amount *may* be present in the intracluster medium. However it is the contents of a volume containing the galaxy, now and at formation, that concern us.

For comparison, the primordial baryon fraction from observations at high redshift is $\sim 15\%$. Presumably these baryons were present initially, when the galaxy formed. Indeed modelling of disk formation requires an initial baryon fraction of $\sim 10 - 15\%$ in order for sufficient cooling to have occurred to form the disk. The "missing" galactic baryons amount to a baryon fraction comparable to what is observed, namely around $\sim 5 - 10\%$ of the dark halo.

If these baryons are indeed in the halo, one possibility is that they are in the form of MACHOs, compact massive objects. The current limit on halo MACHOs is $\lesssim$ 20% of the dark halo mass. The constrained mass range is $10^{-8}\mathrm{M}_\odot$ -10 $\mathrm{M}_\odot$. If the detection claimed by the MACHO collaboration is accepted, the preferred MACHO mass is $\sim 0.5\,\mathrm{M}_\odot$, which favours old white dwarfs: current searches for halo white dwarfs as faint, high proper motion red dwarfs are inconclusive. More generally, one might tolerate a wider mass range for the MACHOs. Theory does not exclude either primordial brown dwarfs ($0.01 - 0.1\,\mathrm{M}_\odot$), primordial black holes (mass $\gtrsim 10^{-16}\mathrm{M}_\odot$) or cold dense H_2 clumps $\lesssim 1\,\mathrm{M}_\odot$ (invoked in the Milky Way halo to account, for example for extreme scattering events [3] or unidentified Scuba objects [4]). Such solutions are difficult to justify in any plausible theoretical framework. There is one intriguing counter-example to the subdominance of H_2 that is typical of spiral galaxies. Diffuse cold molecular gas observed in H_2 rotational emission in the outer disk of NGC 891 could conceivably account for the "missing" baryons, amounting to some 10 times the HI in mass [5], although there

is no evidence for H_2 via H_2 absorption in another metal-poor galaxy [6] nor in galaxies with normal abundances where constraints come from CO emission [7]. The principle alternative is early ejection from the halo.

This could have occurred as a wind in the early, vigorously star-forming phase of galaxy evolution. This inference has additional credence from the facts that the mass in hitherto undetected baryons is comparable to that observed in stars and that the WHIM is observed to exist outside of rich clusters where it is enriched. Moreover it could have a mass comparable to that both in stars and in the colder intergalactic medium.

Ejection via early winds is inferred in the enriched intracluster medium. Observations of Mpc-scale "holes" around Lyman break galaxies, detected via studying absorption of the IGM towards background quasars, support an explanation in terms of early winds from L_* galaxies. The so-called cooling catastrophe in galaxy formation theory, which results in overly luminous massive galaxies can be avoided if early winds eject an amount of baryons comparable to that retained in stars.

Hence the "'missing" baryons *could* be in the WHIM, which would be correspondingly enriched. Unfortunately such strong winds are not supported by hydrodynamical simulations. These use supernovae as the energy source that drives the wind. However the current multiphase simulations lack sufficient fine-scale resolution, as discussed below.

IV. Hierarchical galaxy formation

Galaxy formation and dark matter are intimately related. Before I discuss the connection, I first review the current status of galaxy formation theory. The ab initio approach to large-scale structure has undergone a revolution in the past twenty years, with an improved understanding of the initial conditions of structure formation. This has come about as a confluence of theory and observation. Growth from inflation-boosted quantum fluctuations provides the current paradigm that sets the point of departure for virtually all theories of large-scale structure. The theory of structure growth made one notable prediction that has been verified with outstanding success. This was the existence of fossil cosmic microwave background temperature fluctuations imprinted on the last scattering surface of the cosmic microwave background. The fluctuations are on angular scales that correspond to the comoving scales of the observed large-scale structure in the galaxy distribution. The WMAP satellite, adding unprecedented precision to many earlier experiments, most notably those of BOOMERANG, MAXIMA and DASI, has verified

to within a factor of order unity one of the most remarkable predictions of cosmology, thereby confirming the growth of structure via the gravitational instability of primordial density fluctuations

With the initial conditions specified, it became possible to simulate galaxy formation. Three distinct approaches have emerged: numerical, semi-analytical and hybrid. The fully numerical approach canot yet cope with the complexities of star formation, but has been instrumental in guiding us towards an understanding of the dark matter distribution. The semi-analytical approach has had most success, because it can cope with a wide dynamic range via the extended Press-Schechter formalism, to which is added a prescription for star formation based on baryonic dissipation and plausible but empirical rules. The hybrid approach, combining N-body simulations with a star formation prescription, is particularly useful for its predictive power in observational cosmology, as it is ideal for constructing mock catalogues of galaxies.

There have been some notable successes in the theory of semi-analytic galaxy formation. These include an understanding of the large-scale clustering of galaxies via the primordial density fluctuation power spectrum $P(k)$, including the two-point correlation function $\xi(r)$ and its higher moments, the predictions of the existence of filaments and sheets in the galaxy distribution and of the morphologies of galaxy clusters, the derivations of the cluster and galaxy mass functions, and the predictions of large-scale velocity fields and weak lensing optical depths. On smaller scales, the predictions of galaxy rotation curves and of strong lensing by massive galaxies and galaxy clusters are generally considered to be successes of the theory. Global results that have motivated many observations which are in general agreement with the theory include the cosmic star formation history and the distribution and evolution of HI clouds in the intergalactic medium.

There are complications, however, that demonstrate that we have not yet converged on the ultimate theory of galactic disk formation. There is no fundamental understanding of the Tully-Fisher correlation that relates galaxy luminosity to maximum rotational velocity. Models consistently give too high a normalisation of mass at a given rotation velocity, due to the predominance of dark matter in the model galaxies. There also is some question as to whether the slope is well understood both for samples of nearby disk galaxies which have been carefully corrected for inclination effects, and for distant disk galaxies when projected forward in time for comparison with current epoch samples. This may seem to be a detail, however the fundamental problems are twofold: most of the initial angular momentum in the theoretical models is lost to the dark halo as the disk forms, and the distribution of observed angular momen-

tum is skewed towards high angular momentum in contrast to the initial distribution predicted by the simulations.

Nor is there any fundamental star formation theory for dynamically hot systems such as elliptical galaxies. Appeal must be made to phenomenology. Tidal interactions and mergers are found in simulations to be very effective at concentrating gas into the inner hundreds of parsecs. Ultraluminous infrared galaxies are observed to have star formation rates of hundreds or even thousands of solar masses per year, as inferred if the stars formed in a monolithic collapse of the system. Post-starburst near-infrared light profiles are also suggestive of forming spheroids. Since the ultraluminous infrared galaxies are almost inevitably associated with ongoing mergers or strong tidal interactions with nearby galaxies, it therefore seems entirely plausible that these conditions are capable of driving intense bursts of star formation at the prodigious star formation rates that are observed. Measurements of the molecular gas masses in several such systems at high redshift demonstrate that a very high efficiency indeed of star formation is required, with some $10^{10}\mathrm{M}_{\odot}$ of stars being inferred to form in 10^7 years [39]. This is probably consistent with hierarchical galaxy formation provided that the efficiency of star formation was very high during the first major merger. Why star formation was so efficient is not understood, however.

Colours and spectra of elliptical galaxies at redshift of unity or beyond are suggestive of a very early formation epoch, at least for the stars [13] if not for overall assembly. The cosmic star formation history is likely to be dominated by the precursors of today's ellipticals at $z \gtrsim 2$. Of course such a probe, which relies on galaxy surveys, is rest frame UV flux-limited. However the extragalactic diffuse background light from FIR to optical/UV wavelengths provides a glimpse of all the star formation that ever occurred in the universe. It seems likely that forming dust-shrouded ellipticals dominate the far infrared background above 400μm [14].

V. Unresolved issues in galaxy formation theory

One of the greatest puzzles in galaxy formation theory concerns the distribution of the dark matter. The cold dark matter concentration is predicted from N-body simulations to follow a density profile:

$$\rho = \frac{A}{r^{\gamma}(1 + r/r_s)^{3-\gamma}}.$$

Here, r_s is a scale factor that is incorporated into the concentration parameter, $c \equiv r_v/r_s$, where r_v is either the virial radius or the radius at an overdensity, spherically-averaged, of 200. The profile slope parameter

γ is measured in high resolution N-body simulations ([15], [16]) to be $\gamma \approx 1.2 \pm 0.3$, and the normalisation parameter A reflects the epoch of formation, typically defined to be when half of the present mass was at overdensity of 200.

Unfortunately, observations seem to be in mild disagreement with this predicted profile [17]. A low CDM concentration is observed in low surface brightness dwarf galaxies where the rotation curve is well measured. The predicted dark matter cusp is not usually seen; the typical profile has a soft core, although the interpretation is compounded by issues of disk inclination, of the HI distribution which is usually used to measure the rotation curve, and of the possible mismatch between baryon and CDM potential well depths.

The case for any discrepancy is weakened by the claim that the high resolution numerical simulations extrapolate to a core rasther than to a cusp [43]. This result however is disputed by other simulators who find a central slope $\gamma = 0.16 \pm 0.14$ that holds to 0.3% of the virial radius [19]. In the Milky Way, a low concentration of nonbaryonic dark matter is inferred, with the argument being forcefully made that no more than 10 percent of the total mass interior to the solar circle can be non-baryonic. Theory predicts something like 50 percent for a CDM-dominated universe. However the gravitational microlensing optical depth towards the bulge of the galaxy is used to assess the stellar contribution to the inner rotation curve, and this is uncertain by a factor of ~ 3. This uncertainty has allowed modellers to fit the rotation curve with an NFW ($\gamma = 1$) halo initial density profile that is further concentrated by the adiabatic response of the halo to baryon dissipation.

Another issue is that of dark matter clumpiness. Large numbers of dwarf galaxy halos are predicted at masses comparable to those of the dwarf galaxies in the Local Group, exceeding the observed numbers by an order of magnitude or more. If these systems formed stars, they would be in gross disagreement with observations. If the angular momentum of the baryons is mostly lost to the dark halo as the baryons contract to form the disk, according to simulations, then disk sizes of spiral galaxies are predicted to be smaller by about a factor of 5 than observed. The baryons are clumped and lose angular momentum as a consequence of dynamical friction on the dark matter.

A related prediction is that of the galaxy luminosity function. If the mass in stars tracks that in dark matter, far too many small galaxies are predicted. Too many massive galaxies are also predicted. This has been noted both in isolated groups of galaxies at the L_* level [22] and for the field luminosity function, where an excessive frequency of super-L_* galaxies is expected if a modern value for the initial baryon density

is adopted [23]. The problem arises because the baryons fall into the dark matter potential wells, cool and eventually form stars. There are simply too many cold baryons. If one begins with the baryon fraction predicted by primordial nucleosynthesis of about 15 percent, one ends up with about twice as many baryons as are seen even for the Milky Way galaxy. This issue has been aggravated by recent studies which show that many of the accreting baryons enter the disk cold, without shocking to the virial temperature [24]. This appears to be the dominant form of accretion both for low mass galaxies and at high redshift.

VI. Resurrecting CDM

It would seem that cold dark matter has certain difficulties to overcome. One approach is to tinker with the particle physics by modifying the dark matter, for example by introducing self-interacting or fluid dark matter. This approach is not only non-compelling from the physics perspective but it has also resulted in about as many new difficulties as it purports to resolve. Another strategy is to modify gravity. The less said about this the better: it seems to this author that one should only modify the laws of fundamental physics in the case of true desperation. We are not there yet.

A more promising approach is via astrophysics. The dark matter distribution is inevitably modified by the impact of astrophysical processes. These include dynamical feedback, such as via a massive, transient, rapidly rotating bar. Such gaseous bars are expected to form in the course of a major merger that preceded the first episode of star formation in the protogalaxy, and later would settle into the galactic disk. Indeed up to half of spiral galaxies have significant stellar bars. The initial tumbling of the bar is slowed by dynamical friction on the dark matter. This provides a substantial heat source that is capable of softening the CDM cusp into an isothermal core [25], but see [26] for an independent appraisal of bar-halo angular momentum exchange. The converse consequence is that to explain the observed stellar bars that are generally in rapid rotation, one needs either a deficiency of dark matter, less than 10 percent of the total mass within the region where the bar is observed, or to argue that the observed bars are young. Cold gas infall to disks produces cold stellar disks that can subsequently become bar-unstable [27]. The jury is still out on the history and secular evolution of bars.

A more radical astrophysical approach appeals to the formation of supermassive black holes in the protogalaxy. These must have formed contemporaneously with the oldest stars, as evidenced by the remarkable

correlation between spheroid velocity dispersion and supermassive black hole mass that extends over more than 3 orders of magnitude. Gas accretion onto the supermassive black hole is inevitable in the gas-rich protogalaxy, and provokes violent outflows. It is these outflows that are viewed in the spectra of quasars, the most luminous objects in the universe, and which are powered by accretion onto supermassive black holes. These massive outflows of baryons could have provoked efficient star formation and preferentially expelled the low angular momentum gas. This is a promising, if largely unexplored, source of feedback into the protogalaxy, which offers a potential clue as to why disks and more generally galaxies are the sizes they are, why spheroids formed with great efficiency, why half of the baryons have apparently been expelled from massive galaxies, and why only high angular momentum gas remained to form the disk. As for the impact on the dark matter, rapid loss of more than half the mass in the inner core of the galaxy should leave an impact by softening the dark matter profile.

It is tempting to argue that the dark halo forms as a consequence of substructure mergers, thereby resolving part of the substructure problem if the dwarfs that do not merge are stripped of gas as they become tidally disrupted in the inner halo. The Sagittarius dwarf galaxy provides dramatic evidence of ongoing disruption: a substantial fraction of its mass has already been stripped [29]. Dynamical and chemical evidence from studying tidal debris in the halo of our galaxy suggests that disrupted dwarfs might have contributed up to as much as $\sim 10\%$ of the stellar halo [30]. Gravitational lensing provides evidence for the survival of substructure in massive galaxy halos on $\sim 10^6 M_\odot$ mass scales, amounting to a few percent of the halo mass [31].

VII. An astrophysical solution: early winds

Whether one opts for dynamical heating by tumbling bars or black hole-driven outflows, the impact on the protogalaxy is likely to be dramatic. Bars drive gas into the centre to form and eventually to fuel the SMBH. Mergers of smaller black holes play a subdominant role, especially in the early stages of black hole growth. Gas accretion onto the Super Massive Balck Hole (SMBH) is responsible for fueling the jets and is inferred to be the dominant growth process. This conclusion follows from consideration of quasar outflows. Simulations demonstrate that twin radio jets interact with a clumpy interstellar medium to drive powerful winds that plausibly provide strong feedback into the protogalactic environment. Resolution of the overcooling problem for massive galaxies may require more energy input then is available with a normal initial

stellar mass function. A more exotic solution is possible, and may even be necessary.

There are a number of reasons for believing that massive winds played an important role in galaxy formation. Enrichment of the intracluster gas is observed to $\sim Z_\odot/3$. This cannot be explained by current epoch star formation activity, or indeed by past activity unless substantial mass ejection occurred. Intracluster magnetic fields are observed at a level that is about 10 percent of the typical galactic value. Ejection of magnetic flux from galaxies early in the lifetime of the galaxies seems to be the most plausible explanation. At a redshift of about 3, the Lyman break galaxies are inferred to have outflows to velocities of $\sim$ 600km/s. More indirectly, absorption against background quasars near these galaxies has revealed evidence for a proximity effect on the intergalactic medium. This is in the form of a deficiency of HI that is observed as an increase in the transparency to Lyα and possibly CIV absorption extending out to about $\sim$ 1 Mpc from the Lyman break galaxies [32].

However numerical simulations of early supernova-driven winds fail to find any evidence for substantial gas ejection from luminous ($\sim L_*$) galaxies [33]. One can ask what is wrong with the hydrodynamic simulations? Certainly, the simulations lack adequate resolution. Rayleigh-Taylor instabilities enhance wind porosity and Kelvin-Helmholtz instabilities enhance wind loading of the cold interstellar medium. Both effects are certain to occur and will enhance the wind efficacity. Yet another omission is that one cannot yet resolve the motions of massive stars before they explode. This means that energy quenching is problematic and the current results are inconclusive for typical massive galaxies.

An interesting observation from studies of nearby starbursts is that the mass outflow rate is of the same order of magnitude as the star formation rate. This result is also motivated by a multiphase interstellar medium in which supernova explosions provide the main energy and momentum source [28], and naturally can account for why around half of the baryons are ejected. This mechanism should be effective for sub-L_* galaxies. It is unlikely however that there is sufficient energy available with a normal initial mass function to drive winds from the most massive forming galaxies. In fact, the simulation initial conditions assume that the winds are driven by supernovae produced by massive stars whose initial mass function is similar to that found in the solar neighbourhood. This is a dangerous assumption, given that we have no fundamental theory of the initial mass function, and that conditions both in massive starbursts and in the early universe may be quite different from anything sampled locally. A top-heavy initial mass function is one way to boost

the specific energy and momentum input by up to an order of magnitude. It has been speculated that a top-heavy initial mass function is necessary to account for the high efficiencies of star formation observed in certain very high redshift ultraluminous infrared sources. This option has also been invoked in order to account for the surprisingly high redshift of reionisation found by the WMAP satellite [34] and for the intracluster gas enrichment [35].

Another possibility is that some of the early supernovae may in fact be hypernovae. A hypernova has up to 10^{53} ergs of kinetic energy. If one supernova in 10 at high redshift is in fact a hypernova, the specific energy input is boosted by as much as an order of magnitude. The case for an enhanced hypernova fraction at high redshift is based on the nucleosynthetic evidence from abundances measured for the oldest stars in our halo. The enhancements of zinc and chromium and deficiency of iron in these stars can be explained in terms of hypernova yields. In hypernovae, the energy output is boosted by infall of the inner rotating core onto a black hole, and the corresponding ejecta mass cuts for precursors of $\sim 25\mathrm{M}_{\odot}$ reflect the observed abundance anomalies relative to standard supernova yields [36]. Some subset of hypernovae are also a possible source of the r-process nuclear enhancement seen in the oldest stars.

Finally, the ubiquitous AGN, as traced by the presence of supermassive black holes that amount to ~ 0.001 of the spheroid mass, would inevitably have been activated in the gas-rich protogalactic environment. The supermassive black hole is presumed to achieve most of its growth by gas accretion from a circumnuclear disk. This would inevitably have been accompanied by intense jets of relativistic plasma that provide a means of exerting strong positive feedback onto the protogalactic environment [37]. Jet propagation into a clumpy interstellar medium destabilises the jet and generates high porosity in the relativistic jet fluid that eventually fills the entire diffuse interstellar medium [38]. A likely consequence is the compression of massive clumps of cold gas and the subsequent triggering of star formation. Such positive feedback is precisely what may be required to account for the high efficiency of star formation inferred for massive ellipticals, as characterized by the observed enhancement in $[\alpha/Fe]$ abundance ratios and by the predominance of red stellar populations at high redshift.

QSO observations attest to the role of AGN in star formation, with 30% of high redshift QSOs being luminous far infrared sources with substantial amounts of molecular gas either detected or inferred [39]. The CO and FIR observations are indicative of high star formation rates of $\lesssim 1000\mathrm{M}_{\odot}$ per year, with the star formation efficiency rising for the

most luminous FIR QSOs [40]. The presence of large amounts of gas, heavy elements and dust in QSOs at $z \gtrsim 6$ argues for a causal connection between the AGN trigger and the high efficiency of star formation.

VIII. Observing CDM via the WIMP LSP

Assuming that the WIMPS once were in thermal equilibrium, one finds that the relic WIMP froze out at

$$n_x < \sigma_{ann} v > t_H \lesssim 1 \Longrightarrow T \lesssim m_\chi/20k.$$

From this, one infers that the relic CDM density is $\Omega_x \sim \sigma_{weak}/\sigma_{ann}$. It is useful to know the mass range of the WIMPs in order to define search parameters. Minimal SUSY has many free parameters, and most of them are generally suppressed in parameter searches. For example, requiring the relic neutralino density to be within mSUGRA greatly reduces the parameter space for possible masses [41]. If the WIMP is a SUSY neutralino, simple scaling arguments yield

$$< \sigma_{ann} v > \propto m_\chi^2 \quad \text{for} \quad m_\chi \ll Z^0$$

and

$$< \sigma_{ann} v > \propto m_\chi^{-2} \quad \text{for} \quad m_\chi \gg Z^0,$$

thereby defining a window of opportunity for dark matter. Stability is assumed for the SUSY LSP to be a WIMP candidate, usually via R-parity conservation. From accelerator limits combined with model expectations, the allowed mass range is conservatively found to satisfy

$$50\text{GeV} \lesssim m_\chi \lesssim 1\text{TeV}.$$

Accelerator limits set a lower bound, and the inclusion of the extra degrees of freedom from coannihilations sets an upper bound. Direct searches may also independently set a model-dependent lower bound.

Indirect searches via halo annihilations of the LSP into $\gamma, \bar{p}, e^+, \nu$ have hitherto been inconclusive. There are hints of an anomalous feature in the high energy e^+ spectrum. However halo detection of e^+ requires clumpiness of order

$$< n^2 > / < n >^2 \sim 100,$$

both to get sufficient flux and to allow the possibility of a nearby clump which might allow the observed spectral feature to be reproduced [42]. Such clumpiness could also boost the predicted gamma ray flux from annihilations into the range observable by EGRET. Clumpiness of this order is indeed predicted by galaxy halo simulations. However this generally applies in the outer halo. The γ-ray flux towards the galactic centre

is observed to have a hard spectrum (as expected for annihilations), but the clumps would not survive the tidal disruptions that are inevitable in the inner galaxy [43]. To account for the observed diffuse gamma ray flux from the direction of the galactic centre, one would need to have a very steep density profile ($\rho \overset{\propto}{\sim} r^{-1.5}$). It has been argued that this would conflict with microlensing observations and the inner rotation curve of the Galaxy. A detailed attempt at modelling the inner core including both rotation curve and microlensing constraints concludes however that the NFW profile combined with adiabatic compression of the dark matter leads to a consistent model with a potentially detectable gamma ray signal. Detection could be accomplished with an atmospheric Cerenkov telescope [45] which has the advantage of good angular resolution and a threshold that could probe neutralinos with masses as low as $\sim$ 100GeV.

IX. The future

There are exciting prospects for addressing many of the challenges facing galaxy formation and dark matter. With regard to directly observing forming galaxies, we can look forward to sampling the galaxy luminosity function at redshifts beyond unity with both SIRTF and ground-based NIR spectroscopy. The theory of multiphase galaxy formation is certain to be greatly refined, incorporating dynamical feedback and the impact of supermassive black holes. We will probe scales down to $\sim 10^6 M_\odot$ via spectroscopic gravitational lensing. Baryonic dark matter will be mapped at UV/SXR wavelengths. In the area of indirect detection of CDM, new experiments will search for high energy halo annihilation signatures in the form of $\gamma, e^+, \bar{p}$ and ν. Over the next 5 years, these experiments will include GLAST, HESS, MAGIC, VERITAS, ICECUBE, ANTARES, PAMELA and AMS. High energy neutrinos from annihilations in the sun (and earth) will be probed, thereby providing a measure of the cold dark matter density at the solar circle.

The Galactic Centre could provide a "smoking gun" with radio synchrotron, γ-ray and ν data: annihilations measure cold dark matter where Milky Way formation began "inside-out", some 12 Gyr ago. Accretion models onto the central black hole fail to give sufficient low frequency radio or gamma ray emission to account for the observed fluxes from SagA*, and it is tempting to invoke a more exotic alternative. For example, the low frequency radio emission can be explained by spike-enhanced self-absorbed synchrotron emission, contrary to recent claims. Even choice of equipartition magnetic fields is allowed with appropriate choice of a suitable initial dark matter density profile. In fact this choice depends on uncertain astrophysics that includes the past history of the

formation of the SMBH, including mergers and associated heating and radiation recoil.

However even if the history of the supermassive black hole at the centre of the galaxy were to disfavour a significant cold dark matter spike, as a consequence of something as mundane as dynamical heating by stellar encounters [44], one might expect lesser spikes to survive around other relic masssive black holes. The central supermassive black hole and the bulge of the galaxy most likely formed from the mergers of protogalactic dwarf galaxies that themselves contained smaller black holes. It is the seed black holes that formed by accretion which should retain initial CDM spikes. Dynamical mergers of black holes, if they occur, result in black hole ejection from shallow potential wells as a consequence both of formation of unstable 3-body systems and radiation recoil in 2-body mergers. This model suggests that there should be relic "naked" intermediate mass black holes in the inner halo [20]. The adiabatic growth of these seed black holes should have generated local spikes in cold dark matter that could have survived and maintained a density profile

$$\rho \propto r^{-\gamma} \Rightarrow \rho \propto r^{-\gamma'}, \quad \text{with} \quad \gamma' = \frac{9-2\gamma}{4-\gamma}.$$

Annihilation fluxes would be enhanced, to a level where such sources could possibly account for a subset of the unidentified EGRET gamma ray sources.

The preceding interpretation rests heavily on the hypothesis that the dark matter consists primarily of the lightest $N = 1$ SUSY neutralinos. This is well motivated, but as has often been emphasized, the most compelling and elegant explanation of any natural phenomenon is often false. Of course, if accelerator evidence were found for SUSY, the odds in favour of a neutralino explanation of dark matter would be dramatically increased.

Another intriguing option is that of $N = 2$ SUSY, which could allow the possibility of two CDM relic candidate particles. One might take the dominant, heavier species to be the conventional WIMP, with a light, subdominant, counterpart with a correspondingly larger cross-section. The mass scaling of the gamma ray flux($\propto < \sigma v > m_\chi^{-2}$) allows the light dark matter candidate to be subdominant for a given flux and relic density (or equivalently, annihilation cross-section), if the cross-section is S-wave suppressed. A recent interpretation [46] of 511 keV line emission from the Galactic centre and bulge region detected by the INTEGRAL gamma ray satellite appeals to MeV dark matter. These particles annihilate via e^+e^- pair production, and the positrons decelerate in the interstellar medium to generate a narrow annihilation

line that simultaneously matches the observed flux and angular profile. There are alternative explanations, and potential tests include searching for 511 keV line emission from the Sagittarius dwarf galaxy as well as from the brightest low mass x-ray binary stars. The former would confirm this idea, the latter would support an alternative, more conventional explanation in terms of astrophysical accelerators. Low mass x-ray binaries are accreting neutron stars that have a spatial distribution which is similar to that of Population II, and generate weak radio jets that are a possible injection source of energetic positrons into the interstellar medium.

Despite our failure to converge on a dark matter candidate, dark matter is here to stay. It is exceedingly difficult to construct a theory of galaxy formation without some compelling evidence for the nature of the dark matter. We assume that the dark matter is cold and stable, and this results in beautiful simulations of cosmic structure that meet many, but by no means all, of the observational challenges. Our hope is that with increasingly refined probes of galaxies near and far, we will be able to construct a strong inferential case for the required properties of the dark matter. Indeed, even now we are not far from this goal in so far as our modelling of large-scale structure is concerned.

On smaller scales, however, the picture, and the corresponding role of dark matter, is much less clear. It is particularly disconcerting that we know so little about the fundamental physics of star formation, despite decades of detailed observations. It is only too tempting to assume that conditions in the distant universe, while being far more extreme than those encountered locally, nevertheless permit us to adopt similar rules and inputs for star formation. We may be easily misled. Galaxy formation moreover rests on knowledge of the initial conditions that seeded structure formation, and that we measure in the cosmic microwave background. Here too it is worth recalling that our conclusions are only as robust as the initial priors. Change these substantially, and new modes of fluctuations are allowed that can, for example, permit a much earlier epoch of massive galaxy formation than in the standard model. It is clear that only increasingly refined and precise observations will guide us: if evidence were to be confirmed for a hypothesis that was far from our current prejudices, theory would rapidly adapt. We should bear in mind that Nature has more surprises than we can imagine, otherwise physics would be hopelessly dull.

I thank my colleagues especially at Oxford for their unwavering enthusiasm in discussions about many of the topics covered here. In particular I acknowledge the contributions of Reba Bandyopadhay, Celine Boehm,

Greg Bryan, Julien Devriendt, Ignacio Ferreras, Dan Hooper, Hugues Mathis, Adrianne Slyz and James Taylor.

References

[1] M. Bucher,J. Dunkley, P. G. Ferreira, K. Moodley and C. Skordis, PRL, submitted, astro-ph/0401417 (2004)

[2] S. Penton, J. Stocke and J. M. Shull, ApJ Suppl., in press, astro-ph/0401036 (2004)

[3] M. Walker and M. Wardle, ApJ, 498 (1998) L125

[4] A. Lawrence, MNRAS, 323 (2001) 147L

[5] E. Valentijn and P. van der Werf, ApJ, 522 (1999) L29

[6] ApJ, 538 (2000) L77

[7] A. Boselli, J. Lequeux and G. Gavazzi, A&A, 384 (2002) 33

[8] D. Thilker et al., ApJ Letters, in press, astro-ph/0311571 (2003)

[9] G. Kauffmann et al., MNRAS, 341 (2003) 54

[10] T. Kranz, A. Slyz and H. Rix, ApJ, 586 (2003) 143

[11] C. Conselice et al., in IAU Symposium 220 "Dark Matter in Galaxies," Sydney, July 2003, to be published by ASP, astro-ph/0312352 (2003)

[12] F. Bertoldi et al., A&A Letters, in press, astro-ph/0307408 (2003)

[13] P. G. van Dokkum et al., ApJ, 587 (2003) L83 -L88

[14] C. Balland, J. E. G. Devriendt and J. Silk, MNRAS, 343 (2003) 107

[15] T. Fukushige, A. Kawai and J. Makino, ApJ, submitted, astro-ph/0306203 (2003)

[16] J. F. Navarro et al., MNRAS, submitted, astro-ph/0311231 (2003)

[17] M.R. Merrifield, in IAU Symposium 220 "Dark Matter in Galaxies," Sydney, July 2003, to be published by ASP, astro-ph/0310497 (2003)

[18] F. Stoehr, MNRAS, submitted, astro-ph/0403077 (2004)

[19] J. Diemand, B. Moore and J. Stadel, MNRAS, submitted, astro-ph/0402267 (2004)

[20] R. R. Islam, J. E. Taylor and J. Silk, MNRAS, in press, astro-ph/0307171 (2003)

[21] C. Maraston et al., A&A, 400 (2003) 823-840

[22] E. D'Onghia and G. Lake, ApJ Letters, submitted, astro-ph/0309735 (2003)

[23] A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh and S. Cole, ApJ, submitted, astro-ph/0302450 (2003)

[24] Y. Birnboim and A. Dekel, MNRAS, in press, astro-ph/0302161 (2003)

[25] K. Holley-Bockelmann, M. Weinberg. and N. Katz, MNRAS, submitted, astro-ph/0306374 (2003)

[26] E. Athanassoula, in IAU Symposium 220 "Dark Matter in Galaxies", Sydney, July 2003, to be published by ASP, astro-ph/0312155 (2003).

[27] D. L. Block, F. Bournaud, F. Combes, I. Puerari and R. Buta, in IAU Symposium 217 "Recycling Intergalactic and Interstellar Matter", Sydney, July 2003, to be published by ASP (2003)

[28] J. Silk, MNRAS, 343 (2003) 249

[29] S. Majewski et al., Contribution to proceedings of "IAU Symposium 220: Dark Matter in Galaxies", eds. S. Ryder, D.J. Pisano, M. Walker, and K. Freeman, to be published by ASP, astro-ph/0311522 (2003)

[30] E. Tolstoy and K. Venn, to be published in the Proceedings of Joint Discussion 15 of the 2003 IAU, eds. P. Nissen and M. Pettini, astro-ph/0402295 (2004)

[31] R.B. Metcalf, L.A. Moustakas, A.J. Bunker and I.R. Parry, ApJ, submitted, astro-ph/0309738 (2003)

[32] K.L. Adelberger, C.C. Steidel, A.E. Shapley and M. Pettini, ApJ, 584 (2003) 45-75

[33] V. Springel and L. Hernquist, MNRAS, 339 (2003) 289

[34] B. Ciardi, A. Ferrara and S. D. M. White, MNRAS, 344 (2003) L7

[35] L. Portinari, A. Moretti, C. Chiosi, J. Sommer-Larsen, ApJ, in press, astro-ph/0312360 (2003)

[36] H. Umeda and K. Nomoto, ApJ, submitted, astro-ph/0308029 (2003)

[37] W. van Breugel, C. Fragile, P. Anninos and S. Murray, in IAU Symposium 217 "Recycling Intergalactic and Interstellar Matter", Sydney, July 2003, to be published by ASP (2003).

[38] G. Bicknell, C. SAxton, R. Sutherland, S. Midgley and S. Wagner, New Astronomy Reviews, 47 (2003) 537 (2004)

[39] F.Bertoldi et al., A&A, 406 (2003) L55-L58

[40] Y. Gao and P. Solomon, ApJ,in press, astro-ph/0310339 (2003)

[41] J. Edsjo, M. Schelke, P. Ullio and P. Gondolo, JCAP 0304 (2003) 001

[42] D. Hooper, J. E. Taylor and J. Silk, PRD, submitted, hep-ph/0312076 (2003)

[43] F. Stoehr, S. D. M. White, V. Springel, G. Tormen and N. Yoshida, MNRAS, 345 (2003) 1313

[44] D.Merritt, preprint, astro-ph/0311594 (2003)

[45] F. Prada et al., PRL, submitted, astro-ph/0401512 (2004)

[46] C. Boehm et al., PRL, in press, astro-ph/0309686 (2004)

SUPERSYMMETRIC EXTENSION OF THE STANDARD MODEL

Dimitri Kazakov
BLTP, JINR, Dubna and ITEP, Moscow
kazakovd@thsun1.jinr.ru

Abstract The present lectures contain an introduction to supersymmetry, a new symmetry that relates bosons and fermions, in particle physics. The motivation to introduce supersymmetry is discussed. The main notions of supersymmetry are introduced. The supersymmetric extension of the Standard Model - the Minimal Supersymmetric Standard Model - is considered in more detail. Phenomenological features of the MSSM as well as possible experimental signatures of SUSY are described.

Keywords: Supersymmetry, unification, MSSM, hierarchy, R-parity

I. Introduction: What is supersymmetry

Supersymmetry is a *boson-fermion* symmetry that is aimed to unify all forces in Nature including gravity within a singe framework [1]-[4]. Modern views on supersymmetry in particle physics are based on string paradigm, though the low energy manifestations of SUSY can be possibly found at modern colliders and in non-accelerator experiments.

Supersymmetry emerged from the attempts to generalize the Poincaré algebra to mix representations with different spin [1]. It happened to be a problematic task due to the no-go theorems preventing such generalizations [5]. The way out was found by introducing the so-called graded Lie algebras, i.e. adding the anti-commutators to the usual commutators of the Lorentz algebra. Such a generalization, described below, appeared to be the only possible one within relativistic field theory.

If Q is a generator of SUSY algebra, then acting on a boson state it produces a fermion one and vice versa

$$\bar{Q}|boson>=|fermion> \quad \text{and} \quad Q|fermion>=|boson> .$$

Since bosons commute with each other and fermions anticommute, one immediately finds that SUSY generators should also anticommute,

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 349–395.

they must be *fermionic*, i.e. they must change the spin by a half-odd amount and change the statistics. Indeed, the key element of SUSY algebra is

$$\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = 2\sigma^\mu_{\alpha,\dot{\alpha}} P_\mu, \tag{I.1}$$

where Q and $\bar{Q}$ are SUSY generators and P_μ is the generator of translation, the four-momentum.

In what follows we describe SUSY algebra in more detail and construct its representations which are needed to build a SUSY generalization of the Standard Model of fundamental interactions. Such a generalization is based on a softly broken SUSY quantum filed theory and contains the SM as a low energy theory.

Supersymmetry promises to solve some problems of the SM and of Grand Unified Theories. In what follows we describe supersymmetry as a nearest option for the new physics on a TeV scale.

II. Motivation of SUSY in particle physics

II.1 Unification with gravity

The *general idea* is a unification of all forces of Nature including quantum gravity. However, the graviton has spin 2, while the other gauge bosons (photon, gluons, W and Z weak bosons) have spin 1. Therefore, they correspond to different representations of the Poincaré algebra. To mix them one can use supersymmetry transformations. Starting with the graviton state of spin 2 and acting by SUSY generators we get the following chain of states:

$$spin\ 2 \quad \to \quad spin\ 3/2 \quad \to \quad spin\ 1 \quad \to \quad spin\ 1/2 \quad \to \quad spin\ 0.$$

Thus, a partial unification of matter (fermions) with forces (bosons) naturally arises from an attempt to unify gravity with other interactions.

Taking infinitesimal transformations $\delta_\epsilon = \epsilon^\alpha Q_\alpha$, $\bar{\delta}_{\bar{\epsilon}} = \bar{Q}_{\dot{\alpha}} \bar{\epsilon}^{\dot{\alpha}}$, and using eq.(I.1) one gets

$$\{\delta_\epsilon, \bar{\delta}_{\bar{\epsilon}}\} = 2(\epsilon \sigma^\mu \bar{\epsilon}) P_\mu, \tag{II.1}$$

where ϵ is a transformation parameter. Choosing ϵ to be local, i.e. a function of a space-time point $\epsilon = \epsilon(x)$, one finds from eq.(II.1) that an anticommutator of two SUSY transformations is a local coordinate translation. And a theory which is invariant under local coordinate transformation is General Relativity. Thus, making SUSY local, one naturally obtains General Relativity, or a theory of gravity, or supergravity [2].

II.2 Unification of gauge couplings

According to the Grand Unification *hypothesis*, gauge symmetry increases with energy [6]. All known interactions are different branches of a unique interaction associated with a simple gauge group. The unification (or splitting) occurs at high energy. To reach this goal one has to consider how the couplings change with energy. This is described by the renormalization group equations. In the SM the strong and weak couplings associated with non-Abelian gauge groups decrease with energy, while the electromagnetic one associated with the Abelian group on the contrary increases. Thus, it becomes possible that at some energy scale they become equal.

After the precise measurement of the $SU(3) \times SU(2) \times U(1)$ coupling constants, it has become possible to check the unification numerically. The three coupling constants to be compared are

$$\begin{aligned} \alpha_1 &= (5/3)g'^2/(4\pi) = 5\alpha/(3\cos^2\theta_W), \\ \alpha_2 &= g^2/(4\pi) = \alpha/\sin^2\theta_W, \\ \alpha_3 &= g_s^2/(4\pi) \end{aligned} \tag{II.2}$$

where g', g and g_s are the usual $U(1)$, $SU(2)$ and $SU(3)$ coupling constants and α is the fine structure constant. The factor of $5/3$ in the definition of α_1 has been included for proper normalization of the generators.

In the modified minimal subtraction ($\overline{MS}$) scheme, the world averaged values of the couplings at the Z^0 energy are obtained from a fit to the LEP and Tevatron data [7]:

$$\begin{aligned} \alpha^{-1}(M_Z) &= 128.978 \pm 0.027 \\ \sin^2\theta_{\overline{MS}} &= 0.23146 \pm 0.00017 \\ \alpha_s &= 0.1184 \pm 0.0031, \end{aligned} \tag{II.3}$$

that gives

$$\alpha_1(M_Z) = 0.017, \quad \alpha_2(M_Z) = 0.034, \quad \alpha_3(M_Z) = 0.118 \pm 0.003. \tag{II.4}$$

Assuming that the SM is valid up to the unification scale, one can then use the known RG equations for the three couplings. In the leading order they are:

$$\frac{d\tilde{\alpha}_i}{dt} = b_i\tilde{\alpha}_i^2, \qquad \tilde{\alpha}_i = \frac{\alpha_i}{4\pi}, \qquad t = log(\frac{Q^2}{\mu^2}), \tag{II.5}$$

where for the SM the coefficients are $b_i = (41/10, -19/6, -7)$.

The solution to eq.(II.5) is very simple

$$\frac{1}{\tilde{\alpha}_i(Q^2)} = \frac{1}{\tilde{\alpha}_i(\mu^2)} - b_i log(\frac{Q^2}{\mu^2}). \qquad \text{(II.6)}$$

The result is demonstrated in Fig.1 showing the evolution of the inverse of the couplings as a function of the logarithm of energy. In this presentation, the evolution becomes a straight line in first order. The second order corrections are small and do not cause any visible deviation from a straight line. Fig.1 clearly demonstrates that within the SM the coupling constant unification at a single point is impossible. It is excluded by more than 8 standard deviations. This result means that the unification can only be obtained if new physics enters between the electroweak and the Planck scales.

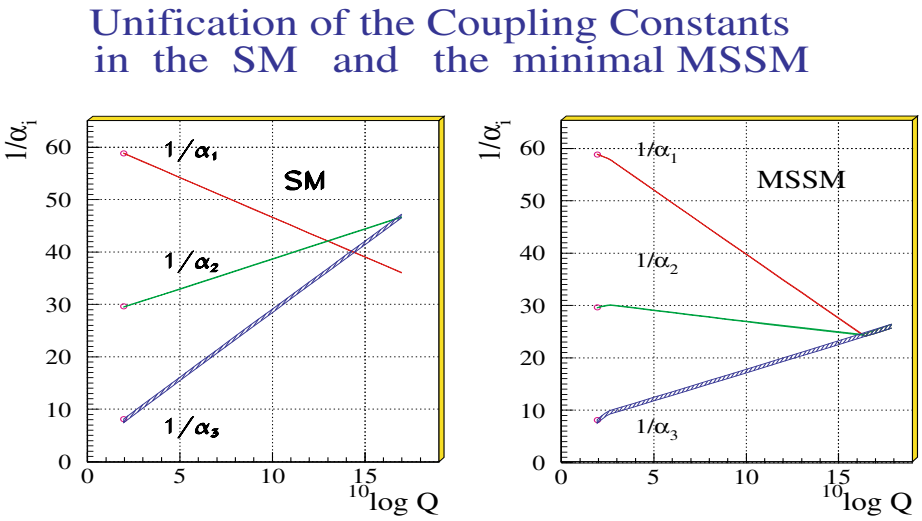

Figure 1. Evolution of the inverse of the three coupling constants in the Standard Model (left) and in the supersymmetric extension of the SM (MSSM) (right).

In the SUSY case, the slopes of the RG evolution curves are modified. The coefficients b_i in eq.(II.5) now are $b_i = (33/5, 1, -3)$. The SUSY particles are assumed to effectively contribute to the running of the coupling constants only for energies above the typical SUSY mass scale. It turns out that within the SUSY model a perfect unification can be obtained as is shown in Fig.1. From the fit requiring unification one finds for the break point M_{SUSY} and the unification point M_{GUT} [8]

$$\begin{aligned} M_{SUSY} &= 10^{3.4\pm0.9\pm0.4}\ GeV, \\ M_{GUT} &= 10^{15.8\pm0.3\pm0.1}\ GeV, \\ \alpha_{GUT}^{-1} &= 26.3 \pm 1.9 \pm 1.0, \end{aligned} \qquad \text{(II.7)}$$

The first error originates from the uncertainty in the coupling constant, while the second one is due to the uncertainty in the mass splittings between the SUSY particles.

This observation was considered as the first "evidence" for supersymmetry, especially since M_{SUSY} was found in the range preferred by the fine-tuning arguments.

II.3 Solution of the hierarchy problem

The appearance of two different scales $V \gg v$ in a GUT theory, namely, M_W and M_{GUT}, leads to a very serious problem which is called the *hierarchy problem*. There are two aspects of this problem.

The first one is the very existence of the hierarchy. To get the desired spontaneous symmetry breaking pattern, one needs

$$\begin{array}{ccccccc} m_H & \sim & v & \sim & 10^2 & \text{GeV} \\ m_\Sigma & \sim & V & \sim & 10^{16} & \text{GeV} \end{array} \qquad \frac{m_H}{m_\Sigma} \sim 10^{-14} \ll 1, \tag{II.8}$$

where H and Σ are the Higgs fields responsible for the spontaneous breaking of the $SU(2)$ and the GUT groups, respectively. The question arises of how to get so small number in a natural way.

The second aspect of the hierarchy problem is connected with the preservation of a given hierarchy. Even if we choose the hierarchy like eq.(II.8) the radiative corrections will destroy it! To see this, consider the radiative correction to the light Higgs mass given by the Feynman diagram shown in Fig.2. This correction proportional to the mass squared

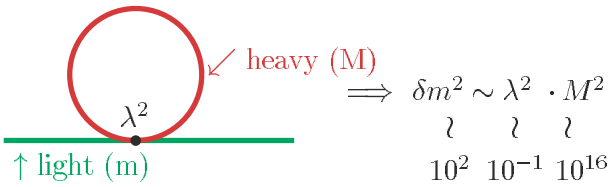

Figure 2. Radiative correction to the light Higgs boson mass

of the heavy particle, obviously, spoils the hierarchy if it is not cancelled. This very accurate cancellation with a precision $\sim 10^{-14}$ needs a fine tuning of the coupling constants.

The only known way of achieving this kind of cancellation of quadratic terms (also known as the cancellation of the quadratic divergencies) is supersymmetry. Moreover, SUSY automatically cancels quadratic corrections in all orders of PT. This is due to the contributions of superpartners of ordinary particles. The contribution from boson loops cancels those from the fermion ones because of an additional factor (-1) coming

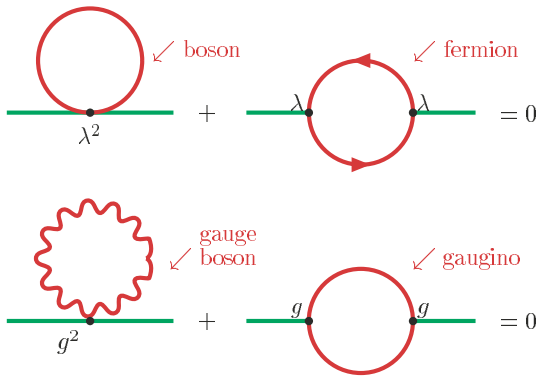

Figure 3. Cancellation of quadratic terms (divergencies)

from Fermi statistics, as shown in Fig.3. One can see here two types of contribution. The first line is the contribution of the heavy Higgs boson and its superpartner. The strength of interaction is given by the Yukawa coupling λ. The second line represents the gauge interaction proportional to the gauge coupling constant g with the contribution from the heavy gauge boson and heavy gaugino.

In both the cases the cancellation of quadratic terms takes place. This cancellation is true up to the SUSY breaking scale, M_{SUSY}, which should not be very large (≤ 1 TeV) to make the fine-tuning natural. Indeed, let us take the Higgs boson mass. Requiring for consistency of perturbation theory that the radiative corrections to the Higgs boson mass do not exceed the mass itself gives

$$\delta M_h^2 \sim g^2 M_{SUSY}^2 \sim M_h^2. \qquad \text{(II.9)}$$

So, if $M_h \sim 10^2$ GeV and $g \sim 10^{-1}$, one needs $M_{SUSY} \sim 10^3$ GeV in order that the relation (II.9) is valid. Thus, we again get the same rough estimate of $M_{SUSY} \sim 1$ TeV as from the gauge coupling unification above.

That is why it is usually said that supersymmetry solves the hierarchy problem. We show below how SUSY can also explain the origin of the hierarchy.

II.4 Astrophysics and Cosmology

The shining matter is not the only one in the Universe. Considerable amount consists of the so-called dark matter. The direct evidence for the presence of the dark matter are the rotation curves of galaxies (see Fig.4). To explain these curves one has to assume the existence of galactic halo

made of non-shining matter which takes part in gravitational interaction. According to the latest data [9] the matter content of the Universe is

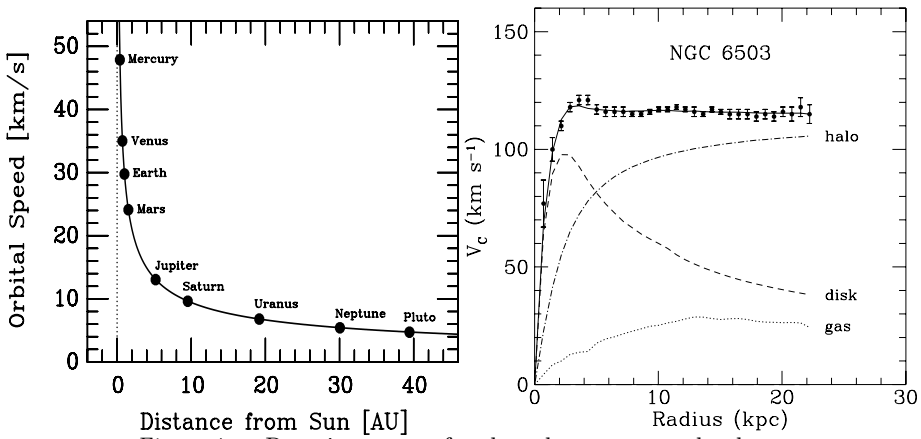

Figure 4. Rotation curves for the solar system and galaxy

the following:

$$\Omega h^2 = 1 \quad \Leftrightarrow \quad \rho = \rho_{crit}$$

$$\Omega_{vacuum} \approx 73\%, \quad \Omega_{DarkMatter} \approx 23\%, \quad \Omega_{Baryon} \approx 4\%$$

There are two possible types of the dark matter: the hot one, consisting of light relativistic particles and the cold one, consisting of massive weakly interacting particles (WIMPs). The hot dark matter might consist of neutrinos, however, this leads to problems with galaxy formation. As for the cold dark matter, it has no candidates within the SM. At the same time, SUSY provides an excellent candidate for the cold dark matter, namely neutralino, the lightest superparticle.

II.5 Beyond GUTs: superstring

Another motivation for supersymmetry follows from even more radical changes of basic ideas related to the ultimate goal of construction of consistent unified theory of everything. At the moment the only viable conception is the superstring theory [10]. In the superstring theory, strings are considered as fundamental objects, closed or open, and are nonlocal in nature. Ordinary particles are considered as string excitation modes. String interactions, which are local, generate proper interactions of usual particles, including gravitational ones.

To be consistent, the string theory should be conformal invariant in D-dimensional target space and have a stable vacuum. The first requirement is valid in classical theory but may be violated by quantum

anomalies. Cancellation of quantum anomalies takes place when space-time dimension of a target space equals to a critical one which is $D_c = 26$ for bosonic string and $D_c = 10$ for a fermionic one.

The second requirement is that the massless string excitations (the particles of the SM) are stable. This assumes the absence of tachyons, the states with imaginary mass, which can be guaranteed only in supersymmetric string theories!

III. Basics of supersymmetry

III.1 Algebra of SUSY

Combined with the usual Poincaré and internal symmetry algebra the Super-Poincaré Lie algebra contains additional SUSY generators Q^i_α and $\bar{Q}^i_{\dot{\alpha}}$ [3]

$$
\begin{aligned}
&[P_\mu, P_\nu] = 0,\\
&[P_\mu, M_{\rho\sigma}] = i(g_{\mu\rho}P_\sigma - g_{\mu\sigma}P_\rho),\\
&[M_{\mu\nu}, M_{\rho\sigma}] = i(g_{\nu\rho}M_{\mu\sigma} - g_{\nu\sigma}M_{\mu\rho} - g_{\mu\rho}M_{\nu\sigma} + g_{\mu\sigma}M_{\nu\rho}),\\
&[B_r, B_s] = iC^t_{rs}B_t,\\
&[B_r, P_\mu] = [B_r, M_{\mu\sigma}] = 0,\\
&[Q^i_\alpha, P_\mu] = [\bar{Q}^i_{\dot{\alpha}}, P_\mu] = 0,\\
&[Q^i_\alpha, M_{\mu\nu}] = \tfrac{1}{2}(\sigma_{\mu\nu})^\beta_\alpha Q^i_\beta, \quad [\bar{Q}^i_{\dot{\alpha}}, M_{\mu\nu}] = -\tfrac{1}{2}\bar{Q}^i_{\dot{\beta}}(\bar{\sigma}_{\mu\nu})^{\dot{\beta}}_{\dot{\alpha}},\\
&[Q^i_\alpha, B_r] = (b_r)^i_j Q^j_\alpha, \quad [\bar{Q}^i_{\dot{\alpha}}, B_r] = -\bar{Q}^j_{\dot{\alpha}}(b_r)^i_j,\\
&\{Q^i_\alpha, \bar{Q}^j_{\dot{\beta}}\} = 2\delta^{ij}(\sigma^\mu)_{\alpha\dot{\beta}}P_\mu,\\
&\{Q^i_\alpha, Q^j_\beta\} = 2\epsilon_{\alpha\beta}Z^{ij}, \quad Z_{ij} = a^r_{ij}b_r, \quad Z^{ij} = Z^+_{ij},\\
&\{\bar{Q}^i_{\dot{\alpha}}, \bar{Q}^j_{\dot{\beta}}\} = -2\epsilon_{\dot{\alpha}\dot{\beta}}Z^{ij}, \quad [Z_{ij}, anything] = 0,\\
&\alpha, \dot{\alpha} = 1, 2 \quad i, j = 1, 2, \ldots, N.
\end{aligned}
\tag{III.1}
$$

Here P_μ and $M_{\mu\nu}$ are four-momentum and angular momentum operators, respectively, B_r are the internal symmetry generators, Q^i and $\bar{Q}^i$ are the spinorial SUSY generators and Z_{ij} are the so-called central charges; $\alpha, \dot{\alpha}, \beta, \dot{\beta}$ are the spinorial indices. In the simplest case one has one spinor generator Q_α (and the conjugated one $\bar{Q}_{\dot{\alpha}}$) that corresponds to an ordinary or N=1 supersymmetry. When $N > 1$ one has an extended supersymmetry.

A natural question arises: how many SUSY generators are possible, i.e. what is the value of N? To answer this question, consider massless states. Let us start with the ground state labeled by energy and helicity, i.e. projection of a spin on the direction of momenta, and let it be annihilated by Q_i

$$\text{Vacuum} = |E, \lambda>, \qquad Q_i|E, \lambda> = 0.$$

Then one and more particle states can be constructed with the help of a creation operators as

State	Expression	# of States
vacuum	$\|E,\lambda>$	1
$1-$particle state	$\bar{Q}_i\|E,\lambda>=\|E,\lambda+1/2>_i$	$\begin{pmatrix} N \\ 1 \end{pmatrix} = N$
$2-$particle state	$\bar{Q}_i\bar{Q}_j\|E,\lambda>=\|E,\lambda+1>_{ij}$	$\begin{pmatrix} N \\ 2 \end{pmatrix} = \frac{N(N-1)}{2}$
...	...	...
$N-$particle state	$\bar{Q}_1\bar{Q}_2...\bar{Q}_N\|E,\lambda>=\|E,\lambda+\frac{N}{2}>$	$\begin{pmatrix} N \\ N \end{pmatrix} = 1$

Total # of states: $\sum_{k=0}^{N} \begin{pmatrix} N \\ k \end{pmatrix} = 2^N = 2^{N-1}$ bosons $+ 2^{N-1}$ fermions.

The energy E is not changed, since according to (III.1) the operators $\bar{Q}_i$ commute with the Hamiltonian.

Thus, one has a sequence of bosonic and fermionic states and the total number of bosons equals that of fermions. This is a generic property of any supersymmetric theory. However, in CPT invariant theories the number of states is doubled, since CPT transformation changes the sign of helicity. Hence, in CPT invariant theories, one has to add the states with opposite helicity to the above mentioned ones.

Consider some examples. Let us take $N = 1$ and $\lambda = 0$. Then one has the following set of states:

$$N=1 \quad \lambda=0 \qquad \begin{array}{lccc} \text{helicity} & 0 & 1/2 \\ \# \text{ of states} & 1 & 1 \end{array} \overset{CPT}{\Longrightarrow} \begin{array}{lccc} \text{helicity} & 0 & -1/2 \\ \# \text{ of states} & 1 & 1 \end{array}$$

Hence, a complete $N = 1$ multiplet is

$$N=1 \qquad \begin{array}{lccc} \text{helicity} & -1/2 & 0 & 1/2 \\ \# \text{ of states} & 1 & 2 & 1 \end{array}$$

which contains one complex scalar and one spinor with two helicity states.

This is an example of the so-called self-conjugated multiplet. There are also self-conjugated multiplets with $N > 1$ corresponding to extended supersymmetry. Two particular examples are the $N = 4$ super Yang-Mills multiplet and the $N = 8$ supergravity multiplet

$$N=4 \quad \begin{array}{lcccccc} \text{SUSY YM} & \text{helicity} & -1 & -1/2 & 0 & 1/2 & 1 \\ \lambda=-1 & \# \text{ of states} & 1 & 4 & 6 & 4 & 1 \end{array}$$

$N=8$	SUGRA	-2	$-3/2$	-1	$-1/2$	0	$1/2$	1	$3/2$	2
	$\lambda=-2$	1	8	28	56	70	56	28	8	1

One can see that the multiplets of extended supersymmetry are very rich and contain a vast number of particles.

The constraint on the number of SUSY generators comes from a requirement of consistency of the corresponding QFT. The number of supersymmetries and the maximal spin of the particle in the multiplet are related by

$$N \leq 4S,$$

where S is the maximal spin. Since the theories with spin greater than 1 are non-renormalizable and the theories with spin greater than 5/2 have no consistent coupling to gravity, this imposes a constraint on the number of SUSY generators

$$\begin{array}{ll} N \leq 4 & \text{for renormalizable theories (YM),} \\ N \leq 8 & \text{for (super)gravity.} \end{array}$$

In what follows, we shall consider simple supersymmetry, or $N = 1$ supersymmetry, contrary to extended supersymmetries with $N > 1$. In this case, one has two types of supermultiplets: the so-called chiral multiplet with $\lambda = 0$, which contains two physical states (ϕ, ψ) with spin 0 and 1/2, respectively, and the vector multiplet with $\lambda = 1/2$, which also contains two physical states (λ, A_μ) with spin 1/2 and 1, respectively.

III.2 Superspace and superfields

An elegant formulation of supersymmetry transformations and invariants can be achieved in the framework of superspace [4]. Superspace differs from the ordinary Euclidean (Minkowski) space by adding of two new coordinates, θ_α and $\bar{\theta}_{\dot{\alpha}}$, which are Grassmannian, i.e. anticommuting, variables

$$\{\theta_\alpha, \theta_\beta\} = 0, \quad \{\bar{\theta}_{\dot{\alpha}}, \bar{\theta}_{\dot{\beta}}\} = 0, \quad \theta_\alpha^2 = 0, \quad \bar{\theta}_{\dot{\alpha}}^2 = 0, \quad \alpha, \beta, \dot{\alpha}, \dot{\beta} = 1, 2.$$

Thus, we go from space to superspace

$$\begin{array}{ccc} \mathit{Space} & \Rightarrow & \mathit{Superspace} \\ x_\mu & & x_\mu, \theta_\alpha, \bar{\theta}_{\dot{\alpha}} \end{array}$$

A SUSY group element can be constructed in superspace in the same way as an ordinary translation in the usual space

$$G(x, \theta, \bar{\theta}) = e^{i(-x^\mu P_\mu + \theta Q + \bar{\theta}\bar{Q})}. \qquad \text{(III.2)}$$

It leads to a supertranslation in superspace

$$\begin{array}{rcl} x_\mu & \to & x_\mu + i\theta\sigma_\mu\bar{\varepsilon} - i\varepsilon\sigma_\mu\bar{\theta}, \\ \theta & \to & \theta + \varepsilon, \quad \bar{\theta} \to \bar{\theta} + \bar{\varepsilon}, \end{array} \tag{III.3}$$

where ε and $\bar{\varepsilon}$ are Grassmannian transformation parameters. From eq.(III.3) one can easily obtain the representation for the supercharges (III.1) acting on the superspace

$$Q_\alpha = \frac{\partial}{\partial\theta_\alpha} - i\sigma^\mu_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\partial_\mu, \quad \bar{Q}_{\dot{\alpha}} = -\frac{\partial}{\partial\bar{\theta}_{\dot{\alpha}}} + i\theta_\alpha\sigma^\mu_{\alpha\dot{\alpha}}\partial_\mu. \tag{III.4}$$

To define the fields on a superspace, consider representations of the Super-Poincaré group (III.1) [3]. The simplest one is a scalar superfield $F(x,\theta,\bar{\theta})$ which is SUSY invariant. Its Taylor expansion in θ and $\bar{\theta}$ has only several terms due to the nilpotent character of Grassmannian parameters. However, this superfield is a reducible representation of SUSY. To get an irreducible one, we define a *chiral* superfield which obeys the equation

$$\bar{D}_{\dot{\alpha}}F = 0, \qquad \text{where } \bar{D}_{\dot{\alpha}} = -\frac{\partial}{\partial\bar{\theta}\,\dot{\alpha}} - i(\theta\sigma^\mu)_{\dot{\alpha}}\partial_\mu \tag{III.5}$$

is a superspace covariant derivative.

For the chiral superfield Grassmannian Taylor expansion looks like ($y = x + i\theta\sigma\bar{\theta}$)

$$\begin{aligned} \Phi(y,\theta) &= A(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y) \\ &= A(x) + i\theta\sigma^\mu\bar{\theta}\partial_\mu A(x) + \frac{1}{4}\theta\theta\bar{\theta}\bar{\theta}\Box A(x) \\ &+ \sqrt{2}\theta\psi(x) - \frac{i}{\sqrt{2}}\theta\theta\partial_\mu\psi(x)\sigma^\mu\bar{\theta} + \theta\theta F(x). \end{aligned} \tag{III.6}$$

The coefficients are ordinary functions of x being the usual fields. They are called the *components* of a superfield. In eq.(III.6) one has 2 bosonic (complex scalar field A) and 2 fermionic (Weyl spinor field ψ) degrees of freedom. The component fields A and ψ are called the *superpartners*. The field F is an *auxiliary* field, it has the "wrong" dimension and has no physical meaning. It is needed to close the algebra (III.1). One can get rid of the auxiliary fields with the help of equations of motion.

Thus, a superfield contains an equal number of bosonic and fermionic degrees of freedom. Under SUSY transformation they convert into one another

$$\begin{aligned} \delta_\varepsilon A &= \sqrt{2}\varepsilon\psi, \\ \delta_\varepsilon\psi &= i\sqrt{2}\sigma^\mu\bar{\varepsilon}\partial_\mu A + \sqrt{2}\varepsilon F, \\ \delta_\varepsilon F &= i\sqrt{2}\bar{\varepsilon}\sigma^\mu\partial_\mu\psi. \end{aligned} \tag{III.7}$$

Notice that the variation of the F-component is a total derivative, i.e. it vanishes when integrated over the space-time.

One can also construct an antichiral superfield Φ^+ obeying the equation

$$D_\alpha \Phi^+ = 0, \quad \text{with} \quad D_\alpha = \frac{\partial}{\partial \theta_\alpha} + i(\sigma^\mu \bar{\theta})_\alpha \partial_\mu.$$

The product of chiral (antichiral) superfields Φ^2, Φ^3, etc is also a chiral (antichiral) superfield, while the product of chiral and antichiral ones $\Phi^+\Phi$ is a general superfield.

For any arbitrary function of chiral superfields one has

$$\begin{aligned} \mathcal{W}(\Phi_i) &= \mathcal{W}(A_i + \sqrt{2}\theta\psi_i + \theta\theta F) \\ &= \mathcal{W}(A_i) + \frac{\partial \mathcal{W}}{\partial A_i}\sqrt{2}\theta\psi_i + \theta\theta\left(\frac{\partial \mathcal{W}}{\partial A_i}F_i - \frac{1}{2}\frac{\partial^2 \mathcal{W}}{\partial A_i \partial A_j}\psi_i\psi_j\right). \end{aligned} \tag{III.8}$$

The $\mathcal{W}$ is usually referred to as a superpotential which replaces the usual potential for the scalar fields.

To construct the gauge invariant interactions, one needs a real vector superfield $V = V^+$. It is not chiral but rather a general superfield with the following Grassmannian expansion:

$$\begin{aligned} V(x,\theta,\bar{\theta}) &= C(x) + i\theta\chi(x) - i\bar{\theta}\bar{\chi}(x) + \frac{i}{2}\theta\theta[M(x) + iN(x)] \\ &- \frac{i}{2}\bar{\theta}\bar{\theta}[M(x) - iN(x)] - \theta\sigma^\mu\bar{\theta}v_\mu(x) + i\theta\theta\bar{\theta}[\lambda(x) + \frac{i}{2}\bar{\sigma}^\mu\partial_\mu\chi(x)] \\ &- i\bar{\theta}\bar{\theta}\theta[\lambda + \frac{i}{2}\sigma^\mu\partial_\mu\bar{\chi}(x)] + \frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}[D(x) + \frac{1}{2}\Box C(x)]. \end{aligned} \tag{III.9}$$

The physical degrees of freedom corresponding to a real vector superfield V are the vector gauge field v_μ and the Majorana spinor field λ. All other components are unphysical and can be eliminated. Indeed, under the Abelian (super)gauge transformation the superfield V is transformed as

$$V \quad \to \quad V + \Phi + \Phi^+,$$

where Φ and Φ^+ are some chiral superfields. In components it looks like

$$\begin{aligned} C &\to C + A + A^*, \\ \chi &\to \chi - i\sqrt{2}\psi, \\ M + iN &\to M + iN - 2iF, \\ v_\mu &\to v_\mu - i\partial_\mu(A - A^*), \\ \lambda &\to \lambda, \\ D &\to D, \end{aligned} \tag{III.10}$$

and corresponds to ordinary gauge transformations for physical components. According to eq.(III.10), one can choose a gauge (the Wess-Zumino gauge) where $C = \chi = M = N = 0$, leaving one with only physical degrees of freedom except for the auxiliary field D. In this gauge

$$\begin{aligned} V &= -\theta\sigma^\mu\bar{\theta}v_\mu(x) + i\theta\theta\bar{\theta}\bar{\lambda}(x) - i\bar{\theta}\bar{\theta}\theta\lambda(x) + \frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}D(x), \\ V^2 &= -\frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}v_\mu(x)v^\mu(x), \\ V^3 &= 0, \quad etc. \end{aligned} \tag{III.11}$$

One can define also a field strength tensor (as analog of $F_{\mu\nu}$ in gauge theories)

$$W_\alpha = -\frac{1}{4}\bar{D}^2 e^V D_\alpha e^{-V}, \quad \bar{W}_{\dot{\alpha}} = -\frac{1}{4}D^2 e^V \bar{D}_\alpha e^{-V}, \tag{III.12}$$

which is a polynomial in the Wess-Zumino gauge. (Here Ds are the supercovariant derivatives.)

The strength tensor is a chiral superfield

$$\bar{D}_{\dot{\beta}}W_\alpha = 0, \quad D_\beta\bar{W}_{\dot{\alpha}} = 0.$$

In the Wess-Zumino gauge it is a polynomial over component fields:

$$W_\alpha = T^a\left(-i\lambda^a_\alpha + \theta_\alpha D^a - \frac{i}{2}(\sigma^\mu\bar{\sigma}^\nu\theta)_\alpha F^a_{\mu\nu} + \theta^2(\sigma^\mu D_\mu\bar{\lambda}^a)_\alpha\right), \tag{III.13}$$

where

$$F^a_{\mu\nu} = \partial_\mu v^a_\nu - \partial_\nu v^a_\mu + f^{abc}v^b_\mu v^c_\nu, \quad D_\mu\bar{\lambda}^a = \partial\bar{\lambda}^a + f^{abc}v^b_\mu\bar{\lambda}^c.$$

In Abelian case eqs.(III.12) are simplified and take form

$$W_\alpha = -\frac{1}{4}\bar{D}^2 D_\alpha V, \quad \bar{W}_{\dot{\alpha}} = -\frac{1}{4}D^2\bar{D}_\alpha V.$$

III.3 Construction of SUSY Lagrangians

Let us start with the Lagrangian which has no local gauge invariance. In the superfield notation SUSY invariant Lagrangians are the polynomials of superfields. Having in mind that for component fields one should have ordinary terms and the above mentioned property of SUSY invariance of the highest dimension components of a superfield, the general SUSY invariant Lagrangian has the form

$$\mathcal{L} = \Phi^+_i\Phi_i|_{\theta\theta\bar{\theta}\bar{\theta}} + [(\lambda_i\Phi_i + \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}g_{ijk}\Phi_i\Phi_j\Phi_k)|_{\theta\theta} + h.c.]. \tag{III.14}$$

Hereafter the vertical line means the corresponding term of a Taylor expansion.

The first term is a kinetic term. It contains both the chiral and antichiral superfields Φ_i and Φ_i^+, respectively, and is a function of Grassmannian parameters θ and $\bar{\theta}$. Being expanded over θ and $\bar{\theta}$ it leads to the usual kinetic terms for the corresponding component fields.

The terms in the bracket form the superpotential. It is composed of the chiral fields only (plus the hermitian conjugated counterpart composed of antichiral superfields) and is a chiral superfield. Since the products of a chiral superfield and antichiral one produce a general superfield, they are not allowed in a superpotential. The last coefficient of its expansion over the parameter θ is supersymmetrically invariant and gives the usual potential after getting rid of the auxiliary fields.

The Lagrangian (III.14) can be written in a much more elegant way in superspace. The same way as an ordinary action is an integral over space-time of Lagrangian density, in supersymmetric case the action is an integral over the superspace. The space-time Lagrangian density then is [3, 4]

$$\mathcal{L} = \int d^2\theta d^2\bar{\theta}\ \Phi_i^+\Phi_i + \int d^2\theta\ [\lambda_i\Phi_i + \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}y_{ijk}\Phi_i\Phi_j\Phi_k] + h.c. \tag{III.15}$$

where the first part is a kinetic term and the second one is a superpotential $\mathcal{W}$. Here instead of taking the proper components we use integration over the superspace according to the rules of Grassmannian integration [11]

$$\int\ d\theta_\alpha = 0, \qquad \int \theta_\alpha\ d\theta_\beta = \delta_{\alpha\beta}.$$

Performing explicit integration over the Grassmannian parameters, we get from eq.(III.15)

$$\begin{aligned} \mathcal{L} &= i\partial_\mu\bar{\psi}_i\bar{\sigma}^\mu\psi_i + A_i^*\Box A_i + F_i^*F_i \qquad \text{(III.16)} \\ &+ [\lambda_iF_i + m_{ij}(A_iF_j - \frac{1}{2}\psi_i\psi_j) + y_{ijk}(A_iA_jF_k - \psi_i\psi_jA_k) + h.c.]. \end{aligned}$$

The last two terms are the interaction ones. To obtain a familiar form of the Lagrangian, we have to solve the constraints

$$\frac{\partial\mathcal{L}}{\partial F_k^*} = F_k + \lambda_k^* + m_{ik}^*A_i^* + y_{ijk}^*A_i^*A_j^* = 0, \qquad \text{(III.17)}$$

$$\frac{\partial\mathcal{L}}{\partial F_k} = F_k^* + \lambda_k + m_{ik}A_i + y_{ijk}A_iA_j = 0. \qquad \text{(III.18)}$$

Expressing the auxiliary fields F and F^* from these equations, one finally gets

$$\begin{aligned}\mathcal{L} &= i\partial_\mu\bar{\psi}_i\bar{\sigma}^\mu\psi_i + A_i^*\Box A_i - \frac{1}{2}m_{ij}\psi_i\psi_j - \frac{1}{2}m_{ij}^*\bar{\psi}_i\bar{\psi}_j \\ &\quad -y_{ijk}\psi_i\psi_j A_k - y_{ijk}^*\bar{\psi}_i\bar{\psi}_j A_k^* - V(A_i, A_j), \qquad \text{(III.19)}\end{aligned}$$

where the scalar potential $V = F_k^* F_k$. We will return to the discussion of the form of the scalar potential in SUSY theories later.

Consider now the gauge invariant SUSY Lagrangians. They should contain gauge invariant interaction of the matter fields with the gauge ones and the kinetic term and the self-interaction of the gauge fields.

Let us start with the gauge field kinetic terms. In the Wess-Zumino gauge one has

$$W^\alpha W_\alpha|_{\theta\theta} = -2i\lambda\sigma^\mu D_\mu\bar{\lambda} - \frac{1}{2}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}D^2 + i\frac{1}{4}F^{\mu\nu}F^{\rho\sigma}\epsilon_{\mu\nu\rho\sigma}, \quad \text{(III.20)}$$

where $D_\mu = \partial_\mu + ig[v_\mu,]$ is the usual covariant derivative and the last, the so-called topological θ term, is the total derivative.

The gauge invariant Lagrangian now has a familiar form

$$\begin{aligned}\mathcal{L} &= \frac{1}{4}\int d^2\theta\ W^\alpha W_\alpha + \frac{1}{4}\int d^2\bar{\theta}\ \bar{W}^{\dot{\alpha}}\bar{W}_{\dot{\alpha}} \\ &= \frac{1}{2}D^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\lambda\sigma^\mu D_\mu\bar{\lambda}. \qquad \text{(III.21)}\end{aligned}$$

To obtain a gauge-invariant interaction with matter chiral superfields, consider their gauge transformation (Abelian)

$$\Phi \rightarrow e^{-ig\Lambda}\Phi, \quad \Phi^+ \rightarrow \Phi^+ e^{ig\Lambda^+}, \quad V \rightarrow V + i(\Lambda - \Lambda^+),$$

where Λ is a gauge parameter (chiral superfield).

It is clear now how to construct both the SUSY and gauge invariant kinetic term (compare with the covariant derivative in a usual gauge theory)

$$\Phi_i^+\Phi_i|_{\theta\theta\bar{\theta}\bar{\theta}} \Rightarrow \Phi_i^+ e^{gV}\Phi_i|_{\theta\theta\bar{\theta}\bar{\theta}} \qquad \text{(III.22)}$$

A complete SUSY and gauge invariant Lagrangian then looks like

$$\begin{aligned}\mathcal{L}_{inv} &= \frac{1}{4}\int d^2\theta\ W^\alpha W_\alpha + \frac{1}{4}\int d^2\bar{\theta}\ \bar{W}^{\dot{\alpha}}\bar{W}_{\dot{\alpha}} + \int d^2\theta d^2\bar{\theta}\ \Phi_i^+ e^{gV}\Phi_i \\ &+ \int d^2\theta\ (\frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}y_{ijk}\Phi_i\Phi_j\Phi_k) + h.c. \qquad \text{(III.23)}\end{aligned}$$

The non-Abelian generalization is straightforward

$$\begin{aligned}\mathcal{L}_{SUSY\ YM} &= \frac{1}{4}\int d^2\theta\ Tr(W^\alpha W_\alpha) + \frac{1}{4}\int d^2\bar\theta\ Tr(\bar W^\alpha \bar W_\alpha) \qquad \text{(III.24)}\\ &+ \int d^2\theta d^2\bar\theta\ \bar\Phi_{ia}(e^{gV})^a_b\Phi^b_i + \int d^2\theta\ \mathcal{W}(\Phi_i) + \int d^2\bar\theta\ \bar{\mathcal{W}}(\bar\Phi_i),\end{aligned}$$

where $\mathcal{W}$ is a superpotential, which should be invariant under the group of symmetry of a particular model.

In terms of component fields the above Lagrangian takes the form

$$\begin{aligned}\mathcal{L}_{SUSY\ YM} &= -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - i\lambda^a\sigma^\mu D_\mu\bar\lambda^a + \frac{1}{2}D^aD^a \qquad \text{(III.25)}\\ &+ (\partial_\mu A_i - igv^a_\mu T^a A_i)^\dagger(\partial_\mu A_i - igv^a_\mu T^a A_i) - i\bar\psi_i\bar\sigma^\mu(\partial_\mu\psi_i - igv^a_\mu T^a\psi_i)\\ &- D^aA^\dagger_i T^a A_i - i\sqrt{2}A^\dagger_i T^a\lambda^a\psi_i + i\sqrt{2}\bar\psi_i T^a A_i\bar\lambda^a + F^\dagger_i F_i\\ &+ \frac{\partial\mathcal{W}}{\partial A_i}F_i + \frac{\partial\bar{\mathcal{W}}}{\partial A^\dagger_i}F^\dagger_i - \frac{1}{2}\frac{\partial^2\mathcal{W}}{\partial A_i\partial A_j}\psi_i\psi_j - \frac{1}{2}\frac{\partial^2\bar{\mathcal{W}}}{\partial A^\dagger_i\partial A^\dagger_j}\bar\psi_i\bar\psi_j.\end{aligned}$$

Integrating out the auxiliary fields D^a and F_i, one reproduces the usual Lagrangian.

III.4 The scalar potential

Contrary to the SM, where the scalar potential is arbitrary and is defined only by the requirement of the gauge invariance, in supersymmetric theories it is completely defined by the superpotential. It consists of the contributions from the D-terms and F-terms. The kinetic energy of the gauge fields (recall eq.(III.21) yields the $1/2D^aD^a$ term, and the matter-gauge interaction (recall eq.(III.23) yields the $gD^aT^a_{ij}A^*_iA_j$ one. Together they give

$$\mathcal{L}_D = \frac{1}{2}D^aD^a + gD^aT^a_{ij}A^*_iA_j. \qquad \text{(III.26)}$$

The equation of motion reads

$$D^a = -gT^a_{ij}A^*_iA_j. \qquad \text{(III.27)}$$

Substituting it back into eq.(III.26) yields the D-term part of the potential

$$\mathcal{L}_D = -\frac{1}{2}D^aD^a \quad \Longrightarrow V_D = \frac{1}{2}D^aD^a, \qquad \text{(III.28)}$$

where D is given by eq.(III.27).

The F-term contribution can be derived from the matter field self-interaction eq.(III.16). For a general type superpotential W one has

$$\mathcal{L}_F = F_i^* F_i + (\frac{\partial W}{\partial A_i} F_i + h.c.). \tag{III.29}$$

Using the equations of motion for the auxiliary field F_i

$$F_i^* = -\frac{\partial W}{\partial A_i} \tag{III.30}$$

yields

$$\mathcal{L}_F = -F_i^* F_i \quad \Longrightarrow V_F = F_i^* F_i, \tag{III.31}$$

where F is given by eq.(III.30). The full potential is the sum of the two contributions

$$V = V_D + V_F. \tag{III.32}$$

Thus, the form of the Lagrangian is practically fixed by symmetry requirements. The only freedom is the field content, the value of the gauge coupling g, Yukawa couplings y_{ijk} and the masses. Because of the renormalizability constraint $V \leq A^4$ the superpotential should be limited by $\mathcal{W} \leq \Phi^3$ as in eq.(III.15). All members of a supermultiplet have the same masses, i.e. bosons and fermions are degenerate in masses. This property of SUSY theories contradicts the phenomenology and requires supersymmetry breaking.

III.5 Spontaneous breaking of SUSY

Since supersymmetric algebra leads to mass degeneracy in a supermultiplet, it should be broken to explain the absence of superpartners at present energies. There are several ways of supersymmetry breaking. It can be broken either explicitly or spontaneously. Performing SUSY breaking one has to be careful not to spoil the cancellation of quadratic divergencies which allows one to solve the hierarchy problem. This is achieved by spontaneous breaking of SUSY.

Apart from non-supersymmetric theories in SUSY models the energy is always nonnegative definite. Indeed, according to quantum mechanics

$$E =< 0|\ H\ |0 >$$

and due to SUSY algebra eq.(III.1) $\{Q_\alpha, \bar{Q}_{\dot{\beta}}\} = 2(\sigma^\mu)_{\alpha\dot{\beta}} P_\mu$, taking into account that $tr(\sigma^\mu P_\mu) = 2P_0$, one gets

$$E = \frac{1}{4} \sum_{\alpha=1,2} < 0|\{Q_\alpha, \bar{Q}_\alpha\}|0 >= \frac{1}{4} \sum_\alpha |Q_\alpha|0 > |^2 \geq 0.$$

Hence

$$E =< 0|\ H\ |0 >\neq 0 \quad if\ and\ only\ if \quad Q_\alpha|0 >\neq 0.$$

Therefore, supersymmetry is spontaneously broken, i.e. vacuum is not invariant ($Q_\alpha|0 >\neq 0$), *if and only if* the minimum of the potential is positive (*i.e.* $E > 0$) .

Spontaneous breaking of supersymmetry is achieved in the same way as the electroweak symmetry breaking. One introduces the field whose vacuum expectation value is nonzero and breaks the symmetry. However, due to a special character of SUSY, this should be a superfield whose auxiliary F and D components acquire nonzero v.e.v.'s. Thus, among possible spontaneous SUSY breaking mechanisms one distinguishes the F and D ones.

i) Fayet-Iliopoulos (D-term) mechanism [12].
In this case the, the linear D-term is added to the Lagrangian

$$\Delta\mathcal{L} = \xi V|_{\theta\theta\bar\theta\bar\theta} = \xi \int d^4\theta\ V. \qquad \text{(III.33)}$$

It is gauge and SUSY invariant by itself; however, it may lead to spontaneous breaking of both of them depending on the value of ξ. We show in Fig.5a the sample spectrum for two chiral matter multiplets. The

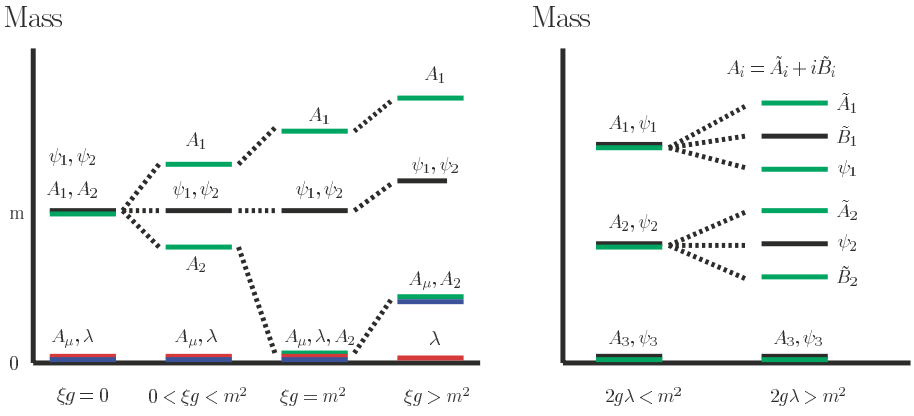

Figure 5. Spectrum of spontaneously broken SUSY theories

drawback of this mechanism is the necessity of $U(1)$ gauge invariance. It can be used in SUSY generalizations of the SM but not in GUTs.

The mass spectrum also causes some troubles since the following sum rule is always valid

$$\sum_{bosonic\ states} m_i^2 = \sum_{fermionic\ states} m_i^2, \qquad \text{(III.34)}$$

which is bad for phenomenology.

ii) O'Raifeartaigh (F-term) mechanism [13].
In this case, several chiral fields are needed and the superpotential should be chosen in a way that trivial zero v.e.v.s for the auxiliary F-fields be absent. For instance, choosing the superpotential to be

$$\mathcal{W}(\Phi) = \lambda\Phi_3 + m\Phi_1\Phi_2 + g\Phi_3\Phi_1^2,$$

one gets the equations for the auxiliary fields

$$\begin{aligned} F_1^* &= mA_2 + 2gA_1A_3, \\ F_2^* &= mA_1, \\ F_3^* &= \lambda + gA_1^2, \end{aligned}$$

which have no solutions with $< F_i >= 0$ and SUSY is spontaneously broken. The sample spectrum is shown in Fig.5b.

The drawbacks of this mechanism is a lot of arbitrariness in the choice of potential. The sum rule (III.34) is also valid here.

Unfortunately, none of these mechanisms explicitly works in SUSY generalizations of the SM. None of the fields of the SM can develop nonzero v.e.v.s for their F or D components without breaking $SU(3)$ or $U(1)$ gauge invariance since they are not singlets with respect to these groups. This requires the presence of extra sources of spontaneous SUSY breaking, which we consider below. They are based, however, on the same F and D mechanisms.

IV. SUSY generalization of the Standard Model. The MSSM

As has been already mentioned, in SUSY theories the number of bosonic degrees of freedom equals that of fermionic. At the same time, in the SM one has 28 bosonic and 90 fermionic degrees of freedom (with massless neutrino, otherwise 96). So the SM is to a great extent non-supersymmetric. Trying to add some new particles to supersymmetrize the SM, one should take into account the following observations:

- There are no fermions with quantum numbers of the gauge bosons;
- Higgs fields have nonzero v.e.v.s; hence they cannot be superpartners of quarks and leptons since this would induce spontaneous violation of baryon and lepton numbers;

• One needs at least two complex chiral Higgs multiplets to give masses to Up and Down quarks.

The latter is due to the form of a superpotential and chirality of matter superfields. Indeed, the superpotential should be invariant under the $SU(3) \times SU(2) \times U(1)$ gauge group. If one looks at the Yukawa interaction in the Standard Model, one finds that it is indeed $U(1)$ invariant since the sum of hypercharges in each vertex equals zero. In the last term this is achieved by taking the conjugated Higgs doublet $\tilde{H} = i\tau_2 H^\dagger$ instead of H. However, in SUSY H is a chiral superfield and hence a superpotential, which is constructed out of chiral fields, can contain only H but not $\tilde{H}$ which is an antichiral superfield.

Another reason for the second Higgs doublet is related to chiral anomalies. It is known that chiral anomalies spoil the gauge invariance and, hence, the renormalizability of the theory. They are canceled in the SM between quarks and leptons in each generation. However, if one introduces a chiral Higgs superfield, it contains higgsinos, which are chiral fermions, and contain anomalies. To cancel them one has to add the second Higgs doublet with the opposite hypercharge. Therefore, the Higgs sector in SUSY models is inevitably enlarged, it contains an even number of doublets.

Conclusion: In SUSY models supersymmetry associates *known* bosons with *new* fermions and *known* fermions with *new* bosons.

IV.1 The field content

Consider the particle content of the Minimal Supersymmetric Standard Model [14]. According to the previous discussion, in the minimal version we double the number of particles (introducing a superpartner to each particle) and add another Higgs doublet (with its superpartner).

Thus, the characteristic feature of any supersymmetric generalization of the SM is the presence of superpartners (see Fig.6) [15]. If supersymmetry is exact, superpartners of ordinary particles should have the same masses and have to be observed. The absence of them at modern energies is believed to be explained by the fact that their masses are very heavy, that means that supersymmetry should be broken. Hence, if the energy of accelerators is high enough, the superpartners will be created.

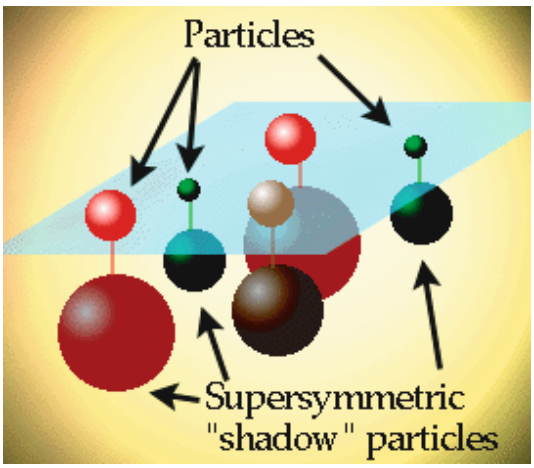

Figure 6. The shadow world of SUSY particles

The particle content of the MSSM then appears as

Particle Content of the MSSM

Superfield	Bosons		Fermions		$SU(3)$	$SU(2$	$U_Y(1)$
Gauge							
$\mathbf{G^a}$	gluon	g^a	gluino	$\tilde{g}^a$	8	0	0
$\mathbf{V^k}$	Weak	$W^k \ (W^{\pm}, Z)$	wino, zino	$\tilde{w}^k \ (\tilde{w}^{\pm}, \tilde{z})$	1	3	0
$\mathbf{V'}$	Hypercharge	$B \ (\gamma)$	bino	$\tilde{b}(\tilde{\gamma})$	1	1	0
Matter							
$\mathbf{L_i}$	sleptons	$\tilde{L}_i = (\tilde{\nu}, \tilde{e})_L$	leptons	$L_i = (\nu, e)_L$	1	2	-1
$\mathbf{E_i}$		$\tilde{E}_i = \tilde{e}_R$		$E_i = e_R$	1	1	2
$\mathbf{Q_i}$	squarks	$\tilde{Q}_i = (\tilde{u}, \tilde{d})_L$	quarks	$Q_i = (u, d)_L$	3	2	1/3
$\mathbf{U_i}$		$\tilde{U}_i = \tilde{u}_R$		$U_i = u^c_R$	3^*	1	$-4/3$
$\mathbf{D_i}$		$\tilde{D}_i = \tilde{d}_R$		$D_i = d^c_R$	3^*	1	2/3
Higgs							
$\mathbf{H_1}$	Higgses	H_1	higgsinos	$\tilde{H}_1$	1	2	-1
$\mathbf{H_2}$		H_2		$\tilde{H}_2$	1	2	1

Hereafter, tilde denotes a superpartner of an ordinary particle.

The presence of an extra Higgs doublet in SUSY model is a novel feature of the theory. In the MSSM one has two doublets with the quantum numbers (1,2,-1) and (1,2,1), respectively:

$$H_1 = \begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix} = \begin{pmatrix} v_1 + \frac{S_1 + iP_1}{\sqrt{2}} \\ H_1^- \end{pmatrix}, \; H_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} = \begin{pmatrix} H_2^+ \\ v_2 + \frac{S_2 + iP_2}{\sqrt{2}} \end{pmatrix},$$

where v_i are the vacuum expectation values of the neutral components.

Hence, one has 8=4+4=5+3 degrees of freedom. As in the case of the SM, 3 degrees of freedom can be gauged away, and one is left with 5 physical states compared to 1 in the SM. Thus, in the MSSM, as actually in any of two Higgs doublet models, one has five physical Higgs bosons: two CP-even neutral, one CP-odd neutral and two charged. We consider the mass eigenstates below.

IV.2 Lagrangian of the MSSM

The Lagrangian of the MSSM consists of two parts; the first part is SUSY generalization of the Standard Model, while the second one represents the SUSY breaking as mentioned above.

$$\mathcal{L} = \mathcal{L}_{SUSY} + \mathcal{L}_{Breaking}, \tag{IV.1}$$

where

$$\mathcal{L}_{SUSY} = \mathcal{L}_{Gauge} + \mathcal{L}_{Yukawa} \tag{IV.2}$$

and

$$\begin{aligned}\mathcal{L}_{Gauge} &= \sum_{SU(3),SU(2),U(1)} \frac{1}{4}\left(\int d^2\theta\ TrW^\alpha W_\alpha + \int d^2\bar\theta\ Tr\bar W^{\dot\alpha}\bar W_{\dot\alpha}\right) \\ &\quad + \sum_{Matter}\int d^2\theta d^2\bar\theta\ \Phi_i^\dagger e^{g_3\hat V_3 + g_2\hat V_2 + g_1\hat V_1}\Phi_i, \qquad \text{(IV.3)} \\ \mathcal{L}_{Yukawa} &= \int d^2\theta\ (\mathcal{W}_R + \mathcal{W}_{NR}) + h.c. \qquad \text{(IV.4)}\end{aligned}$$

The index R in a superpotential refers to the so-called R-parity [16] which adjusts a "+" charge to all the ordinary particles and a "−" charge to their superpartners. The first part of $\mathcal{W}$ is R-symmetric

$$W_R = \epsilon_{ij}(y^U_{ab}Q^j_a U^c_b H^i_2 + y^D_{ab}Q^j_a D^c_b H^i_1 + y^L_{ab}L^j_a E^c_b H^i_1 + \mu H^i_1 H^j_2), \tag{IV.5}$$

where $i,j = 1,2,3$ are the $SU(2)$ and $a,b = 1,2,3$ are the generation indices; colour indices are suppressed. This part of the Lagrangian almost exactly repeats that of the SM except that the fields are now the superfields rather than the ordinary fields of the SM. The only difference is the last term which describes the Higgs mixing. It is absent in the SM since there is only one Higgs field there.

The second part is R-nonsymmetric

$$W_{NR} = \epsilon_{ij}(\lambda^L_{abd}L^i_a L^j_b E^c_d + \lambda^{L\prime}_{abd}L^i_a Q^j_b D^c_d + \mu'_a L^i_a H^j_2) + \lambda^B_{abd}U^c_a D^c_b D^c_d. \tag{IV.6}$$

These terms are absent in the SM. The reason is very simple: one can not replace the superfields in eq.(IV.6) by the ordinary fields like in eq.(IV.5) because of the Lorentz invariance. These terms have a different property, they violate either lepton (the first 3 terms in eq.(IV.6)) or baryon number (the last term). Since both effects are not observed in Nature, these terms must be suppressed or be excluded. One can avoid such terms if one introduces special symmetry called the R-symmetry. This is the global $U(1)_R$ invariance

$$U(1)_R: \quad \theta \to e^{i\alpha}\theta, \quad \Phi \to e^{in\alpha}\Phi, \tag{IV.7}$$

which is reduced to the discrete group Z_2, called the R-parity. The R-parity quantum number is given by $R = (-1)^{3(B-L)+2S}$ for particles with spin S. Thus, all the ordinary particles have the R-parity quantum number equal to $R = +1$, while all the superpartners have R-parity quantum number equal to $R = -1$. The R-parity obviously forbids the W_{NR} terms. However, it may well be that these terms are present, though experimental limits on the couplings are very severe

$$\lambda^L_{abc}, \quad \lambda^{L\prime}_{abc} < 10^{-4}, \qquad \lambda^B_{abc} < 10^{-9}.$$

IV.3 Properties of interactions

If one assumes that the R-parity is preserved, then the interactions of superpartners are essentially the same as in the SM, but two of three particles involved into an interaction at any vertex are replaced by superpartners. The reason for it is the R-parity. Conservation of the R-parity has two consequences

- the superpartners are created in pairs;
- the lightest superparticle (LSP) is stable. Usually it is photino $\tilde{\gamma}$, the superpartner of a photon with some admixture of neutral higgsino.

Typical vertices are shown in Figs.7. The tilde above a letter denotes the corresponding superpartner. Note that the coupling is the same in all the vertices involving superpartners.

IV.4 Creation and decay of superpartners

The above-mentioned rule together with the Feynman rules for the SM enables one to draw diagrams describing creation of superpartners. One of the most promising processes is the e^+e^- annihilation (see Fig.8).

The usual kinematic restriction is given by the c.m. energy $m^{max}_{sparticle} \leq \frac{\sqrt{s}}{2}$. Similar processes take place at hadron colliders with electrons and positrons being replaced by quarks and gluons.

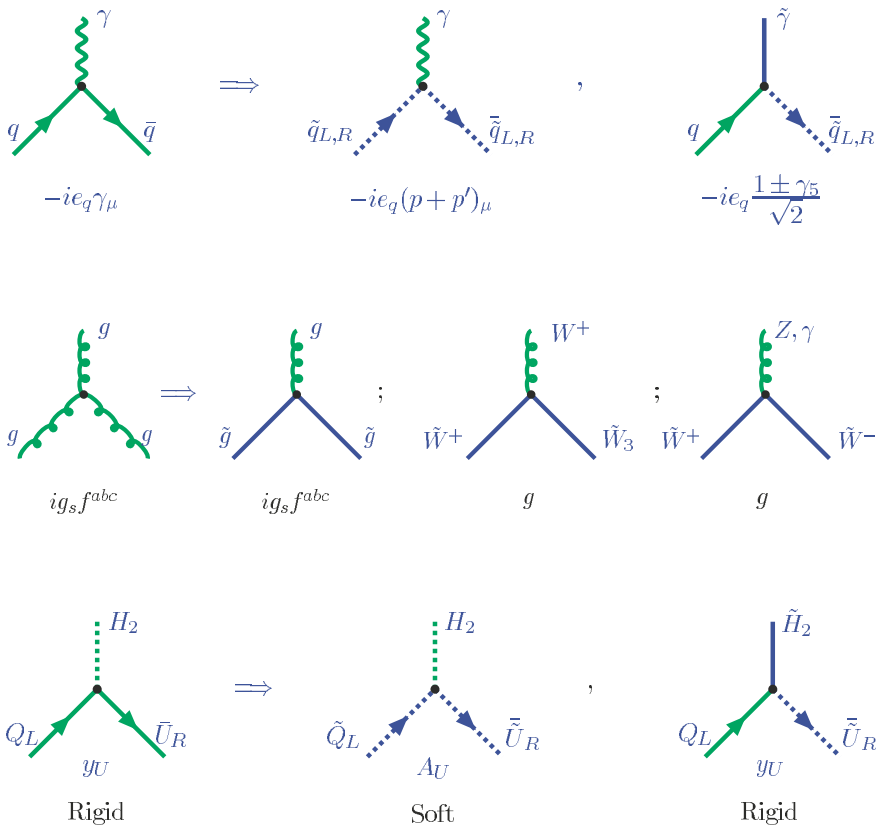

Figure 7. Gauge-matter interaction, Gauge self-interaction and Yukawa-type interaction

Creation of superpartners can be accompanied by creation of ordinary particles as well. We consider various experimental signatures for e^+e^- and hadron colliders below. They crucially depend on SUSY breaking pattern and on the mass spectrum of superpartners.

The decay properties of superpartners also depend on their masses. For the quark and lepton superpartners the main processes are shown in Fig.9.

When the R-parity is conserved, new particles will eventually end up giving neutralinos (the lightest superparticle) whose interactions are comparable to those of neutrinos and they leave undetected. Therefore, their signature would be missing energy and transverse momentum.

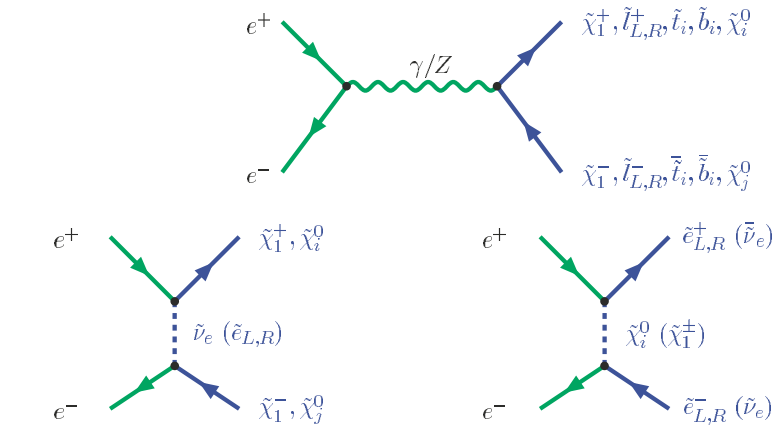

Figure 8. Creation of superpartners

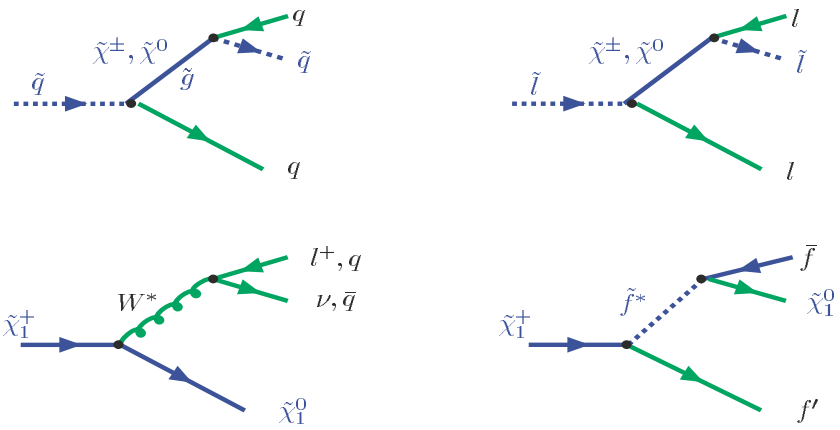

Figure 9. Decay of superpartners

Thus, if supersymmetry exists in Nature and if it is broken somewhere below 1 TeV, then it will be possible to detect it in the nearest future.

V. Breaking of SUSY in the MSSM

Since none of the fields of the MSSM can develop non-zero v.e.v. to break SUSY without spoiling the gauge invariance, it is supposed that spontaneous supersymmetry breaking takes place via some other fields. The most common scenario for producing low-energy supersymmetry breaking is called the *hidden sector* one [17]. According to this scenario, there exist two sectors: the usual matter belongs to the "visible" one, while the second, "hidden" sector, contains fields which lead to breaking of supersymmetry. These two sectors interact with each other by ex-

change of some fields called *messengers*, which mediate SUSY breaking from the hidden to the visible sector. There might be various types of messenger fields: gravity, gauge, etc. The hidden sector is the weakest part of the MSSM. It contains a lot of ambiguities and leads to uncertainties of the MSSM predictions considered below.

So far four main mechanisms there are known to mediate SUSY breaking from a hidden to a visible sector:

- Gravity mediation (SUGRA) [18];
- Gauge mediation [19];
- Anomaly mediation [20];
- Gaugino mediation [21].

All four mechanisms of soft SUSY breaking are different in details but are common in results. Predictions for the sparticle spectrum depend on the mechanism of SUSY breaking. For comparison of four above-mentioned mechanisms we show in Fig.10 the sample spectra as the ratio to the gaugino mass M_2 [22].

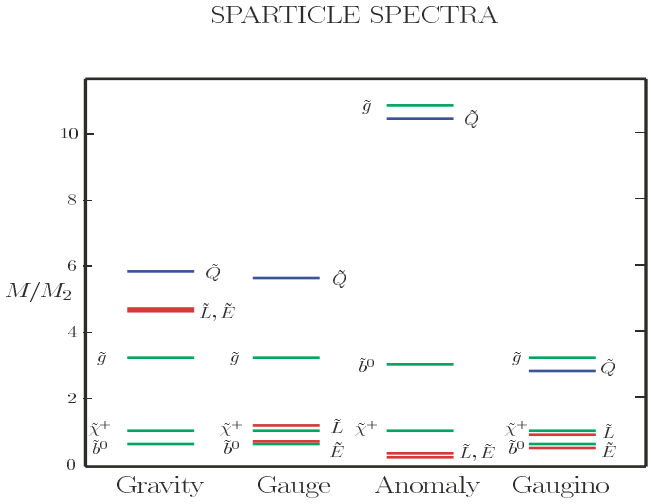

Figure 10. Superparticle spectra for various mediation mechanisms

In what follows, to calculate the mass spectrum of superpartners, we need an explicit form of SUSY breaking terms. For the MSSM and without the R-parity violation one has

$$
\begin{aligned}
-\mathcal{L}_{Breaking} &= \sum_i m_{0i}^2 |\varphi_i|^2 + \left(\frac{1}{2} \sum_\alpha M_\alpha \tilde{\lambda}_\alpha \tilde{\lambda}_\alpha + B H_1 H_2 \right. \qquad \text{(V.1)} \\
&+ \left. A^U_{ab} \tilde{Q}_a \tilde{U}^c_b H_2 + A^D_{ab} \tilde{Q}_a \tilde{D}^c_b H_1 + A^L_{ab} \tilde{L}_a \tilde{E}^c_b H_1 + h.c. \right),
\end{aligned}
$$

where we have suppressed the $SU(2)$ indices. Here φ_i are all scalar fields, $\tilde{\lambda}_\alpha$ are the gaugino fields, $\tilde{Q}, \tilde{U}, \tilde{D}$ and $\tilde{L}, \tilde{E}$ are the squark and slepton fields, respectively, and $H_{1,2}$ are the SU(2) doublet Higgs fields.

Eq.(V.1) contains a vast number of free parameters which spoils the prediction power of the model. To reduce their number, we adopt the so-called *universality* hypothesis, i.e., we assume the universality or equality of various soft parameters at a high energy scale, namely, we put all the spin 0 particle masses to be equal to the universal value m_0, all the spin 1/2 particle (gaugino) masses to be equal to $m_{1/2}$ and all the cubic and quadratic terms, proportional to A and B, to repeat the structure of the Yukawa superpotential (IV.5). This is an additional requirement motivated by the supergravity mechanism of SUSY breaking. Universality is not a necessary requirement and one may consider nonuniversal soft terms as well. However, it will not change the qualitative picture presented below; so for simplicity, in what follows we consider the universal boundary conditions. In this case, eq.(V.1) takes the form

$$\begin{aligned} -\mathcal{L}_{Breaking} &= m_0^2 \sum_i |\varphi_i|^2 + \Bigg(\frac{1}{2} m_{1/2} \sum_\alpha \tilde{\lambda}_\alpha \tilde{\lambda}_\alpha \qquad \text{(V.2)} \\ &+ A[y^U_{ab} \tilde{Q}_a \tilde{U}^c_b H_2 + y^D_{ab} \tilde{Q}_a \tilde{D}^c_b H_1 + y^L_{ab} \tilde{L}_a \tilde{E}^c_b H_1] + B[\mu H_1 H_2] + h.c. \Bigg) , \end{aligned}$$

The soft terms explicitly break supersymmetry. As will be shown later, they lead to the mass spectrum of superpartners different from that of ordinary particles. Remind that the masses of quarks and leptons remain zero until $SU(2)$ invariance is spontaneously broken.

V.1 The soft terms and the mass formulas

There are two main sources of the mass terms in the Lagrangian: the D terms and soft ones. With given values of $m_0, m_{1/2}, \mu, Y_t, Y_b, Y_\tau, A$, and B one can construct the mass matrices for all the particles. Knowing them at the GUT scale, one can solve the corresponding RG equations, thus linking the values at the GUT and electroweak scales. Substituting these parameters into the mass matrices, one can predict the mass spectrum of superpartners [23, 24].

Gaugino-higgsino mass terms. The mass matrix for gauginos, the superpartners of the gauge bosons, and for higgsinos, the superpartners of the Higgs bosons, is nondiagonal, thus leading to their mixing. The mass terms look like

$$\mathcal{L}_{Gaugino-Higgsino} = -\frac{1}{2} M_3 \bar{\lambda}_a \lambda_a - \frac{1}{2} \bar{\chi} M^{(0)} \chi - (\bar{\psi} M^{(c)} \psi + h.c.), \quad \text{(V.3)}$$

where $\lambda_a, a = 1, 2, \ldots, 8$, are the Majorana gluino fields and

$$\chi = \begin{pmatrix} \tilde{B}^0 \\ \tilde{W}^3 \\ \tilde{H}_1^0 \\ \tilde{H}_2^0 \end{pmatrix}, \quad \psi = \begin{pmatrix} \tilde{W}^+ \\ \tilde{H}^+ \end{pmatrix} \tag{V.4}$$

are, respectively, the Majorana neutralino and Dirac chargino fields.

The neutralino mass matrix is

$$M^{(0)} = \begin{pmatrix} M_1 & 0 & -M_Z \cos\beta \sin_W & M_Z \sin\beta \sin_W \\ 0 & M_2 & M_Z \cos\beta \cos_W & -M_Z \sin\beta \cos_W \\ -M_Z \cos\beta \sin_W & M_Z \cos\beta \cos_W & 0 & -\mu \\ M_Z \sin\beta \sin_W & -M_Z \sin\beta \cos_W & -\mu & 0 \end{pmatrix}, \tag{V.5}$$

where $\tan\beta = v_2/v_1$ is the ratio of two Higgs v.e.v.s and $\sin_W = \sin\theta_W$ is the usual sinus of the weak mixing angle. The physical neutralino masses $M_{\tilde{\chi}_i^0}$ are obtained as eigenvalues of this matrix after diagonalization.

For charginos one has

$$M^{(c)} = \begin{pmatrix} M_2 & \sqrt{2} M_W \sin\beta \\ \sqrt{2} M_W \cos\beta & \mu \end{pmatrix}. \tag{V.6}$$

This matrix has two chargino eigenstates $\tilde{\chi}_{1,2}^{\pm}$ with mass eigenvalues

$$\begin{aligned} M_{1,2}^2 &= \frac{1}{2}\left[M_2^2 + \mu^2 + 2M_W^2 \right. \\ &\left. \mp \sqrt{(M_2^2 - \mu^2)^2 + 4M_W^4 \cos^2 2\beta + 4M_W^2 (M_2^2 + \mu^2 + 2M_2 \mu \sin 2\beta)} \right]. \end{aligned} \tag{V.7}$$

Squark and slepton masses. Non-negligible Yukawa couplings cause a mixing between the electroweak eigenstates and the mass eigenstates of the third generation particles. The mixing matrices for $\tilde{m}_t^2, \tilde{m}_b^2$ and $\tilde{m}_\tau^2$ are

$$\begin{pmatrix} \tilde{m}_{tL}^2 & m_t(A_t - \mu \cot\beta) \\ m_t(A_t - \mu \cot\beta) & \tilde{m}_{tR}^2 \end{pmatrix}, \tag{V.8}$$

$$\begin{pmatrix} \tilde{m}_{bL}^2 & m_b(A_b - \mu \tan\beta) \\ m_b(A_b - \mu \tan\beta) & \tilde{m}_{bR}^2 \end{pmatrix}, \tag{V.9}$$

$$\begin{pmatrix} \tilde{m}_{\tau L}^2 & m_\tau(A_\tau - \mu \tan\beta) \\ m_\tau(A_\tau - \mu \tan\beta) & \tilde{m}_{\tau R}^2 \end{pmatrix} \tag{V.10}$$

with

$$\begin{aligned}
\tilde{m}^2_{tL} &= \tilde{m}^2_Q + m_t^2 + \frac{1}{6}(4M_W^2 - M_Z^2)\cos 2\beta, \\
\tilde{m}^2_{tR} &= \tilde{m}^2_U + m_t^2 - \frac{2}{3}(M_W^2 - M_Z^2)\cos 2\beta, \\
\tilde{m}^2_{bL} &= \tilde{m}^2_Q + m_b^2 - \frac{1}{6}(2M_W^2 + M_Z^2)\cos 2\beta, \\
\tilde{m}^2_{bR} &= \tilde{m}^2_D + m_b^2 + \frac{1}{3}(M_W^2 - M_Z^2)\cos 2\beta, \\
\tilde{m}^2_{\tau L} &= \tilde{m}^2_L + m_\tau^2 - \frac{1}{2}(2M_W^2 - M_Z^2)\cos 2\beta, \\
\tilde{m}^2_{\tau R} &= \tilde{m}^2_E + m_\tau^2 + (M_W^2 - M_Z^2)\cos 2\beta
\end{aligned}$$

and the mass eigenstates are the eigenvalues of these mass matrices. For the light generations the mixing is negligible.

The first terms here ($\tilde{m}^2$) are the soft ones, which are calculated using the RG equations starting from their values at the GUT (Planck) scale. The second ones are the usual masses of quarks and leptons and the last ones are the D terms of the potential.

V.2 The Higgs potential

As has already been mentioned, the Higgs potential in the MSSM is totally defined by superpotential (and the soft terms). Due to the structure of $\mathcal{W}$ the Higgs self-interaction is given by the D-terms while the F-terms contribute only to the mass matrix. The tree level potential is

$$\begin{aligned}
V_{tree}(H_1, H_2) &= m_1^2|H_1|^2 + m_2^2|H_2|^2 - m_3^2(H_1 H_2 + h.c.) \\
&+ \frac{g^2 + g^{'2}}{8}(|H_1|^2 - |H_2|^2)^2 + \frac{g^2}{2}|H_1^+ H_2|^2, \quad \text{(V.11)}
\end{aligned}$$

where $m_1^2 = m_{H_1}^2 + \mu^2, m_2^2 = m_{H_2}^2 + \mu^2$. At the GUT scale $m_1^2 = m_2^2 = m_0^2 + \mu_0^2,\ m_3^2 = -B\mu_0$. Notice that the Higgs self-interaction coupling in eq.(V.10) is fixed and defined by the gauge interactions as opposed to the SM.

The potential (V.10), in accordance with supersymmetry, is positive definite and stable. It has no nontrivial minimum different from zero. Indeed, let us write the minimization condition for the potential (V.10)

$$\begin{aligned}
\frac{1}{2}\frac{\delta V}{\delta H_1} &= m_1^2 v_1 - m_3^2 v_2 + \frac{g^2 + g'^2}{4}(v_1^2 - v_2^2)v_1 = 0, \quad \text{(V.12)} \\
\frac{1}{2}\frac{\delta V}{\delta H_2} &= m_2^2 v_2 - m_3^2 v_1 + \frac{g^2 + g'^2}{4}(v_1^2 - v_2^2)v_2 = 0, \quad \text{(V.13)}
\end{aligned}$$

where we have introduced the notation

$$< H_1 > \equiv v_1 = v\cos\beta,\ \ < H_2 > \equiv v_2 = v\sin\beta,\ \ v^2 = v_1^2 + v_2^2,\ \ \tan\beta \equiv \frac{v_2}{v_1}.$$

Solution of eqs.(V.12),(V.13) can be expressed in terms of v^2 and $\sin 2\beta$

$$v^2 = \frac{4(m_1^2 - m_2^2\tan^2\beta)}{(g^2+g'^2)(\tan^2\beta - 1)}, \quad \sin 2\beta = \frac{2m_3^2}{m_1^2+m_2^2}. \tag{V.14}$$

One can easily see from eq.(V.14) that if $m_1^2 = m_2^2 = m_0^2 + \mu_0^2$, v^2 happens to be negative, i.e. the minimum does not exist. In fact, real positive solutions to eqs.(V.12),(V.13) exist only if the following conditions are satisfied:

$$m_1^2 + m_2^2 > 2m_3^2, \qquad m_1^2 m_2^2 < m_3^4, \tag{V.15}$$

which is not the case at the GUT scale. This means that spontaneous breaking of the $SU(2)$ gauge invariance, which is needed in the SM to give masses for all the particles, does not take place in the MSSM.

This strong statement is valid, however, only at the GUT scale. Indeed, going down with energy, the parameters of the potential (V.10) are renormalized. They become the "running" parameters with the energy scale dependence given by the RG equations. The running of the parameters leads to a remarkable phenomenon known as *radiative spontaneous symmetry breaking* to be discussed below.

Provided conditions (V.15) are satisfied, the mass matrices at the tree level are
CP-odd components P_1 and P_2 :

$$\mathcal{M}^{odd} = \left.\frac{\partial^2 V}{\partial P_i \partial P_j}\right|_{H_i = v_i} = \begin{pmatrix} \tan\beta & 1 \\ 1 & \cot\beta \end{pmatrix} m_3^2, \tag{V.16}$$

CP-even neutral components S_1 and S_2:

$$\mathcal{M}^{ev} = \left.\frac{\partial^2 V}{\partial S_i \partial S_j}\right| = \begin{pmatrix} \tan\beta & -1 \\ -1 & \cot\beta \end{pmatrix} m_3^2 + \begin{pmatrix} \cot\beta & -1 \\ -1 & \tan\beta \end{pmatrix} M_Z \frac{\sin 2\beta}{2}, \tag{V.17}$$

Charged components H^- and H^+:

$$\mathcal{M}^{ch} = \left.\frac{\partial^2 V}{\partial H_i^+ \partial H_j^-}\right|_{H_i = v_i} = \begin{pmatrix} \tan\beta & 1 \\ 1 & \cot\beta \end{pmatrix} (m_3^2 + M_W \cos\beta \sin\beta). \tag{V.18}$$

Diagonalizing the mass matrices, one gets the mass eigenstates:

$$\begin{cases} G^0 &= -\cos\beta P_1 + \sin\beta P_2, & \textit{Goldstone boson } \to Z_0, \\ A &= \sin\beta P_1 + \cos\beta P_2, & \textit{Neutral } CP = -1 \textit{ Higgs}, \end{cases}$$

$$\begin{cases} G^+ &= -\cos\beta (H_1^-)^* + \sin\beta H_2^+, & \textit{Goldstone boson } \to W^+, \\ H^+ &= \sin\beta (H_1^-)^* + \cos\beta H_2^+, & \textit{Charged Higgs}, \end{cases}$$

$$\begin{cases} h &= -\sin\alpha S_1 + \cos\alpha S_2, & \textit{SM Higgs boson } CP = 1, \\ H &= \cos\alpha S_1 + \sin\alpha S_2, & \textit{Extra heavy Higgs boson}, \end{cases}$$

where the mixing angle α is given by

$$\tan 2\alpha = \tan 2\beta \left(\frac{m_A^2 + M_Z^2}{m_A^2 - M_Z^2} \right).$$

The physical Higgs bosons acquire the following masses [14]:

$$\begin{aligned} \text{CP-odd neutral Higgs } A: &\qquad m_A^2 = m_1^2 + m_2^2, \\ \text{Charge Higgses } H^\pm: &\qquad m_{H^\pm}^2 = m_A^2 + M_W^2, \end{aligned} \tag{V.19}$$

CP-even neutral Higgses H, h:

$$m_{H,h}^2 = \frac{1}{2} \left[m_A^2 + M_Z^2 \pm \sqrt{(m_A^2 + M_Z^2)^2 - 4 m_A^2 M_Z^2 \cos^2 2\beta} \right], \tag{V.20}$$

where, as usual,

$$M_W^2 = \frac{g^2}{2} v^2, \quad M_Z^2 = \frac{g^2 + g'^2}{2} v^2.$$

This leads to the once celebrated SUSY mass relations

$$\begin{gathered} m_{H^\pm} \geq M_W, \quad m_h \leq m_A \leq M_H, \\ m_h \leq M_Z |\cos 2\beta| \leq M_Z, \quad m_h^2 + m_H^2 = m_A^2 + M_Z^2. \end{gathered} \tag{V.21}$$

Thus, the lightest neutral Higgs boson happens to be lighter than the Z boson, which clearly distinguishes it from the SM one. Though we do not know the mass of the Higgs boson in the SM, there are several indirect constraints leading to the lower boundary of $m_h^{SM} \geq 135$ GeV. After including the radiative corrections, the mass of the lightest Higgs boson in the MSSM, m_h, however increases. We consider it in more detail below.

V.3 Renormalization group analysis

To calculate the low energy values of the soft terms, we use the corresponding RG equations. The one-loop RG equations for the rigid MSSM couplings are [25]

$$\begin{aligned}
\frac{d\tilde{\alpha}_i}{dt} &= b_i\tilde{\alpha}_i^2, \quad t \equiv \log Q^2/M_{GUT}^2 \\
\frac{dY_U}{dt} &= -Y_L\left(\frac{16}{3}\tilde{\alpha}_3 + 3\tilde{\alpha}_2 + \frac{13}{15}\tilde{\alpha}_1 - 6Y_U - Y_D\right), \\
\frac{dY_D}{dt} &= -Y_D\left(\frac{16}{3}\tilde{\alpha}_3 + 3\tilde{\alpha}_2 + \frac{7}{15}\tilde{\alpha}_1 - Y_U - 6Y_D - Y_L\right), \\
\frac{dY_L}{dt} &= -Y_L\left(3\tilde{\alpha}_2 + \frac{9}{5}\tilde{\alpha}_1 - 3Y_D - 4Y_L\right),
\end{aligned} \tag{V.22}$$

where we use the notation $\tilde{\alpha} = \alpha/4\pi = g^2/16\pi^2,\ Y = y^2/16\pi^2$.

For the soft terms one finds

$$\begin{aligned}
\frac{dM_i}{dt} &= b_i\tilde{\alpha}_i M_i. \\
\frac{dA_U}{dt} &= \frac{16}{3}\tilde{\alpha}_3 M_3 + 3\tilde{\alpha}_2 M_2 + \frac{13}{15}\tilde{\alpha}_1 M_1 + 6Y_U A_U + Y_D A_D, \\
\frac{dA_D}{dt} &= \frac{16}{3}\tilde{\alpha}_3 M_3 + 3\tilde{\alpha}_2 M_2 + \frac{7}{15}\tilde{\alpha}_1 M_1 + 6Y_D A_D + Y_U A_U + Y_L A_L, \\
\frac{dA_L}{dt} &= 3\tilde{\alpha}_2 M_2 + \frac{9}{5}\tilde{\alpha}_1 M_1 + 3Y_D A_D + 4Y_L A_L, \\
\frac{dB}{dt} &= 3\tilde{\alpha}_2 M_2 + \frac{3}{5}\tilde{\alpha}_1 M_1 + 3Y_U A_U + 3Y_D A_D + Y_L A_L. \\
\frac{d\tilde{m}_Q^2}{dt} &= -\Big[(\frac{16}{3}\tilde{\alpha}_3 M_3^2 + 3\tilde{\alpha}_2 M_2^2 + \frac{1}{15}\tilde{\alpha}_1 M_1^2) \\
&- Y_U(\tilde{m}_Q^2 + \tilde{m}_U^2 + m_{H_2}^2 + A_U^2) - Y_D(\tilde{m}_Q^2 + \tilde{m}_D^2 + m_{H_1}^2 + A_D^2)\Big], \\
\frac{d\tilde{m}_U^2}{dt} &= -\left[(\frac{16}{3}\tilde{\alpha}_3 M_3^2 + \frac{16}{15}\tilde{\alpha}_1 M_1^2) - 2Y_U(\tilde{m}_Q^2 + \tilde{m}_U^2 + m_{H_2}^2 + A_U^2)\right], \\
\frac{d\tilde{m}_D^2}{dt} &= -\left[(\frac{16}{3}\tilde{\alpha}_3 M_3^2 + \frac{4}{15}\tilde{\alpha}_1 M_1^2) - 2Y_D(\tilde{m}_Q^2 + \tilde{m}_D^2 + m_{H_1}^2 + A_D^2)\right], \\
\frac{d\tilde{m}_L^2}{dt} &= -\left[3(\tilde{\alpha}_2 M_2^2 + \frac{1}{5}\tilde{\alpha}_1 M_1^2) - Y_L(\tilde{m}_L^2 + \tilde{m}_E^2 + m_{H_1}^2 + A_L^2)\right], \\
\frac{d\tilde{m}_E^2}{dt} &= -\left[(\frac{12}{5}\tilde{\alpha}_1 M_1^2) - 2Y_L(\tilde{m}_L^2 + \tilde{m}_E^2 + m_{H_1}^2 + A_L^2)\right], \\
\frac{d\mu^2}{dt} &= -\mu^2\left[3(\tilde{\alpha}_2 + \frac{1}{5}\tilde{\alpha}_1) - (3Y_U + 3Y_D + Y_L)\right],
\end{aligned} \tag{V.23}$$

$$\frac{dm_{H_1}^2}{dt} = -\Big[3(\tilde{\alpha}_2 M_2^2 + \frac{1}{5}\tilde{\alpha}_1 M_1^2) - 3Y_D(\tilde{m}_Q^2 + \tilde{m}_D^2 + m_{H_1}^2 + A_D^2) - Y_L(\tilde{m}_L^2 + \tilde{m}_E^2 + m_{H_1}^2 + A_L^2)\Big],$$

$$\frac{dm_{H_2}^2}{dt} = -\Big[3(\tilde{\alpha}_2 M_2^2 + \frac{1}{5}\tilde{\alpha}_1 M_1^2) - 3Y_U(\tilde{m}_Q^2 + \tilde{m}_U^2 + m_{H_2}^2 + A_U^2)\Big].$$

Having all the RG equations, one can now find the RG flow for the soft terms. Taking the initial values of the soft masses at the GUT scale in the interval between $10^2 \div 10^3$ GeV consistent with the SUSY scale suggested by unification of the gauge couplings (II.7) leads to the RG flow of the soft terms shown in Fig.11. [23, 24]

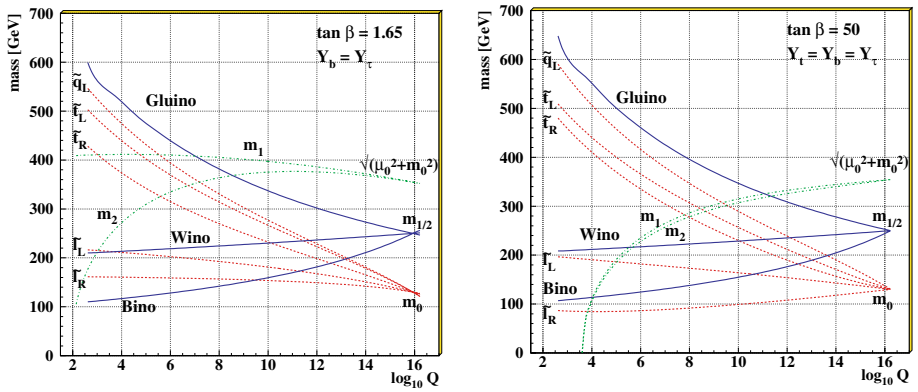

Figure 11. An example of evolution of sparticle masses and soft supersymmetry breaking parameters $m_1^2 = m_{H_1}^2 + \mu^2$ and $m_2^2 = m_{H_2}^2 + \mu^2$ for low (left) and high (right) values of $\tan\beta$

One should mention the following general features common to any choice of initial conditions:

i) The gaugino masses follow the running of the gauge couplings and split at low energies. The gluino mass is running faster than the others and is usually the heaviest due to the strong interaction.

ii) The squark and slepton masses also split at low energies, the stops (and sbottoms) being the lightest due to relatively big Yukawa couplings of the third generation.

iii) The Higgs masses (or at least one of them) are running down very quickly and may even become negative.

Typical dependence of the mass spectra on the initial conditions (m_0) is also shown in Fig.12 [26]. For a given value of $m_{1/2}$ the masses of the lightest particles are practically independent of m_0, while the

heavier ones increase with it monotonically. One can see that the lightest neutralinos and charginos as well as the stop squark may be rather light.

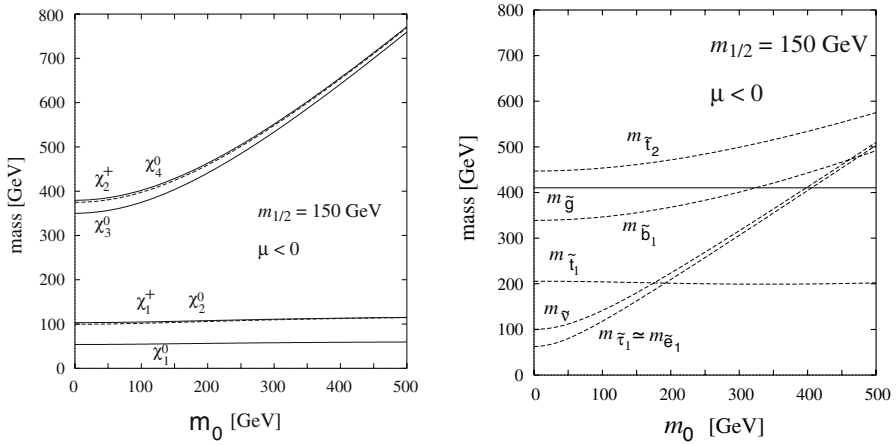

Figure 12. The masses of sparticles as functions of the initial value m_0

V.4 Radiative electroweak symmetry breaking

The running of the Higgs masses leads to the phenomenon known as *radiative electroweak symmetry breaking*. Indeed, one can see in Fig.11 that m_2^2 (or both m_1^2 and m_2^2) decreases when going down from the GUT scale to the M_Z scale and can even become negative. As a result, at some value of Q^2 the conditions (V.15) are satisfied, so that the nontrivial minimum appears. This triggers spontaneous breaking of the $SU(2)$ gauge invariance. The vacuum expectations of the Higgs fields acquire nonzero values and provide masses to quarks, leptons and $SU(2)$ gauge bosons, and additional masses to their superpartners.

In this way one also obtains the explanation of why the two scales are so much different. Due to the logarithmic running of the parameters, one needs a long "running time" to get m_2^2 (or both m_1^2 and m_2^2) to be negative when starting from a positive value of the order of $M_{SUSY} \sim 10^2 \div 10^3$ GeV at the GUT scale.

VI. Constrained MSSM

VI.1 Parameter space of the MSSM

The Minimal Supersymmetric Standard Model has the following free parameters: i) three gauge couplings α_i; ii) three matrices of the Yukawa couplings y_{ab}^i, where $i = L, U, D$; iii) the Higgs field mixing parameter

μ; iv) the soft supersymmetry breaking parameters. Compared to the SM there is an additional Higgs mixing parameter, but the Higgs self-coupling, which is arbitrary in the SM, is fixed by supersymmetry. The main uncertainty comes from the unknown soft terms.

With the universality hypothesis one is left with the following set of 5 free parameters defining the mass scales

$$\mu,\ m_0,\ m_{1/2},\ A \text{ and } B \leftrightarrow \tan\beta = \frac{v_2}{v_1}.$$

While choosing parameters and making predictions, one has two possible ways to proceed:

i) take the low-energy parameters like superparticle masses $\tilde{m}_{t1}, \tilde{m}_{t2}, m_A$, $\tan\beta$, mixings X_{stop}, μ, etc. as input and calculate cross-sections as functions of these parameters.

ii) take the high-energy parameters like the above mentioned 5 soft parameters as input, run the RG equations and find the low-energy values. Now the calculations can be carried out in terms of the initial parameters. The experimental constraints are sufficient to determine these parameters, albeit with large uncertainties.

VI.2 The choice of constraints

When subjecting constraints on the MSSM, perhaps, the most remarkable fact is that all of them can be fulfilled simultaneously. In our analysis we impose the following constraints on the parameter space of the MSSM:

• Gauge coupling constant unification;

This is one of the most restrictive constraints, which we have discussed in Sect 2. It fixes the scale of SUSY breaking of an order of 1 TeV.

• M_Z from electroweak symmetry breaking;

Radiative EW symmetry breaking (see eq.(V.14)) defines the mass of the Z-boson

$$M_Z^2 = 2\frac{m_1^2 - m_2^2\tan^2\beta}{\tan^2\beta - 1}. \qquad \text{(VI.1)}$$

This condition determines the value of μ for given values of m_0 and $m_{1/2}$.

• Yukawa coupling constant unification;

The masses of top, bottom and τ can be obtained from the low energy values of the running Yukawa couplings via

$$m_t = y_t\ v\sin\beta, \quad m_b = y_b\ v\cos\beta, \quad m_\tau = y_\tau\ v\cos\beta. \qquad \text{(VI.2)}$$

They can be translated to the pole masses with account taken of the radiative corrections. The requirement of bottom-tau Yukawa coupling

unification strongly restricts the possible solutions in m_t versus $\tan\beta$ plane [27] as it can be seen from Fig.13.

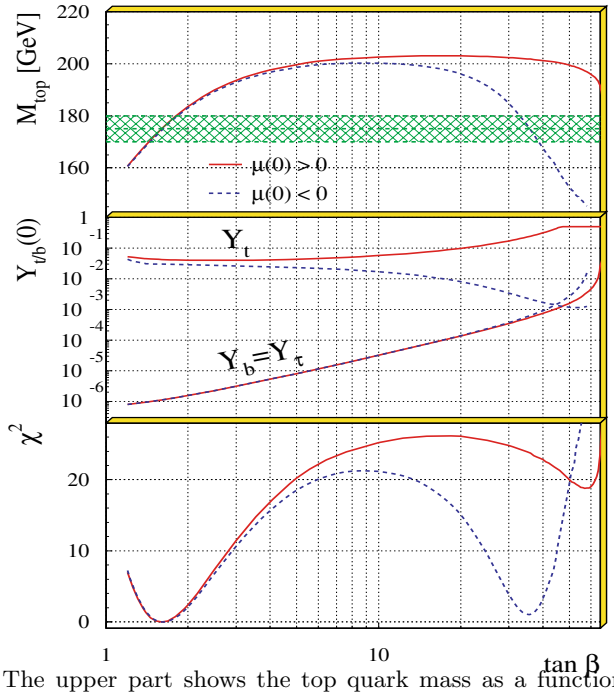

Figure 13. The upper part shows the top quark mass as a function of $\tan\beta$ for $m_0 = 600$ GeV, $m_{1/2} = 400$ GeV. The middle part shows the corresponding values of the Yukawa couplings at the GUT scale and the lower part of the χ^2 values.

• Precision measurement of decay rates;

We take the branching ratio $BR(b \to s\gamma)$ which has been measured by the CLEO [28] collaboration and later by ALEPH [29] and yields the world average of $BR(b \to s\gamma) = (3.14 \pm 0.48) \cdot 10^{-4}$. The Standard Model contribution to this process gives slightly lower result, thus leaving window for SUSY. This requirement imposes severe restrictions on the parameter space, especially for the case of large $\tan\beta$.

• Anomalous magnetic moment of muon.

Recent measurement of the anomalous magnetic moment indicates small deviation from the SM of the order of 2 σ. The deficiency may be easily filled with SUSY contribution, which is proportional to μ. This requires positive sign of μ that kills half of the parameter space of the MSSM [30].

• Experimental lower limits on SUSY masses;

SUSY particles have not been found so far and from the searches at LEP one knows the lower limit on the charged lepton and chargino masses of about half of the centre of mass energy [31]. The lower limit on the neutralino masses is smaller. There exist also limits on squark and gluino

masses from the hadron colliders [32]. These limits restrict the minimal values for the SUSY mass parameters.

- Dark Matter constraint;

Recent very precise astrophysical data restrict the amount of the Dark matter in the Universe up to 23%. Assuming $h_0 > 0.4$ one finds that the contribution of each relic particle species χ has to obey $\Omega_\chi h_0^2 \sim 0.1 \div 0.3$,. This serves as a very severe bound on SUSY parameters [33].

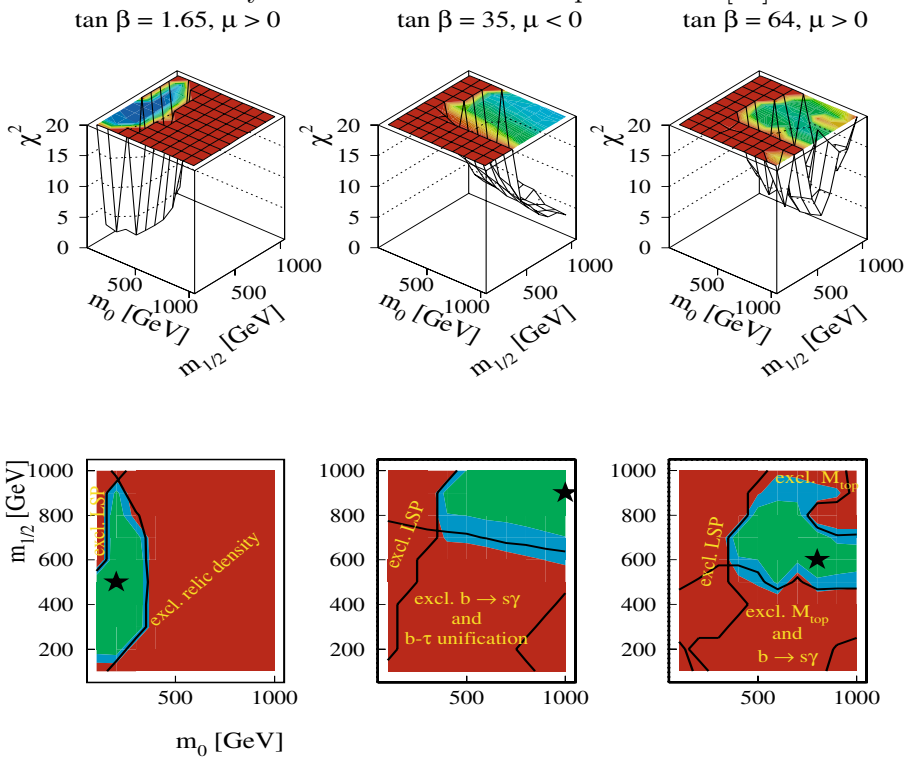

Figure 14. The χ^2-distribution for low and high $\tan\beta$ solutions. The different shades in the projections indicate steps of $\Delta\chi^2 = 4$. The stars indicate the optimum solution. Contours enclose domains by the particular constraints used in the analysis.

Having in mind the above mentioned constraints one can find the most probable region of the parameter space by minimizing the χ^2 function [24]. We first choose the value of the Higgs mixing parameter μ from the requirement of radiative EW symmetry breaking, then we take the values of $\tan\beta$ from the requirement of Yukawa coupling unification (see Fig.13). One finds two possible solutions: low $\tan\beta$ solution corresponding to $\tan\beta \approx 1.7$ and high $\tan\beta$ solution corresponding to $\tan\beta \approx 30 \div 60$.

What is left are the values of the soft parameters A, m_0 and $m_{1/2}$. However, the role of the trilinear coupling A is not essential. In what

follows, we consider the plane $m_0, m_{1/2}$ and find the allowed region in this plane. Each point at this plane corresponds to a fixed set of parameters and allows one to calculate the spectrum, the cross-sections, etc.

We present the allowed regions of the parameter space for low and high $\tan\beta$ scenarios in Fig.14. This plot demonstrates the role of various constraints in the χ^2 function and the contours shown correspond to the different constraints included in the analysis [34].

VI.3 The mass spectrum of superpartners

When the parameter set is fixed, one can calculate the mass spectrum of superpartners. Below we show the predicted mass spectrum corresponding to the best fit values indicated by stars in Fig.14 (see Table 1) [24].

SUSY masses in [GeV]		
Symbol	low $\tan\beta$	high $\tan\beta$
$\tilde{\chi}_1^0(\tilde{B})$, $\tilde{\chi}_2^0(\tilde{W}^3)$	214, 413	170, 322
$\tilde{\chi}_3^0(\tilde{H}_1)$,$\tilde{\chi}_4^0(\tilde{H}_2)$	1028, 1016	481, 498
$\tilde{\chi}_1^\pm(\tilde{W}^\pm)$, $\tilde{\chi}_2^\pm(\tilde{H}^\pm)$	413, 1026	322, 499
$\tilde{g}$	1155	950
$\tilde{e}_L$, $\tilde{e}_R$	303, 270	663, 621
$\tilde{\nu}_L$	290	658
$\tilde{q}_L$, $\tilde{q}_R$	1028, 936	1040, 1010
$\tilde{\tau}_1$, $\tilde{\tau}_2$	279, 403	537, 634
$\tilde{b}_1$, $\tilde{b}_2$	953, 1010	835, 915
$\tilde{t}_1$, $\tilde{t}_2$	727, 1017	735, 906
h, H	95, 1344	119, 565
A, $H^\pm$	1340, 1344	565, 571

Table 1. Values of the SUSY mass spectra for the low and high $\tan\beta$ solutions.

VI.4 Experimental signatures at e^+e^- colliders

Experiments are finally beginning to push into a significant region of supersymmetry parameter space. We know the sparticles and their couplings, but we do not know their masses and mixings. Given the mass spectrum one can calculate the cross-sections and consider the possibilities of observing new particles at modern accelerators. Otherwise, one can get restrictions on unknown parameters.

We start with e^+e^- colliders and, first of all, with LEP II. In the leading order creation of superpartners is given by the diagrams shown in Fig.8 above. For a given center of mass energy the cross-sections depend on the mass of created particles and vanish at the kinematic boundary. Experimental signatures are defined by the decay modes which vary with the mass spectrum. The main ones are summarized below.

Production	Key Decay Modes	Signatures
• $\tilde{l}_{L,R}\tilde{l}_{L,R}$	$\tilde{l}^{\pm}_R \to l^{\pm}\tilde{\chi}^0_i \searrow$ cascade	acomplanar pair of
	$\tilde{l}^{\pm}_L \to l^{\pm}\tilde{\chi}^0_i \nearrow$ decays	charged leptons + $\not{E}_T$
• $\tilde{\nu}\tilde{\nu}$	$\tilde{\nu} \to l^{\pm}\tilde{\chi}^0_1$	$\not{E}_T$
• $\tilde{\chi}^{\pm}_1\tilde{\chi}^{\pm}_1$	$\tilde{\chi}^{\pm}_1 \to \tilde{\chi}^0_1 l^{\pm}\nu,\ \tilde{\chi}^0_1 q\bar{q}'$	isol lept + 2 jets + $\not{E}_T$
	$\tilde{\chi}^{\pm}_1 \to \tilde{\chi}^0_2 f\bar{f}'$	pair of acomplanar
	$\tilde{\chi}^{\pm}_1 \to l\tilde{\nu}_l \to l\nu_l\tilde{\chi}^0_1$	leptons + $\not{E}_T$
	$\tilde{\chi}^{\pm}_1 \to \nu_l\tilde{l} \to \nu_l l\tilde{\chi}^0_1$	4 jets + $\not{E}_T$
• $\tilde{\chi}^0_i\tilde{\chi}^0_j$	$\tilde{\chi}^0_i \to \tilde{\chi}^0_1 X, \tilde{\chi}^0_j \to \tilde{\chi}^0_1 X'$	$X = \nu_l\bar{\nu}_l$ invisible
		$= \gamma, 2l$, 2 jets
		$2l + \not{E}_T, l + 2j + \not{E}_T$
• $\tilde{t}_i\tilde{t}_j$	$\tilde{t}_1 \to c\tilde{\chi}^0_1$	2 jets + $\not{E}_T$
	$\tilde{t}_1 \to b\tilde{\chi}^{\pm}_1 \to bf\bar{f}'\tilde{\chi}^0_1$	2 b jets + 2 leptons + $\not{E}_T$
		2 b jets + lepton + $\not{E}_T$
• $\tilde{b}_i\tilde{b}_j$	$\tilde{b}_i \to b\tilde{\chi}^0_1$	2 b jets + $\not{E}_T$
	$\tilde{b}_i \to b\tilde{\chi}^0_2 \to bf\bar{f}'\tilde{\chi}^0_1$	2 b jets + 2 leptons + $\not{E}_T$
		2 b jets + 2 jets + $\not{E}_T$

A characteristic feature of all possible signatures is the missing energy and transverse momenta, which is a trade mark of a new physics.

Numerous attempts to find superpartners at LEP II gave no positive result thus imposing the lower bounds on their masses [31]. Typical LEP II limits on the masses of superpartners are

$$\begin{array}{lll} m_{\chi^0_1} > 40\ GeV & m_{\tilde{e}_{L,R}} > 105\ GeV & m_{\tilde{t}} > 90\ GeV \\ m_{\chi^{\pm}_1} > 100\ GeV & m_{\tilde{\mu}_{L,R}} > 100\ GeV & m_{\tilde{b}} > 80\ GeV \\ & m_{\tilde{\tau}_{L,R}} > 80\ GeV & \end{array} \qquad \text{(VI.3)}$$

VI.5 Experimental signatures at hadron colliders

Experimental signatures at hadron colliders are similar to those at e^+e^- machines; however, here one has much wider possibilities. Besides the usual annihilation channel identical to e^+e^- one with the obvious replacement of electrons by quarks (see Fig.8), one has numerous processes of gluon fusion, quark-antiquark and quark-gluon scattering (see Fig.15).

Experimental SUSY signatures at the Tevatron (and LHC) are

Production	Key Decay Modes	Signatures
• $\tilde{g}\tilde{g}, \tilde{q}\tilde{q}, \tilde{g}\tilde{q}$	$\left.\begin{array}{l}\tilde{g} \to q\bar{q}\tilde{\chi}_1^0 \\ \quad q\bar{q}'\tilde{\chi}_1^\pm \\ \quad g\tilde{\chi}_1^0\end{array}\right\} m_{\tilde{q}} > m_{\tilde{g}}$	$\not{E}_T$ + multijets (+leptons)
	$\left.\begin{array}{l}\tilde{q} \to q\tilde{\chi}_i^0 \\ \tilde{q} \to q'\tilde{\chi}_i^\pm\end{array}\right\} m_{\tilde{g}} > m_{\tilde{q}}$	
• $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$	$\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 l^\pm\nu,\ \tilde{\chi}_2^0 \to \tilde{\chi}_1^0 ll$	Trilepton + $\not{E}_T$
	$\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 q\bar{q}', \tilde{\chi}_2^0 \to \tilde{\chi}_1^0 ll,$	Dilepton + jet + $\not{E}_T$
• $\tilde{\chi}_1^+\tilde{\chi}_1^-$	$\tilde{\chi}_1^+ \to l\tilde{\chi}_1^0 l^\pm\nu$	Dilepton + $\not{E}_T$
• $\tilde{\chi}_i^0\tilde{\chi}_i^0$	$\tilde{\chi}_i^0 \to \tilde{\chi}_1^0 X, \tilde{\chi}_i^0 \to \tilde{\chi}_1^0 X'$	$\not{E}_T$ + Dilept+(jets)+lept
• $\tilde{t}_1\tilde{t}_1$	$\tilde{t}_1 \to c\tilde{\chi}_1^0$	2 acollinear jets + $\not{E}_T$
	$\tilde{t}_1 \to b\tilde{\chi}_1^\pm, \tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 q\bar{q}'$	single lepton + $\not{E}_T + b's$
	$\tilde{t}_1 \to b\tilde{\chi}_1^\pm, \tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 l^\pm\nu,$	Dilepton + $\not{E}_T + b's$
• $\tilde{ll}, \tilde{l\nu}, \tilde{\nu\nu}$	$\tilde{l}^\pm \to l \pm \tilde{\chi}_i^0, \tilde{l}^\pm \to \nu_l\tilde{\chi}_i^\pm$	Dilepton + $\not{E}_T$
	$\tilde{\nu} \to \nu\tilde{\chi}_1^0$	Single lept + $\not{E}_T$ + jets
		$\not{E}_T$

Note again the characteristic missing energy and transverse momenta events. Contrary to e^+e^- colliders, at hadron machines the background is extremely rich and essential.

VI.6 The lightest superparticle

One of the crucial questions is the properties of the lightest superparticle. Different SUSY breaking scenarios lead to different experimental signatures and different LSP.

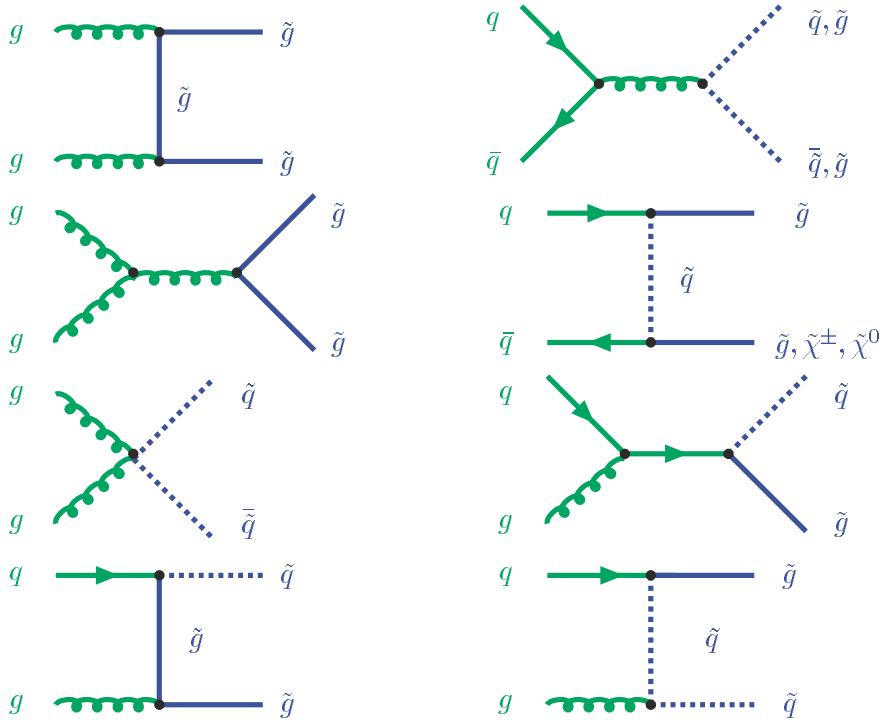

Figure 15. Gluon fusion, $q\bar{q}$ scattering, quark-gluon scattering

• Gravity mediation

In this case, the LSP is the lightest neutralino $\tilde{\chi}_1^0$, which is almost 90% photino for a low $\tan\beta$ solution and contains more higgsino admixture for high $\tan\beta$. The usual signature for LSP is missing energy; $\tilde{\chi}_1^0$ is stable and is the best candidate for the cold dark matter in the Universe. Typical processes, where the LSP is created, end up with jets + $\not{E}_T$, or leptons + $\not{E}_T$, or both jest + leptons + $\not{E}_T$.

• Gauge mediation

In this case the LSP is the gravitino $\tilde{G}$ which also leads to missing energy. The actual question here is what the NLSP, the next lightest particle, is. There are two possibilities:

i) $\tilde{\chi}_1^0$ is the NLSP. Then the decay modes are: $\tilde{\chi}_1^0 \to \gamma\tilde{G},\ h\tilde{G},\ Z\tilde{G}$. As a result, one has two hard photons + $\not{E}_T$, or jets + $\not{E}_T$.

ii) $\tilde{l}_R$ is the NLSP. Then the decay mode is $\tilde{l}_R \to \tau\tilde{G}$ and the signature is a charged lepton and the missing energy.

• Anomaly mediation

In this case, one also has two possibilities:

i) $\tilde{\chi}_1^0$ is the LSP and wino-like. It is almost degenerate with the NLSP.

ii) $\tilde{\nu}_L$ is the LSP. Then it appears in the decay of chargino $\tilde{\chi}^+ \to \tilde{\nu} l$ and the signature is the charged lepton and the missing energy.

• R-parity violation

In this case, the LSP is no longer stable and decays into the SM particles. It may be charged (or even colored) and may lead to rare decays like neutrinoless double β-decay, etc.

Experimental limits on the LSP mass follow from non-observation of the corresponding events. Modern lower limit is around 40 GeV .

VII. The Higgs boson mass in the MSSM

One of the hottest topics in the SM now is the search for the Higgs boson. It is also a window to a new physics. Below we consider properties of the Higgs boson in the MSSM.

It has already been mentioned that in the MSSM the mass of the lightest Higgs boson is predicted to be less than the Z-boson mass. This is, however, the tree level result and the masses acquire the radiative corrections. With account taken of the one-loop radiative corrections the lightest Higgs mass is

$$m_h^2 \approx M_Z^2 \cos^2 2\beta + \frac{3g^2 m_t^4}{16\pi^2 M_W^2} \log \frac{\tilde{m}_{t_1}^2 \tilde{m}_{t_2}^2}{m_t^4}. \qquad \text{(VII.1)}$$

One finds that the one-loop correction is positive and increases the mass value. Two loop corrections have the opposite effect but are smaller [36].

The Higgs mass depends mainly on the following parameters: the top mass, the squark masses, the mixing in the stop sector and $\tan\beta$. The maximum Higgs mass is obtained for large $\tan\beta$, for a maximum value of the top and squark masses and a minimum value of the stop mixing.

The lightest Higgs boson mass m_h is shown as a function of $\tan\beta$ in Fig. 16 [35]. The shaded band corresponds to the uncertainty from the stop mass and stop mixing for $m_t = 175$ GeV. The upper and lower lines correspond to m_t=170 and 180 GeV, respectively.

Combining all the uncertainties the results for the Higgs mass in the CMSSM can be summarized as follows:

• The low $\tan\beta$ scenario ($\tan\beta < 3.3$) of the CMSSM is excluded by the lower limit on the Higgs mass of 113.3 GeV [7].

• For the high $\tan\beta$ scenario the Higgs mass is found to be [35]:

$$m_h = 115 \pm 3 \text{ (stopm) } \pm 1.5 \text{ (stopmix) } \pm 2 \text{ (theory) } \pm 5 \text{ (topm) GeV},$$

where the errors are the estimated standard deviations around the central value.

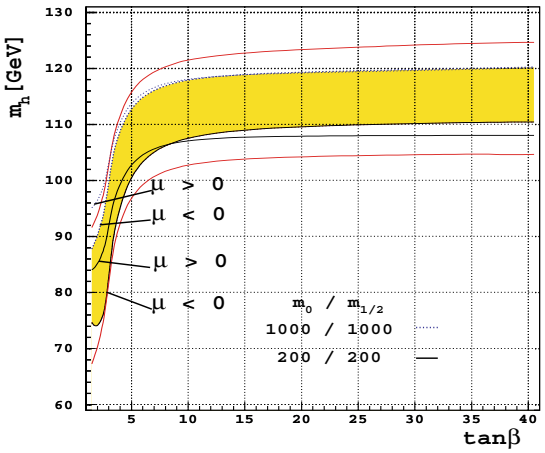

Figure 16. The mass of the lightest Higgs boson in the MSSM as a function of $\tan\beta$

However, these SUSY limits on the Higgs mass may not be so restricting if non-minimal SUSY models are considered. However, more sophisticated models do not change the generic feature of SUSY theories, the presence of the light Higgs boson.

VIII. Perspectives of SUSY observation

With the LEP shut down, further attempts to discover supersymmetry are connected with the Tevatron and LHC hadron colliders.

Tevatron

Tevatron Run II has the c.m. energy of 2 TeV with planned luminosity almost 10 times greater than in RUN I. However, since it is a hadron collider, not the full energy goes into collision taken away by those quarks in a proton that do not take part in the interaction. Due to a severe background, this collider needs time to reach the integrated luminosity required for SUSY discovery.

We show in Table 2 [37] the discovery reach of the Tevatron for squarks of the third generation. Modern exclusion areas are also shown in plots in Fig.17 [38]. One can see that they are still far from the expected masses given in Table 1.

LHC

The LHC hadron collider is the ultimate machine for a new physics at the TeV scale. Its c.m. energy is planned to be 14 TeV with very high luminosity up to a few hundred fb^{-1}. The LHC is supposed to cover the

Decay ($Br = 100\%$)	Subsequent Decay	Final State of $\tilde{b}_1\bar{\tilde{b}}_1$ or $\tilde{t}_1\bar{\tilde{t}}_1$	Discovery Reach @20 fb$^{(}-1$)	(Run I)
$\tilde{b}_1 \to b\tilde{\chi}_1^0$		$bb\not{E}_T$	260 GeV/c^2	(146 GeV/c^2)
$\tilde{t}_1 \to c\tilde{\chi}_1^0$		$cc\not{E}_T$	220 GeV/c^2	(116 GeV/c^2)
$\tilde{t}_1 \to bl\tilde{\nu}$	$\tilde{\nu} \to \nu\tilde{\chi}_1^0$	$l^+l^-b\not{E}_T$	240 GeV/c^2	(140 GeV/c^2)
$\tilde{t}_1 \to bl\tilde{\nu}\tilde{\chi}_1^0$		$l^+l^-b\not{E}_T$	-	(129 GeV/c^2)
$\tilde{t}_1 \to b\tilde{\chi}_1^\pm$	$\tilde{\chi}_1^\pm \to W^{(*)}\tilde{\chi}_1^0$	$l^+l^-b\not{E}_T$;	210 GeV/c^2	(-)
$\tilde{t}_1 \to bW\tilde{\chi}_1^0$		$l^+l^-bj\not{E}_T$	190 GeV/c^2	(-)

Table 2. Discovery reaches on $M_{\tilde{b}}$ and $M_{\tilde{t}}$ expected in Run II.

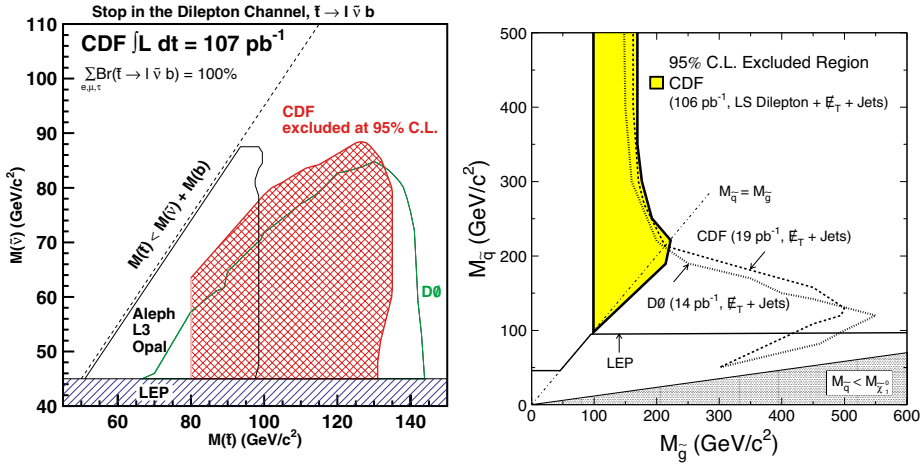

Figure 17. Exclusion plots for squarks and sneutrinos (left) and squarks and gluino (right) at Tevatron

wide range of parameters of the MSSM (see Fig.18 [39]) and discover the superpartners with the masses below 2 TeV [40]. This will be a crucial test for the MSSM and the low energy supersymmetry.

IX. Conclusion

Supersymmetry is now the most popular extension of the Standard Model. It promises us that new physics is round the corner at a TeV scale to be exploited at new machines of this decade. If our expectations

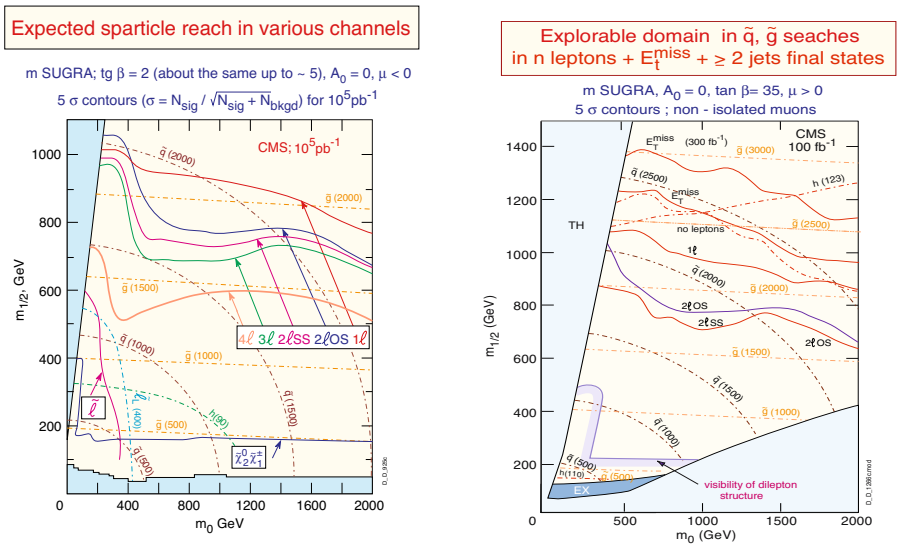

Figure 18. Expected sparticle reach at LHC

are correct, very soon we will face new discoveries, the whole world of supersymmetric particles will show up and the table of fundamental particles will be enlarged in increasing rate. This would be a great step in understanding the microworld.

Acknowledgements

Financial support from CNRS, NATO and Collectivité Territoriale Corse as well as RFBR grant # 02-02-16889, Russian MIST grant # 2339.2003.2 is kindly acknowledged.

References

[1] Y. A. Golfand and E. P. Likhtman, *JETP Letters* **13** (1971) 452; D. V. Volkov and V. P. Akulov, *JETP Letters* **16** (1972) 621; J. Wess and B. Zumino, *Phys. Lett.* **B49** (1974) 52.

[2] P. Fayet and S. Ferrara, *Phys. Rep.* **32** (1977) 249; M. F. Sohnius, *Phys. Rep.* **128** (1985) 41; H. P. Nilles, *Phys. Rep.* **110** (1984) 1; H. E. Haber and G. L. Kane, *Phys. Rep.* **117** (1985) 75; A. B. Lahanas and D. V. Nanopoulos, *Phys. Rep.* **145** (1987) 1.

[3] J. Wess and J. Bagger, *"Supersymmetry and Supergravity"*, Princeton Univ. Press, 1983.

[4] S. J. Gates, M. Grisaru, M. Roček and W. Siegel, *"Superspace or One Thousand and One Lessons in Supersymmetry"*, Benjamin & Cummings, 1983.

[5] S. Coleman and J .Mandula, *Phys.Rev.* **159** (1967) 1251.

[6] G. G. Ross, *"Grand Unified Theories"*, Benjamin & Cummings, 1985.

[7] D. E. Groom *et al.*, "Review of Particle Physics", Eur. Phys. J. **C15** (2000) 1.

[8] U. Amaldi, W. de Boer and H. Fürstenau, *Phys. Lett.* **B260** (1991) 447.

[9] C.L. Bennett et al., 2003, ApJS, 148, 1

[10] M. B. Green, J. H. Schwarz and E. Witten, *"Superstring Theory"*, Cambridge, UK: Univ. Press, 1987. *Cambridge Monographs On Mathematical Physics.*

[11] F. A. Berezin, *"The Method of Second Quantization"*, Moscow, Nauka, 1965.

[12] P. Fayet and J. Illiopoulos, *Phys. Lett.* **B51** (1974) 461.

[13] L. O'Raifeartaigh, *Nucl. Phys.* **B96** (1975) 331.

[14] H. E. Haber, *"Introductory Low-Energy Supersymmetry"*, Lectures given at TASI 1992, (SCIPP 92/33, 1993), hep-ph/9306207.
D. I. Kazakov, *"Beyond the Standard Model (In search of supersymmetry)"*, Lectures at the European school on high energy physics, CERN-2001-003, hep-ph/0012288.

[15] http://atlasinfo.cern.ch/Atlas/documentation/EDUC/physics14.html

[16] P.Fayet, *Nucl. Phys.* **B90**(1975) 104; A.Salam and J.Srathdee, *Nucl. Phys.* **B87**(1975) 85.

[17] L. Hall, J. Lykken and S. Weinberg, *Phys. Rev.* **D27** (1983) 2359; S. K. Soni and H. A. Weldon, *Phys. Lett.* **B126** (1983) 215; I. Affleck, M. Dine and N. Seiberg, *Nucl. Phys.* **B256** (1985) 557.

[18] H. P. Nilles, *Phys. Lett.* **B115** (1982) 193; A. H. Chamseddine, R. Arnowitt and P. Nath, *Phys. Rev. Lett.* **49** (1982) 970; *Nucl. Phys.* **B227** (1983) 121; R. Barbieri, S. Ferrara and C. A. Savoy, *Phys. Lett.* **B119** (1982) 343.

[19] M. Dine and A. E. Nelson, *Phys. Rev.* **D48** (1993) 1277, M. Dine, A. E. Nelson and Y. Shirman, *Phys. Rev.* **D51** (1995) 1362.

[20] L. Randall and R. Sundrum, *Nucl. Phys.* **B557** (1999) 79; G. F. Giudice, M. A. Luty, H. Murayama and R. Rattazzi, *JHEP*, **9812** (1998) 027.

[21] D. E. Kaplan, G. D. Kribs and M. Schmaltz, *Phys. Rev.* **D62** (2000) 035010; Z. Chacko, M. A. Luty, A. E. Nelson and E. Ponton, *JHEP*, **0001** (2000) 003.

[22] M. E. Peskin, *"Theoretical summary lecture for EPS HEP99"*, hep-ph/0002041.

[23] G. G. Ross and R. G. Roberts, *Nucl. Phys.* **B377** (1992) 571.
V. Barger, M. S. Berger and P. Ohmann, *Phys. Rev.* **D47** (1993) 1093.

[24] W. de Boer, R. Ehret and D. Kazakov, *Z. Phys.* **C67** (1995) 647;
W. de Boer et al., *Z. Phys.* **C71** (1996) 415.

[25] L. E. Ibáñez, C. Lopéz and C. Muñoz, *Nucl. Phys.* **B256** (1985) 218.

[26] W. Barger, M. Berger, P. Ohman, *Phys. Rev.* **D49** (1994) 4908.

[27] V.Barger, M.S. Berger, P.Ohmann and R.Phillips, *Phys. Lett.* **B314** (1993) 351.
P. Langacker and N. Polonsky, *Phys. Rev.* **D49** (1994) 1454.
S. Kelley, J. L. Lopez and D.V. Nanopoulos, *Phys. Lett.* **B274** (1992) 387.

[28] S.Ahmed et al. (CLEO Collaboration), CLEO CONF 99/10, hep-ex/9908022.

[29] R. Barate et al. (ALEPH Collaboration), *Phys. Lett.* **B429** (1998) 169.

[30] W.de Boer, M.Huber, C.Sander, D.I.Kazakov, *Phys.Lett.* **B515** (2001) 283.

[31] ALEPH Collaboration, *Phys.Lett.* **B499** (2001) 67.

[32] S. Abel *et al.* [SUGRA Working Group Collaboration], *Report of the SUGRA working group for run II of the Tevatron*, hep-ph/0003154.

[33] M. Drees and M. M. Nojiri, *Phys. Rev.* **D47** (1993) 376;
J. L. Lopez, D. V. Nanopoulos and H. Pois, *Phys. Rev.* **D47** (1993) 2468;
P. Nath and R. Arnowitt, *Phys. Rev. Lett.* **70** (1993) 3696.

[34] W.de Boer, H.J.Grimm, A.Gladyshev, D.Kazakov, *Phys. Lett.* **B438** (1998) 281.

[35] W.de Boer, M.Huber, A.Gladyshev, D.Kazakov, *Eur.Phys.J.* **C20** (2001) 689.

[36] S. Heinemeyer, W. Hollik and G. Weiglein, *Phys. Lett.* **B455** (1999) 179; *Eur. Phys. J.* **C9** (1999) 343.

[37] T.Kamon, hep-ex/0301019, Proc. of IX Int. Conf. "SUSY-01", WS 2001, p.196.

[38] CDF Collaboration (D. Acosta et al.), *Phys.Rev.Lett.* **90** (2003) 251801; CDF Collaboration (T. Affolder et al.), *Phys.Rev.Lett.* **87** (2003) 251803.

[39] http://CMSinfo.cern.ch/Welcome.html/CMSdocuments/CMSplots

[40] N.V.Krasnikov and V.A.Matveev, "Search for new physics at LHC", hep-ph/0309200.

GAUGINO CONDENSATION AND SUSY BREAKDOWN

Hans Peter Nilles
Physikalisches Institut, Universität Bonn,
Nussallee 12, D-53115 Bonn, Germany
nilles@th.physik.uni-bonn.de

Abstract We review the mechanism of gaugino condensation in the framework of the $d = 10$ heterotic string and its $d = 11$ extension of Horava and Witten. In particular we emphasize the relation between the gaugino condensate and the flux of the antisymmetric tensor fields of higher dimensional supergravity. Its potential role for supersymmetry breakdown and moduli stabilization is investigated.

Keywords: Supersymmetry, Supergravity, String theory, Gaugino condensation, Spontaneous breakdown of supersymmetry

I. Introduction

The topic of my lectures at the Cargèse summer school 2003 was a general introduction to the breakdown of supersymmetry in field- and string-theory. In this written up version I decided to concentrate on some particular aspects of this mechanism: gaugino condensation in the framework of heterotic string- and M-theory. This allows a more detailed discussion of supergravity in $d = 10$ and $d = 11$ dimensions and the reduction to the $d = 4$ case.

The mechanism of gaugino condensation is believed to play a crucial role for moduli stabilization and SUSY breakdown in string theory. Conceived as a mechanism for hidden sector supersymmetry breakdown in supergravity extensions of the standard model of strong and electroweak interactions [93, 94, 40] it found a natural setting in the framework of

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 397–433.

the $E_8 \times E_8$ heterotic string theory [47] (see ref. [28, 32]) as well as the M-theory of Horava and Witten [54] (suggested in [53, 102]).

One of the attractive features of the mechanism is a specific cancellation of $H-$flux and the gaugino condensate in the low energy effective potential. This allows a somewhat controlled discussion of the vacuum energy at least at the classical level and it is the direct consequence of the properties of the higher dimensional supergravity action. It also emphasizes the importance of Chern-Simons-terms in $H-$flux of $d = 10$ supergravity [29].

In these lectures we shall not attempt to construct a fully realistic model but try to explain the mechanism in its simplest form. In section 2 we will discuss supergravity in $d = 10$ and then define the supermultiplets relevant for the $d = 4$ discussion. The mechanism of gaugino condensation is introduced in section 4 followed by a determination of the $d = 4$ effective action using the method of reduction and truncation [116]. Section 6 contains some aspects of the theory beyond the classical level. In section 7 we give a detailed discussion of the $d = 11$ heterotic M-theory and the determination of the low energy effective actions. Some aspects of the mechanism that were rather obscure in the $d = 10$ theory (like the cancellation of gaugino bilinears due to Chern-Simons flux) become obvious in this generalized picture [103]. Section 8 will then focus on the specific properties of supersymmetry breakdown at the hidden wall. The last section discusses some recent developments that lead to a revived interest of this mechanism during the last year.

Before we start let me make a technical comment. Traditionally there are two different ways to include a gaugino condensate in the effective action. The first one [40, 28] uses explicitly the $F-$terms of the supersymmetry transformation laws. These include gaugino bilinears multiplied by the derivative of the gauge kinetic function. These gaugino bilinears are then replaced by the (field dependent) renormalization group invariant scale and included in the standard fashion in the scalar potential. This is the procedure which we follow in these lectures. An alternative method [32, 29] postulates the gaugino bilinear as a new term in an effective superpotential. The qualitative results of the two mechanisms are the same. Some quantitative differences will be mentioned where it applies.

II. Supergravity in $d = 10$

We shall here discuss the effective action of superstring theories in the supergravity field theory framework. For the known ($N = 1$) superstring theories this is $N = 1$ supergravity in $d = 10$ coupled to pure $E_8 \times E_8$

or 0(32) gauge multiplets. The spectrum of this theory is given by the supergravity multiplet $(g_{MN}, \psi_{M\alpha}, B_{MN}, \lambda_\alpha, \varphi)$ where $M, N = 0, \ldots, 9$ are world indices and α is a Majorana-Weyl spinor index, as well as the gauge multiplet (A_M^A, χ_α^A) where $A = 1, \ldots, 496$ lables the adjoint representation of $E_8 \times E_8$ or 0(32). In the Type I theory, these correspond to the massless closed (open) string states respectively. The action of such a theory, including terms up to two derivatives, is unique and given by [16]:

$$\begin{aligned} e_{10}^{-1}\mathcal{L} = & -\frac{1}{2}R - \frac{i}{2}\bar{\psi}^M \Gamma^{MNP} D_N(\omega)\psi_P + \frac{9}{16}\left(\frac{\partial_M \varphi}{\varphi}\right)^2 + \\ & + \frac{3}{4}\varphi^{-3/2} H_{MNP} H^{MNP} + \frac{i}{2}\bar{\lambda}\Gamma^M D_M \lambda + \frac{3\sqrt{2}}{8}\bar{\psi}_M \left(\frac{\Gamma^P \partial_P \varphi}{\varphi}\right)\Gamma^M \lambda \\ & - \frac{\sqrt{2}}{16}\varphi^{-3/4} H_{MNP}\left(i\bar{\psi}_Q \Gamma^{QMNPR}\psi_R + 6i\bar{\psi}^M \Gamma^N \psi^P + \right. \\ & \left. + \sqrt{2}\bar{\psi}_Q \Gamma^{MNP}\Gamma^Q \lambda - i\bar{\chi}\Gamma^{MNP}\chi\right) - \\ & - \frac{1}{4}\varphi^{-3/4} F_{MN} F^{MN} + \frac{i}{2}\bar{\chi}\Gamma^M D_M(\omega)\chi - \\ & - \frac{i}{4}\varphi^{-3/8}\left(\bar{\chi}\Gamma^M \Gamma^{NP} F_{NP}\right)\left(\psi_M + \frac{i\sqrt{2}}{12}\Gamma_M \lambda\right) \\ & + \text{four fermion interactions} \end{aligned} \tag{II.1}$$

where Γ denote Dirac matrices in $d = 10$ and

$$F_{MN}^A = \frac{1}{2}\partial_{[M} A_{N]}^A + f^{ABC} A_M^B A_N^C \tag{II.2}$$

(written for short as $F = dA + A^2$) denotes the gauge field strength. Supersymmetry requires the field strength H_{MNP} of the antisymmetric tensor field B_{MN} not just to be the curl of B, but

$$H_{MNP}^A = \partial_{[M} B_{NP]} + \omega_{MNP}^{YM} \tag{II.3}$$

where the Chern-Simons term is given by

$$\omega^{YM} = Tr\left(AF - \frac{2}{3}A^3\right) \tag{II.4}$$

i.e., B_{NP} has to transform non-trivially under the $E_8 \times E_8$ [or 0(32)] gauge transformations. This theory as it stands has gravitational anomalies and is too naive an approximation to the anomaly-free superstring theory. The absence of anomalies requires an additional term to (3)[50]:

$$H = dB + \omega^{YM} - \omega^L \tag{II.5}$$

with

$$\omega^L = Tr(\omega R - \frac{2}{3}\omega^3) \tag{II.6}$$

where ω_M^{ab} is the spin connection. ω contains a derivative, thus ω^L contains three and appears squared in the action. This term is purely bosonic and for a supersymmetric action requires additional terms which up to now are only partially known. The action in (1) thus requires further terms in order to be an adequate low-energy limit of string theory. The action (1) was derived by truncating all heavy string states. For a better approximation they should be integrated out, leaving a low-energy theory with higher derivatives and terms in a higher order in α' (the slope parameter). These terms appear in what is usually called "σ -model perturbation theory", not to be confused with the string loop expansion, which, at least in the heterotic case, is an expansion in g, the gauge coupling constant. This expansion in powers of α' is classical at the string level. There might also be world-sheet non-perturbative effects that play a role at this classical level. Looking at (1), one might wonder what g (the gauge coupling constant) is. Observe that the gauge fields have non-minimal gauge kinetic terms. Here g is not an input parameter, but g will be determined dynamically

$$\frac{1}{g^2} =< \varphi^{-3/4} > \tag{II.7}$$

consistent with the expectations in string theory. We have to be aware of the fact that the coupling constant as determined by this naive approximation might be different from that determined by string theory. This approximation is probably only useful in defining the important interactions at low energies. In order to ask more fundamental questions, like the determination of the fundamental coupling constants, the approximation probably has to be improved. This can already be seen when we discuss compactification. One possible way is to compactify on a six-torus T^6 , leading to $N = 4$ supergravity in $d = 4$, which does not resemble known $d = 4$ phenomenology. One might therefore ask the question for more non-trivial compactifications (still postponing the question of why these should be more likely than the trivial ones). Defining $\phi = (3/4)\log\varphi$ and neglecting fermionic terms, the equation of motion for ϕ is:

$$\Box\phi = \exp(-\phi)\left[F_{MN}^2 + \exp(-\phi)H_{MNP}^2\right] \tag{II.8}$$

Integrating $\Box\phi$ over a compact manifold without boundary leads to a vanishing result. The right-hand side is positive definite and therefore has to vanish. This implies trivial compactification unless $\phi \to \infty$, which

is outside the validity of our approximation. The addition of ω^L in H does not change the situation, but this term requires supersymmetric completion which necessitates the presence of R^2 terms. They actually appear in the Euler combination

$$-\exp(-\phi)\left[R^2_{MNPQ} - 4R^2_{MN} + R^2\right] \tag{II.9}$$

on the right-hand side of (8), ensuring the absence of ghosts. With these terms from the α' expansion, non-trivial compactification is possible: R^2 can be compensated by F^2, and this implies a breakdown of gauge symmetries in the presence of compactification [14]. Notice, however, that the scale of compactification is not yet fixed. There exists an independent argument confirming this result. For the H field to be well defined, the integral of the curl of H over a compact manifold without boundary should vanish:

$$\int\limits_{C_4} dH = \int\limits_{C_4} [TrF \wedge F - TrR \wedge R] = 0 \tag{II.10}$$

leading to a compensation of F and R in extra dimensions (for details of this discussion please consult the available textbooks [46, 107]).These results are very encouraging. If $E_8 \times E_8$ or $0(32)$ were to remain unbroken in $d = 4$, they would not be able to lead to chiral fermions. The discussed constraints involve integrated quantities and could have various solutions. Only the simplest possibility – a vanishing integrand – can be studied easily [116]. It implies a direct identification of F and R. The spin connection $\omega_m^{ab}(m = 4, \ldots, 9; a, b = 1. \ldots, 6)$ can be viewed as a gauge field of an $0(6)$ subgroup of the Lorentz group $0(9, 1)$, identified with A_m^A in an $0(6)$ subgroup of $E_8 \times E_8$ or $0(32)$ in order to fulfil the constraints. The question of a remaining supersymmetry in $d = 4$ is related to the holonomy group of the compact manifold, which in turn is a subgroup of $0(6)$. I shall not explain this relation here in detail, but just give a heuristic argument. The gravitino ψ_M^α transforms like a 4 of $0(6)$. $N = 1$ supersymmetry will be present in $d = 4$ if the decomposition of the 4 with respect to the holonomy group contains exactly one singlet. If there are more singlets, one will have extended supersymmetries, e.g., in the case of the torus the holonomy group is trivial and $4 = 1+1+1+1$, resulting in $N = 4$ supersymmetry. The simplest choice for $N = 1$ is to have $SU(3)$ holonomy, which leads to $4 = 1 + 3$ and $6 = 3 + \bar{3}$, and is used in the Calabi-Yau approach. But there are certainly more possibilities, even with discrete subgroups of $SU(3)$ corresponding to certain orbifolds. For simplicity, I shall here assume $SU(3)$ holonomy. With this identification of ω and A at least one $SU(3)$ subgroup of $0(32)$ or $E_8 \times E_8$

will break down during compactification. In the case of 0(32), this will lead to $0(26)\times U(1)$ with possible zero modes in the decomposition of the adjoint of 0(32), giving exclusively real representations of 0(32). Based on this argument, one usually concludes that 0(32) will not lead to a phenomenologically successful model, although not all possibilities have yet been studied. The situation in the case of $E_8 \times E_8$ looks better. A decomposition of the adjoint of E_8 with respect to $E_6 \times SU(3)$ leads to $248 = (78,1)+(27,3)+(\overline{27},\bar{3}+(1,8)$ and contains chiral representations. Moreover, E_6 is one of the more successful candidates for a grand unified gauge group with a family of quarks and leptons in 27, the number of these zero modes being defined by topological properties of the compact manifold. Here is then the usual starting point for the construction of "superstring-inspired models".

III. Towards $d = 4$

We have first to discuss the possible zero modes. Let us define indices $M = (\mu, m)$ $(\mu = 0,\ldots,3;\ m = 4,\ldots,9)$ and start with the metric

$$g_{MN} = \left(\begin{array}{c|c} g_6^{-1/2}\hat{g}_{\mu\nu} & \\ \hline & g_{mn} \end{array}\right) \tag{III.11}$$

where $g_6 = \det g_{mn}$ is used to redefine $g_{\mu\nu}$ in order to have usual kinetic terms for the graviton. The integral over extra dimensions

$$\int d^6y\sqrt{-g_6} = R_c^6 \sim \frac{1}{M_c^6} \tag{III.12}$$

defines the average radius of compactification. Defining then $g_{mn} = \exp(\sigma)\hat{g}_{mn}$ with a scalar field σ, one can normalize $\int d^6y\sqrt{-\hat{g}_6} = M_P^{-6}$ and $\exp(\sigma)$ defines the radius of compactification in units of the Planck length. Depending on the topological properties of the manifold, g_{mn} gives rise to zero modes that are scalars in $d = 4$ (we will not discuss off-diagonal terms in g_{MN} like $g_{\mu m}$ that give rise to gauge bosons depending on the isometries of the manifold). g_{mn} corresponds to a symmetric tensor of 0(6) with respect to the $SU(3)$ subgroup discussed earlier; we have $21 = 1 + 8 + 6 + \bar{6}$. With the notation $m = (i, \bar{j})$, the latter correspond to modes of $g_{i\bar{j}}, g_{ij}, g_{\bar{i}\bar{j}}$, while σ is the singlet.

Turning to the gravitino ψ_M^α, we can view α as an eight-dimensional index which transforms as a 4 of 0(6) and a Weyl spinor of $0(3,1)$. ψ_μ^α corresponds to spin-3/2 particles in $d = 4$ with $N_{max} = 4$ as already discussed. ψ_m^α can give rise to spin-$\frac{1}{2}$ zero modes. To obtain canonical kinetic terms for the gravitino, as in the case of the metric, a rescaling

$$\tilde{\psi}_\mu = \exp(-3\sigma/4)\psi_\mu \tag{III.13}$$

is required.

The antisymmetric tensor field B_{MN} could give rise to $B_{\mu\nu}, B_{m\nu}$ and B_{mn} (corresponding to the Betti numbers b_0, b_1 and b_2). A zero mode from $B_{\mu\nu}$ corresponds to one pseudoscalar degree of freedom θ defined through a duality transformation

$$H_{\mu\nu\varrho}\epsilon^{\mu\nu\varrho\sigma} = \varphi^{3/2} exp(-6\sigma)\partial^{\sigma}\theta + \ldots \qquad \text{(III.14)}$$

$B_{m\nu}$ could give rise to extra gauge bosons which (although possibly interesting) we shall not discuss here. B_{mn} will again correspond to pseudoscalars in $d = 4$. A decomposition with respect to $SU(3)$ gives $15 = 1 + 3 + \bar{3} + 8$ with the singlet corresponding to the "trace" $\eta = \epsilon^{mn}B_{mn}$ and $B_{\bar{i}\bar{j}}$, and B_{ij} and $B_{i\bar{j}}$ corresponding to $3, \bar{3}$, and 8 respectively. All these modes appear in the action only through the field strength H implying derivative couplings, i.e., they show axion-like behaviour. From the λ, ϕ members of the supergravity multiplet, we expect additional spin $\frac{1}{2}(0)$ particles in $d = 4$.

The discussion of the zero modes of A_M^A involves some complication because of the identification of ω_m^{ab} and A_m^A in an $SU(3)$ subgroup. A_μ^A will of course, give rise to gauge bosons in the adjoint representations of the unbroken gauge group, e.g., $A = 1, \ldots, 78$ for E_6. A_m^A will give rise to scalars in $d = 4$, and we are mostly interested in those transforming as 27 (or $\overline{27}$) under E_6. Let us therefore write $A = (a, i)$ or $(\bar{a}, \bar{i})$ $a = 1, \cdots, 27$. The states $C^b = A_{\bar{i}}^{b,i}$ and $B^{\bar{b}} = A_i^{\bar{b},\bar{i}}$ then transform as $27, \overline{27}$ with respect to E_6 and are singlets under the diagonal subgroup $SU(3)$ of the product of $SU(3) \subset 0(6)$ and $SU(3) \subset E_8$. These bosons will have supersymmetric partners from the zero modes of χ_α^A. The number of the possible zero model is of course entirely defined by the topological properties of the manifold under consideration.

We can now have a first look at the possible interactions of these zero model in $d = 4$ starting from the $d = 10$ action given in (1). Of course, in general we expect here not only the influence of topological properties, but also the explicit form of the metric of the compact manifold will become important. Nonetheless we will be able to obtain some non-trivial results that are rather independent of the special form of the metric. We will do that exclusively in the framework of $N = 1$ supergravity in $d = 4$, firstly because of the reasons given in Section 2, and secondly because this theory is simpler than the non-supersymmetric case.

$N = 1$ supergravity in $d = 4$ (with action including terms up to two derivatives [21]) is defined through two complex-valued functions f, G of the chiral superfields ϕ_i. The first is an analytic function $f(\phi_i)$ defining the gauge kinetic terms $f(\phi_i)W^\alpha W_\alpha$. In a component language,

f appears in many places, but it can be extracted most efficiently from

$$Ref(\varphi_i)F_{\mu\nu}F^{\mu\nu} + Imf(\varphi_i)\epsilon_{\mu\nu\varrho\sigma}F^{\mu\nu}F^{\varrho\sigma} \tag{III.15}$$

where φ_i denotes the (complex) scalar component of the superfield ϕ_i. The second function is the so-called Kähler potential

$$G(\phi_i, \phi_i^*) = K(\phi_i, \phi_i^*) + \log|W(\phi_i)|^2 \tag{III.16}$$

Unlike f, G is not analytic and contains the left-handed chiral superfields along with their complex conjugates. The second term in (16) contains the analytic function $W(\phi_i)$: the superpotential. The action in component form usually contains G in complicated form; the scalar kinetic terms, e.g., are

$$G_i^j(\partial_\mu\varphi^i)(\partial^\mu\varphi_j^*); \;\; G_i^j \equiv \frac{\partial^2 G}{\partial\varphi^i\partial\varphi_j^*} \tag{III.17}$$

whereas the scalar potential is given by

$$V = \exp(G)[G_k(G^{-1})_l^k G^l - 3]; \;\; G_k \equiv \frac{\partial G}{\partial\varphi^k}; \;\; G^l \equiv \frac{\partial G}{\partial\varphi_l^*} \tag{III.18}$$

which makes it difficult to extract G once an action is given in component form. There is only one term which allows a rather simple identification of G, and this is a term involving the gravitino

$$e_4 \exp(G/2)\,\bar{\psi}_\mu\gamma^{\mu\nu}\gamma_5\psi_\nu \tag{III.19}$$

which will later be used extensively after the correct redefinitions of the gravitino in $d = 4$ have been performed. Let us now consider the action in $d = 10$ in order to learn something about the possible action in $d = 4$. We start with the gauge kinetic term

$$e_{10}\varphi^{-3/4}F_{MN}F^{MN}. \tag{III.20}$$

Since we are interested in the $F_{\mu\nu}^2$ part, we write

$$e_4 e_6 \varphi^{-3/4}F_{\mu\nu}F_{\varrho\sigma}g^{\mu\varrho}g^{\nu\sigma} \tag{III.21}$$

where, with the definitions given earlier, we would like to extract f from

$$\hat{e}_4 Ref\, F_{\mu\nu}F^{\mu\nu} \tag{III.22}$$

with $\hat{e}_4 = (det\ \hat{g}_{\mu\nu})^{\frac{1}{2}} = \exp(6\sigma)e_4$, and indices are contracted with the "hatted" metric. Integrating the extra six dimensions with the normalization given in (13) using $M_P \equiv 1$, we obtain

$$ReS \equiv Ref = \varphi^{-3/4}\exp(3\sigma) \tag{III.23}$$

as the real part S of the scalar component of a chiral superfield (that will later also be denoted by S). This is a rather amazing result. Remember that at no point in the derivation did we have to know something about the metric of the compact six-dimensional space, so this constitutes a rather model-independent result. Observe that f is usually non-trivial, that its vacuum expectation value (vev) will determine the gauge coupling constant, and that the couplings of E_8 (or E_6) and E_8' coincide.

Let us now discuss the imaginary part of f, to be extracted from $F_{\mu\nu}F_{\rho\sigma}\varepsilon^{\mu\nu\rho\sigma}$. The relevant degree of freedom comes from $B_{\mu\nu}$ as discussed earlier. $B_{\mu\nu}$ couples only through its field strength $H_{\mu\nu\rho}$ and has therefore only derivative couplings. Taking the relevant terms in the $d = 10$ action and integrating the extra dimensions, we obtain

$$\varphi^{-3/2}\exp(6\sigma)H_{\mu\nu\varrho}H^{\mu\nu\varrho} + H_{\mu\nu\varrho}O^{\mu\nu\varrho} \tag{III.24}$$

where $O^{\mu\nu\varrho}$ contains fermion bilinears. H has to satisfy a constraint (neglecting R^2-terms for the moment)

$$\partial_{[\mu}H_{\nu\varrho\sigma]} = -TrF_{[\mu\nu}F_{\varrho\sigma]} \tag{III.25}$$

which we take into account by adding a Lagrange multiplier

$$\theta\epsilon^{\mu\nu\varrho\sigma}\left(\partial_\mu H_{\nu\varrho\sigma} + TrF_{\mu\nu}F_{\varrho\sigma}\right) \tag{III.26}$$

Next we eliminate H via the equations of motion and arrive at an action containing the terms

$$\varphi^{3/2}\exp(-6\sigma)(\partial_\mu\theta)^2 + \theta\epsilon^{\mu\nu\varrho\sigma}Tr(F_{\mu\nu}F_{\varrho\sigma}) \tag{III.27}$$

which tells us that $Imf = \theta$, and for the scalar component of S we obtain

$$S = \varphi^{-3/4}\exp(+3\sigma) + i\theta \tag{III.28}$$

as a mixture of g_{MN} and B_{MN} zero modes. The partner is a combination of ψ_m and λ zero modes which we will not discuss here in detail. Observe that θ couples only with derivatives except for the last term in (27), and that the $d = 4$ action, has a Peccei-Quinn-like symmetry under shifts of θ by a real constant, thus θ couples like an axion. Let me stress again that all these statements about the action and the form of (28) are model-independent and could be derived without explicit knowledge of the metric.

Unfortunately, the situation changes once we try to extract the Kähler potential. As already indicated, the term to investigate is the $d = 4$ "gravitino mass term" (19). The extraction of this term is rather complicated due to several redefinitions of the gravitino field. A general form

has been given in [29]), and we will not repeat the derivation here. Many of the terms appearing there depend explicitly on the metric and spin-connection of the six-dimensional compact space. A model-independent statement can only be made about the structure of the superpotential, because it is an analytic function in the chiral superfields. Symbolically the "gravitino mass term" is obtained as

$$\exp(G/2) = \varphi^{-3/4} \exp(-3\sigma)\Gamma^{mnp} H_{mnp} \qquad \text{(III.29)}$$

and from (16) we can try to read off the superpotential. $W(\phi_i)$ is defined to be an analytic function in the chiral superfields and should not contain derivatives. A first inspection of (29) therefore suggests that a possible candidate for a superpotential is the A^3 term contained in the Yang-Mills Chern-Simons term (4) included in H. This then gives rise to a trilinear superpotential involving the C and B fields defined earlier. At the moment it is not clear whether these are the only possible terms in the superpotential, although at the classical level this seems to be the complete expression. Observe that, for example, the superfield S as defined in (28) cannot appear in the superpotential, since its pseudoscalar component has only derivative couplings. We will come back to these points later. In any case, a more detailed discussion of the Kähler potential requires more information (or approximations) about the $d = 6$ metric. Before we tackle this topic, let me first present a discussion about supersymmetry breakdown in $d = 4$.

IV. Gaugino condensation and supersymmetry breakdown

$N = 1$ supergravity in $d = 4$ still needs the incorporation of supersymmetry breakdown at a scale small compared to the Planck mass. For the phenomenological reasons mentioned earlier, this should appear in a hidden sector only coupled gravitationally to the observable sector. Some superstring models contain such a hidden sector, e.g. the sector that contains the particles transforming non-trivially under the second E_8. Notice that the observable sector (for definiteness called the E_6 sector) only couples gravitationally to the E_8' sector (there are no particles that transform non-trivially both under E_6 and E_8'). Moreover, the E_8' sector contains a $d = 10$ pure super-Yang-Mills multiplet, suggesting a possible breakdown of supersymmetry via gaugino condensates. This breakdown has already been discussed in the framework of supergravity models, both at the level of an effective Lagrangian [93] and at the level of the complete classical action [40]. Assume asymptotically-free gauge

interactions (here E_8' or a subgroup thereof) with a scale

$$\Lambda = \mu \exp\left(-1/b_0 g^2(\mu)\right) \tag{IV.30}$$

which is renormalization-group invariant at the level of the one-loop β-function. In analogy to QCD, which leads to $q\bar{q}$ condensates, we will here assume that the gauge fermions condense at a scale

$$\langle \chi\chi \rangle = \Lambda^3 \tag{IV.31}$$

As long as Λ is small compared to M_p, we assume that gravity will not qualitatively disturb this dynamical mechanism. The question whether such a condensate breaks supersymmetry can be studied by investigating the supersymmetry transformation laws of the fermionic fields of the theory. The non-derivative terms in these transformations will give us the auxiliary fields that serve as order parameters for supersymmetry breakdown. The relevant objects here are the auxiliary fields F_k of the chiral superfields ϕ_k,

$$F_k = \exp(G/2) G_k - \frac{1}{4} f_k(\chi\chi) + \ldots \tag{IV.32}$$

where f is the gauge kinetic function discussed earlier and f_k is its derivative with respect to ϕ_k. A necessary condition for the breakdown of supersymmetry via gaugino condensates is therefore a non-trivial f-function. This condition is fulfilled in the framework of superstring-inspired models [28] [32], since we have seen in the last section that $f = S$ in a rather model-independent way. Whether this is also sufficient for the breakdown of supersymmetry can only be checked by minimizing the potential

$$V = F_k \left(G^{-1}\right)_\ell^k F^\ell - 3 \exp(G) \tag{IV.33}$$

since the different terms in (32) might cancel at the minimum. But let us for the moment assume that only the second term in (32) receives a *vev*. Since $f_s = 1$ in units of M_p, we find a supersymmetry breakdown scale

$$\langle F_s \rangle = M_s^2 \approx \Lambda^3 / M_p \tag{IV.34}$$

and a scale of $\Lambda \sim 10^{13}$ GeV would lead to a gravitino mass in the TeV range. Once we understand why Λ is five orders of magnitude smaller than M_p, we shall understand why $m_{3/2}/M_p \sim 10^{-15}$. Λ now depends on the E_8' gauge coupling and the spectrum of low-energy modes. Identifying g_6 with g_8 would in many circumstances lead to too large a value for Λ, and one might speculate that E_8' should break during compactification. We shall, however, see later that the equality of g_6 and g_8 seems

to be only an artifact of the classical approximation, which is not true in the full theory. Thus the shadow E_8' (or a subgroup thereof) sector of the superstring takes the role of the hidden sector of supergravity models and might explain the smallness of $m_{3/2}$ compared to M_p. But how does this breakdown of SUSY in the hidden sector influence the observable sector? In general, we would expect gaugino masses ($m_{1/2}$), scalar masses (m_0) and the trilinear couplings (Am) to be of the order of magnitude of $m_{3/2}$. A naive inspection shows that this might also be true here. Gaugino masses in the observables sector are in general given by

$$m_{1/2} = f_k \left(G^{-1}\right)_\ell^k F^\ell \tag{IV.35}$$

where f is the gauge kinetic function of the observable sector. With $F^\ell = (1/4) f^\ell < \chi\chi >$ we would therefore obtain $m_{1/2} \sim m_{3/2}$. In the same way we would obtain under these circumstances the soft trilinear couplings $A \sim 1$ and scalar masses of order $m_{3/2}$. To make a quantitative statement about the soft parameters we need a better understanding of the Kähler potential, a question which we want to discuss in the next section. We need this in order to study the explicit form of the effective potential and finds its minimum. This then also has to determine the exact value of Λ in (IV.30) which depends on the coupling constant and is fixed only after the value of g is known.

V. Reduction and Truncation

A first approximation for G (that might simulate an orbifold approximation of interest in this context) is obtained through reduction and truncation [120]. One first compactifies the $d = 10$ theory on a six-torus T^6. The resulting theory is $N = 4$ supersymmetric in $d = 4$. From this theory one truncates unwanted states, to obtain an $N = 1$ theory. From the gauge singlet sector one keeps only those states that transform as singlets under an $SU(3) \subset 0(6)$ of the Lorentz group. Since ψ_μ^α transforms as a 4 of 0(6) and thus as $1 + 3$ under $SU(3)$, we remain with one gravitino. As already explained in Section 3, there are only a few gauge singlets that survive this truncation. For the bosonic modes we have φ, σ from the metric as well as θ and η from the antisymmetric tensor. For the gauge non-singlet fields one has to remember the identification of spin-connection and gauge fields. Here one keeps those states which are singlets under the diagonal subgroup of the product of $SU(3) \subset 0(6)$ and $SU(3) \subset E_8$. This leaves us with one 27 of E_6 in this case, corresponding to $C^b = A_{\bar{i}}^{b,i}$; ($b = 1, \ldots, 27$, cf. Section 3). With this well-defined procedure based on simple reduction on T^6, the component Lagrangian in $d = 4$ can be deduced. From this we can immediately read off $f = S$

and $W = d_{abc}C^aC^bC^c$, which should not be surprising. Moreover, from the "gravitino mass term" formula (29) one obtains

$$G = \log\left(e^{-6\sigma}\varphi^{-3/2}\right) + \log|W|^2 \tag{V.36}$$

The components φ and σ should correxpond to lowest components of chiral superfields. One combination $S = \varphi^{-3/4}\exp(3\sigma) + i\theta$ has already been defined earlier. To define the other combination, the information from (36) is not enough. The charged fields C do not yet appear in the first term of G in (36) and the correct definition of the superfields has yet to be found. This can be done, for example, by using the scalar kinetic terms. It leads to a second superfield in which φ, σ and the C-modes mix

$$T = \exp(\sigma)\varphi^{3/4} + |C_a|^2 + i\eta \tag{V.37}$$

where η is the mode from $\epsilon^{mn}B_{mn}$ as discussed after equation (14), and the Kähler potential from (36) thus reads

$$G = -\log(S + S^*) - 3\log\left(T + T^* - 2|C|^2\right) + \log|W|^2 \tag{V.38}$$

a form already previously mentioned in the framework of supergravity models. The scalar potential derived from this G-function has some remarkable properties

$$V = \frac{1}{16st_c^3}\left[|W|^2 + \frac{t_c}{3}|W'|^2\right] + D^2 - \text{terms} \tag{V.39}$$

where $s = ReS$ and $t_c = ReT - |C_a|^2 = t - |C_a|^2$ and W' is the derivative of W with respect to the C-field. The potential is positive definite ($t_c > 0$ is required by the kinetic terms) and has a minimum with vanishing vacuum energy $V = 0$. This minimum is obtained at $W = W' = 0$ independent of the values of s and t. This implies that at this level the gauge coupling constant and the radius of compactification is not yet fixed. The theory has classical symmetries which allow shifts of the values of s and t, as well as Peccei-Quinn symmetries corresponding to shifts in θ and η. This, of course, makes the use of this approximation as an effective low-energy limit of the superstring very problematic. Certain crucial parameters, like the value of the gauge coupling constant and the scale of compactification, which we believe to be dynamically determined in the full string theory, are not yet fixed. To determine these quantities we would need information beyond the truncated theory.

This remains a relevant question when we discuss the effective potential in the presence of a gaugino condensate. Since the gauge coupling

constant is not determined, Λ in (30) is also unknown. Using (33) and (38), we get for the potential

$$V = \frac{1}{16 s t_c^3}\left[|W - 2(st_c)^{3/2}(\chi\chi)|^2 + \frac{t_c}{3}|W'|^2\right] \qquad \text{(V.40)}$$

where $(\chi\chi)$ depends on g^2 through $\exp(-S/b_0)$. The potential is still positive definite and has a minimum at $V = 0$ which is still degenerate. Now the minimum need not necessarily imply $W = W' = 0$, but we could have a non-trivial *vev* of W. Given fixed $\langle W\rangle \neq 0$ by some yet unknown mechanism, the value of the gauge coupling constant would be fixed. The most natural candidate for such a mechanism would be a nontrivial *vev* (so-called flux) of the antisymmetric tensor field H as defined in (5). At first sight one might have conjectured that the appropriate flux would originate from dB, but it was soon realized that $\langle dB\rangle$ is quantized in units of the Planck scale [108]. Thus $\langle dB\rangle$ should vanish in order to allow for a value of $\langle\chi\chi\rangle$ that is small enough to give a reasonable value for the gravitino mass after supersymmetry breakdown. One therefore concluded that it is the flux of the Chern-Simons terms in (5) which is responsible for the non-zero $H-$flux [29]. For these terms the quantization argument of ref. [108] does not apply and acceptable values for $\langle\chi\chi\rangle$ can be obtained. We shall come back to this question in more detail in the framework of the heterotic M-theory. There it will become obvious from theoretical arguments that it is not $\langle dB\rangle$ but the flux from the Chern-Simons terms that compensates the contribution of the gaugino condensate in the effective potential.

In order to minimize the potential, the theory slides to a coupling constant which, through (30), gives a value of the condensate that exactly cancels the contribution of W. In other words, this means that the dilaton S slides to a value that cancels the vacuum energy in the same way as an axion slides to cancel a possible θ-parameter of a gauge theory [observe that $\exp(-S/b_0)$ contains both s and θ]. Although we do not yet understand the magnitude of supersymmetry breakdown, this mechanism to ensure $E_{\text{vacuum}} = 0$ after $SUSY$ breakdown appears very attractive. We shall still need to convince ourselves that supersymmetry is actually broken, since in (40) a certain cancellation of $\langle W\rangle$ and $\langle\chi\chi\rangle$ appears. In fact it tells us that the auxiliary field F_S of the S-superfield vanishes in the vacuum. Nonetheless, here F_T requires a non-vanishing vev once $\langle W\rangle \neq 0$, and supersymmetry is broken

$$F_T = exp(G/2)G_T \neq 0. \qquad \text{(V.41)}$$

In a next step we have to analyze how the breakdown of $SUSY$ is felt in the observable sector, and this is, of course, model dependent.

Gaugino masses, for example, are given by $m_{1/2} = f_k(G^{-1})^k_\ell F^\ell$, and only f_S is different from zero. In the case at hand we therefore obtain $m_{1/2} = 0$. In fact, the same is also true for the scalar masses. This is an artifact of the special model (a so-called no scale model) and would need further discussions. A first question concerns the stability of this result in perturbation theory which we shall investigate in the next section.

Before we do this let us mention a different way to include the gaugino condensate in the low-energy effective potential [32]. Instead of including $(\chi\chi)$ in F_k directly, as in (32), one might postulate a new contribution to the superpotential proportional to $(\chi\chi) \sim \exp(-S/b_0)$. This leads to a potential very similar although not identical to the one given in (40) (for details see [32, 29]).

VI. Beyond the classical level

Vanishing values of the soft parameters in the observable sector might be a result of the symmetries of the theory, and if yes, whether these symmetries hold to all orders in the perturbative loop expansion. For the heterotic string, this loop expansion is governed by the coupling constant g, which in turn is defined through a *vev* of the dilaton field. This will allow us to construct a definite loop expansion in the dilaton field and still give us restrictions on how the classical symmetries are broken by loop effects. But before we discuss the loop expansion in more general terms, let us examine some aspects at the one-loop level. We can do that because of the mechanism of anomaly cancellation in the $d = 10$ field theory. Green and Schwarz have observed that the cancellation of anomalies [50] requires certain new local counterterms with definite finite coefficients in the one-loop effective action to cancel the gauge non-invariance of present non-local terms. In general, such terms appear with infinite coefficients, but the possible symmetry of the effective action forces us to renormalize the theory in such a way that these gauge-variant local counterterms have a well-defined finite coefficient. An example of such a term is

$$\epsilon B_{VO} Tr(F_{LK} F_{SW}) Tr(F_{AG} F_{EN}) \epsilon^{VOLKSWAGEN} \tag{VI.42}$$

where $\epsilon = 1/720(2\pi)^5$. While this gives rise to many new interaction terms in the $d = 4$ theory, one possible manifestation seems to be of particular importance. Replacing one of the TrF^2 terms by their *vev* in extra dimensions, one arrives at

$$\eta \epsilon^{\mu\nu\varrho\sigma} Tr(F_{\mu\nu} F_{\varrho\sigma}). \tag{VI.43}$$

According to (37), η is the imaginary part of T, and unlike in the classical case it now (in addition to θ) couples to $F\tilde{F}$. Observe that (VI.43) is

gauge-invariant, while (VI.42) is not, but is required by the absence of anomalies in $d = 10$. This shows that the remnants of such terms originate in ten dimensions, and are one of the few places where we could in principle observe whether we live in higher dimensions. (VI.43) suggests that not only θ, but also η couples like an axion. To make sure that this does not just lead to a redefinition of θ at the one-loop level, all anomaly cancellation terms have to be considered. Doing this and satisfying $TrF^2 = TrR^2$ in extra dimensions, one arrives at the result that η couples differently to E_6 and E_8':

$$\epsilon\eta[(F\tilde{F})_8 - (F\tilde{F})_6] \quad \text{and} \quad \theta[(F\tilde{F})_8 + (F\tilde{F})_6] \tag{VI.44}$$

leading to different gauge kinetic functions

$$f = S \pm \epsilon T \tag{VI.45}$$

for the different gauge groups.

This fact has interesting consequences, some of which we will now list.

a) The second axion could be a candidate to solve the strong CP problem of QCD in the observable sector. One axion (like θ alone) would not be sufficient, because it is used to adjust the θ-angle of E_8' and becomes massive. For a relatively recent discussion see [43].

b) Supersymmetry requires the same behaviour of the real parts of S and T as that of the imaginary parts; i.e., ReS and ReT couple differently to E_6 and E_8'. Since the $vev's$ of these fields define the gauge coupling constants, g_6 and g_8' need no longer be equal. This might have consequences for the condensation scale of E_8'.

c) There exist now two axion-dilaton pairs, and this might generalize the relaxation of the cosmological constant to the observable sector in the same way as is appears in the hidden sector [71].

d) Imposition of supersymmetry also requires new terms in the Kähler potential at the one-loop level. We will discuss this later.

e) As expected, these effects at the one-loop level lead to an induced breakdown of supersymmetry in the observable sector once it is broken in the hidden sector. Remember our discussion in Section 5, where the observable sector remained supersymmetric. Gaugino masses are given by

$$m_{1/2} \sim F_T f_T + F_S f_S. \tag{VI.46}$$

At tree level we had $F_S = f_T = 0$ and vanishing gaugino masses. But now we have $f = S + \epsilon T$, and f_T no longer vanishes. As a result, non-trivial gaugino masses (and also non-trivial scalar masses and A-parameters) of order $\epsilon m_{3/2}$ are transmitted to the observable sector.

Of course, in case of a nontrivial ϵ, we should go back to the action and see what happens to the one-loop effective potential. In general we shall expect a nontrivial value of the cosmological constant (typically anti-de Sitter) and in some cases even unbroken supersymmetry. So a discussion of a fully realistic model needs more structure than present in the toy model under consideration.

VII. Heterotic M-Theory

With the discovery of string dualities, there has been a revival of the study of those string theories that might eventually become relevant for our discussion of the low-energy effective supergravity theories. From all the new and interesting results in string dualities, it is the heterotic M–theory of Hořava and Witten [54] (that in $d = 11$ could be regarded as the strong coupling limit of $d = 10$ $E_8 \times E_8$ heterotic string theory) which might have a direct impact on the discussion of the phenomenological aspects of these theories. One of the results concerns the question of the unification of all fundamental coupling constants [117] and the second one the properties of the soft terms (especially the gaugino masses) once supersymmetry is broken [102, 103]. As we shall see in both cases, results that appear problematic in the weakly coupled case (as the formerly discussed heterotic string case will be called from now on) get modified in a satisfactory way, while the overall qualitative picture remains essentially unchanged. In these lectures we shall therefore concentrate on these aspects of the new picture.

The heterotic M–theory is an 11–dimensional theory with the $E_8 \times E_8$ gauge fields living on two 10–dimensional boundaries (walls), respectively, while the gravitational fields can propagate in the bulk as well. A $d = 4$ dimensional theory with $N = 1$ supersymmetry emerges at low energies when 6 dimensions are compactified on a Calabi–Yau manifold. The scales of that theory are M_{11}, the $d = 11$ Planck scale, R_{11} the size of the x^{11} interval, and $V \sim R^6$ the volume of the Calabi–Yau manifold. The quantities of interest in $d = 4$, the Planck mass, the GUT–scale and the unified gauge coupling constant α_{GUT} should be determined through these higher dimensional quantities. The fit of ref. [117] identifies $M_{GUT} \sim 3 \cdot 10^{16}$ GeV with the inverse Calabi–Yau radius R^{-1}. Adjusting $\alpha_{GUT} = 1/25$ gives M_{11} to be a few times larger than M_{GUT}.

On the other hand, the fit of the actual value of the Planck scale can be achieved by the choice of R_{11} and, interestingly enough, R_{11} turns out to be an order of magnitude larger than the fundamental length scale M_{11}^{-1}. A satisfactory fit of the $d = 4$ scales is thus possible, in contrast to the case of the weakly coupled heterotic string where, naively, the string scale seems to be a factor 20 larger than M_{GUT}.

VII.1 The action in $d = 11$

The effective action of the strongly coupled $E_8 \times E_8$ – M–theory in the "downstairs" approach is given by [54] (we take into account the numerical corrections found in [20])

$$\begin{aligned} L \;=\; & \frac{1}{\kappa^2}\int_{M^{11}} d^{11}x\sqrt{g}\left[-\frac{1}{2}R - \frac{1}{2}\bar{\psi}_I\Gamma^{IJK}D_J\psi_K - \frac{1}{48}G_{IJKL}G^{IJKL}\right. \\ & -\frac{\sqrt{2}}{384}\left(\bar{\psi}_I\Gamma^{IJKLMN}\psi_N + 12\bar{\psi}^J\Gamma^{KL}\psi^M\right)\left(G_{JKLM} + \hat{G}_{JKLM}\right) \\ & \left. - \frac{\sqrt{2}}{3456}\epsilon^{I_1I_2\ldots I_{11}}C_{I_1I_2I_3}G_{I_4\ldots I_7}G_{I_8\ldots I_{11}}\right] \qquad \text{(VII.47)} \\ + \; & \frac{1}{4\pi(4\pi\kappa^2)^{2/3}}\int_{M_i^{10}} d^{10}x\sqrt{g}\left[-\frac{1}{4}F^a_{iAB}F_i^{aAB} - \frac{1}{2}\bar{\chi}^a_i\Gamma^A D_A(\hat{\Omega})\chi^a_i\right. \\ & \left. -\frac{1}{8}\bar{\psi}_A\Gamma^{BC}\Gamma^A\left(F^a_{iBC} + \hat{F}^a_{iBC}\right)\chi^a_i + \frac{\sqrt{2}}{48}\left(\bar{\chi}^a_i\Gamma^{ABC}\chi^a_i\right)\hat{G}_{ABC11}\right] \end{aligned}$$

where M^{11} is the $d = 11$ manifold and M_i^{10} (with $i = 1, 2$) its 10–dimensional boundaries. In the lowest approximation M^{11} is just a product $M^4 \times X^6 \times S^1/Z_2$. Compactifying to $d = 4$ in such an approximation we obtain [117, 20]

$$G_N = \frac{\kappa_4^2}{8\pi} = \frac{\kappa^2}{8\pi R_{11}V}\,, \qquad \text{(VII.48)}$$

$$\alpha_{GUT} = \frac{(4\pi\kappa^2)^{2/3}}{V} \qquad \text{(VII.49)}$$

with V the volume of the Calabi–Yau manifold X^6 and $R_{11} = \pi\rho$ the S^1/Z_2 length.

The fundamental mass scale of the 11–dimensional theory is given by $M_{11} = \kappa^{-2/9}$. Let us see which value of M_{11} is favoured in a phenomenological application. For that purpose we identify the Calabi–Yau volume V with the GUT–scale: $V \sim (M_{GUT})^{-6}$. From (VII.49) and the value of

$\alpha_{GUT} = 1/25$ at the grand unified scale, we can then deduce the value of M_{11}

$$V^{1/6} M_{11} = (4\pi)^{1/9} \alpha_{GUT}^{-1/6} \approx 2.3\,, \tag{VII.50}$$

to be a few times larger than the GUT–scale. In a next step we can now adjust the gravitational coupling constant by choosing the appropriate value of R_{11} using (VII.48). This leads to

$$R_{11} M_{11} = \left(\frac{M_{Planck}}{M_{11}}\right)^2 \frac{\alpha_{GUT}}{8\pi(4\pi)^{2/3}} \approx 2.9 \cdot 10^{-4} \left(\frac{M_{Planck}}{M_{11}}\right)^2 . \tag{VII.51}$$

This simple analysis tells us the following:

- In contrast to the weakly coupled case, the correct value of M_{Planck} can be fitted by adjusting the value of R_{11}.
- The numerical value of R_{11}^{-1} turns out to be approximately an order of magnitude smaller than M_{11}.
- Thus the 11th dimension appears to be larger than the dimensions compactified on the Calabi–Yau manifold, and at an intermediate stage the world appears 5–dimensional with two 4–dimensional boundaries (walls).

We thus have the following picture of the evolution and unification of coupling constants. At low energies the world is 4–dimensional and the couplings evolve accordingly with energy: a logarithmic variation of gauge coupling constants and the usual power law behaviour for the gravitational coupling. Around R_{11}^{-1} we have an additional 5th dimension and the power law evolution of the gravitational interactions changes. Gauge couplings are not affected at that scale since the gauge fields live on the walls and do not feel the existence of the 5th dimension. Finally at M_{GUT} the theory becomes 11–dimensional and both gravitational and gauge couplings show a power law behaviour and meet at the scale M_{11}, the fundamental scale of the theory. It is obvious that the correct choice of R_{11} is needed to achieve unification. We also see that, although the theory is weakly coupled at M_{GUT}, this is no longer true at M_{11}. The naive estimate for the evolution of the gauge coupling constants between M_{GUT} and M_{11} goes with the sixth power of the scale. At M_{11} we thus expect unification of the couplings at $\alpha \sim O(1)$. In that sense, the M–theoretic description of the heterotic string gives an interpolation between weak coupling and moderate coupling. In $d = 4$ this is not strong–weak coupling duality in the usual sense. We shall later come back to these questions when we discuss the appearance of a critical

limit on the size of R_{11}. A value of $\alpha \sim O(1)$ (and thus $S \sim O(1)$) at M_{11} might also be favoured in view of the question of the dynamical determination of the *vev* of the dilaton field [73].

VII.2 The effective action in $d = 4$

We now want to perform a compactification from $d = 11$ to $d = 4$. Again we use the method of reduction and truncation. For the metric we write

$$g^{(11)}_{MN} = \begin{pmatrix} c_4 e^{-\gamma} e^{-2\sigma} g_{\mu\nu} & & \\ & e^{\sigma} g_{mn} & \\ & & e^{2\gamma} e^{-2\sigma} \end{pmatrix} \tag{VII.52}$$

with $M, N = 1 \ldots 11$; $\mu, \nu = 1 \ldots 4$; $m, n = 5 \ldots 10$ and $\det(g_{mn})$=1. This is the frame in which the 11–dimensional Einstein action gives the ordinary Einstein action after the reduction to $d = 4$:

$$-\frac{1}{2\kappa^2} \int d^{11}x \sqrt{g^{(11)}} R^{(11)} = -\frac{c_4 \hat{V}_7}{2\kappa^2} \int d^4x \sqrt{g} R + \ldots \tag{VII.53}$$

where $\hat{V}_7 = \int d^7x$ is the coordinate volume of the compact 7–manifold and the scaling factor c_4 describes our freedom to choose the units in $d = 4$. The most popular choice in the literature is $c_4 = 1$. This, however, corresponds to the unphysical situation in which the 4–dimensional Planck mass is determined by the choice of $\hat{V}_7$ which is just a convention. With $c_4 = 1$ one needs further rescaling of the 4–dimensional metric. We instead prefer the choice

$$c_4 = V_7 / \hat{V}_7 \tag{VII.54}$$

where $V_7 = \int d^7x \sqrt{g^{(7)}}$ is the physical volume of the compact 7–manifold. In this way we recover eq. (VII.48) in which the 4–dimensional Planck mass depends on the physical (and not coordinate) volume of the manifold on which we compactify. As a result, if we start from the product of the 4–dimensional Minkowski space and some 7–dimensional compact space (in the leading order of the expansion in $\kappa^{2/3}$) as a ground state in $d = 11$ we obtain the Minkowski space with the standard normalization as the vacuum in $d = 4$.

To find a more explicit formula for c_4 we have to discuss the fields σ and γ in some detail. In the leading approximation, σ is the overall modulus of the Calabi–Yau 6–manifold. We can divide it into a sum of the vacuum expectation value, $\langle\sigma\rangle$, and the fluctuation $\tilde{\sigma}$. In general both parts could depend on all 11 coordinates, but in practice we have to impose some restrictions. The vacuum expectation value cannot depend

on x^μ if the 4–dimensional theory is to be Lorentz–invariant. In the fluctuations we drop the dependence on the compact coordinates corresponding to the higher Kaluza–Klein modes. Furthermore, we know that in the leading approximation $\langle\sigma\rangle$ is just a constant, σ_0 , while corrections depending on the internal coordinates, σ_1, are of the next order in $\kappa^{2/3}$. Thus, we obtain

$$\sigma(x^\mu, x^m, x^{11}) = \langle\sigma\rangle\,(x^m, x^{11}) + \tilde{\sigma}(x^\mu) = \sigma_0 + \sigma_1(x^m, x^{11}) + \tilde{\sigma}(x^\mu)\,. \tag{VII.55}$$

To make the above decomposition unique we define σ_0 by requiring that the integral of σ_1 over the internal space vanishes. The analogous decomposition can be also done for γ. With the above definitions, the physical volume of the compact space is

$$V_7 = \int d^7x\,\left\langle e^{2\sigma}e^{\gamma}\right\rangle = e^{2\sigma_0}e^{\gamma_0}\hat{V}_7 \tag{VII.56}$$

up to corrections of order $\kappa^{4/3}$. Thus, the parameter c_4 can be written as

$$c_4 = e^{2\sigma_0}e^{\gamma_0}\,. \tag{VII.57}$$

The choice of coordinate volumes is just a convention. For example in the case of the Calabi–Yau 6–manifold only the product $e^{3\sigma}\hat{V}_6$ has physical meaning. For definiteness we will use the convention that the coordinate volumes are equal 1 in M_{11} units. Thus, $\left\langle e^{3\sigma}\right\rangle$ describes the Calabi–Yau volume in these units. Using eqs. (VII.50,VII.51) we obtain $e^{3\sigma_0} = VM_{11}^6 \approx (2.3)^6$, $e^{\gamma_0}e^{-\sigma_0} = R_{11}M_{11}$. The parameter c_4 is equal to the square of the 4–dimensional Planck mass in these units.

At the classical level we compactify on $M^4 \times X^6 \times S^1/Z_2$. This means that the vacuum expectation values $\langle\sigma\rangle$ and $\langle\gamma\rangle$ are just constants and eq. (VII.55) reduces to

$$\sigma = \sigma_0 + \tilde{\sigma}(x^\mu), \qquad\qquad \gamma = \gamma_0 + \tilde{\gamma}(x^\mu)\,. \tag{VII.58}$$

In such a situation σ and γ are 4–dimensional fields. We introduce two other 4–dimensional fields D and C_{11} by the relations

$$\frac{1}{4!c_4}e^{6\sigma}G_{11\lambda\mu\nu} = \epsilon_{\lambda\mu\nu\rho}\,(\partial^\rho D)\,, \tag{VII.59}$$

$$C_{11a\bar{b}} = C_{11}\delta_{a\bar{b}} \tag{VII.60}$$

where x^a ($x^{\bar{b}}$) is the holomorphic (antiholomorphic) coordinate of the Calabi–Yau manifold. Now we can define the dilaton and the modulus fields by

$$\mathcal{S} = \frac{1}{(4\pi)^{2/3}}\left(e^{3\sigma} + i24\sqrt{2}D\right)\,, \tag{VII.61}$$

$$\mathcal{T} = \frac{1}{(4\pi)^{2/3}}\left(e^{\gamma} + i6\sqrt{2}C_{11} + C_i^* C_i\right) \qquad \text{(VII.62)}$$

where the observable sector matter fields C_i originate from the gauge fields A_M on the 10–dimensional observable wall (and M is an index in the compactified six dimensions). The Kähler potential takes its standard form

$$K = -\log(\mathcal{S}+\mathcal{S}^*) - 3\log(\mathcal{T}+\mathcal{T}^* - 2C_i^* C_i)\,. \qquad \text{(VII.63)}$$

The imaginary part of $\mathcal{S}$ (Im$\mathcal{S}$) corresponds to the model independent axion, and with the above normalization the gauge kinetic function is $f = \mathcal{S}$. We have also

$$W(C) = d_{ijk}C_iC_jC_k \qquad \text{(VII.64)}$$

Thus the action to leading order is very similar to the weakly coupled case.

Before drawing any conclusion from the formulae obtained above we have to discuss a possible obstruction at the next to leading order. For the 3–index tensor field H in $d = 10$ supergravity to be well defined one has to satisfy $dH = \mathrm{tr}F_1^2 + \mathrm{tr}F_2^2 - \mathrm{tr}R^2 = 0$ cohomologically. Here F_i denotes the E_8 field strength at the corresponding wall. In the simplest case of the standard embedding one assumes $\mathrm{tr}F_1^2 = \mathrm{tr}R^2$ locally and the gauge group is broken to $E_6 \times E_8$. Since in the M–theory case the two different gauge groups live on the two different boundaries (walls) of space–time such a cancellation point by point is no longer possible [117]. We expect nontrivial vacuum expectation values ($vev's$) of

$$(dG) \propto \sum_i \delta(x^{11} - x_i^{11})\left(\mathrm{tr}F_i^2 - \frac{1}{2}\mathrm{tr}R^2\right) \qquad \text{(VII.65)}$$

at least on one boundary (x_i^{11} is the position of the i–th boundary). In the case of the standard embedding we would have $\mathrm{tr}F_1^2 - \frac{1}{2}\mathrm{tr}R^2 = \frac{1}{2}\mathrm{tr}R^2$ on one and $\mathrm{tr}F_2^2 - \frac{1}{2}\mathrm{tr}R^2 = -\frac{1}{2}\mathrm{tr}R^2$ on the other boundary. This might pose a severe problem since a nontrivial vev of G might be in conflict with supersymmetry ($G_{11ABC} = H_{ABC}$). The supersymmetry transformation law in $d = 11$ reads

$$\delta\psi_M = D_M\eta + \frac{\sqrt{2}}{288}G_{IJKL}\left(\Gamma_M^{IJKL} - 8\delta_M^I\Gamma^{JKL}\right)\eta + \ldots \qquad \text{(VII.66)}$$

Supersymmetry will be broken unless e.g. the derivative term $D_M\eta$ compensates the nontrivial vev of E. Witten has shown [117] that such a cancellation can occur and constructed the solution in the linearized

approximation (linear in the expansion parameter $\kappa^{2/3}$). This solution requires some modification of the metric on M^{11}:

$$g_{MN}^{(11)} = \begin{pmatrix} (1+b)\eta_{\mu\nu} & & \\ & (g_{ij}+h_{ij}) & \\ & & (1+\gamma') \end{pmatrix} . \qquad \text{(VII.67)}$$

M^{11} is no longer a direct product $M^4 \times X^6 \times S^1/Z_2$ because b, h_{ij} and γ' depend now on the compactified coordinates. The volume of X^6 depends on x^{11} [117]:

$$\frac{\partial}{\partial x^{11}} V = -\frac{\sqrt{2}}{8} \int d^6x \sqrt{g} \omega^{AB} \omega^{CD} G_{ABCD} \qquad \text{(VII.68)}$$

where the integral is over the Calabi–Yau manifold X^6 and ω is the corresponding Kähler form. The parameter $(1+b)$ is the scale factor of the Minkowski 4–manifold and depends on x^{11} in the following way

$$\frac{\partial}{\partial x^{11}} b = \frac{1}{2} \frac{\partial}{\partial x^{11}} \log v_4 = \frac{\sqrt{2}}{24} \omega^{AB} \omega^{CD} G_{ABCD} \qquad \text{(VII.69)}$$

where v_4 is the physical volume for some fixed coordinate volume in M^4. In our simple reduction and truncation method with the metric $g_{MN}^{(11)}$ given by eq. (VII.52) we can reproduce the x^{11} dependence of V and v_4. The volume of X^6 is determined by σ:

$$\frac{\partial}{\partial x^{11}} \log V = \frac{\partial}{\partial x^{11}} \left(3 \langle \sigma \rangle\right) = 3 \frac{\partial}{\partial x^{11}} \sigma \qquad \text{(VII.70)}$$

while the scale factor of M^4 can be similarly expressed in terms of the fields σ and γ:

$$\frac{\partial}{\partial x^{11}} \log v_4 = -\frac{\partial}{\partial x^{11}} \left(2 \langle \gamma \rangle + 4 \langle \sigma \rangle\right) = -\frac{\partial}{\partial x^{11}} (2\gamma + 4\sigma) . \qquad \text{(VII.71)}$$

Substituting $\langle \sigma \rangle$ by σ in the above two equations is allowed because, due to our decomposition (VII.55), only the *vev* of σ depends on the internal coordinates (the same is true for γ). The scale factor b calculated in ref. [117] also depends on the Calabi–Yau coordinates. Such a dependence cannot be reproduced in our simple reduction and truncation compactification so we have to average eq. (VII.69) over X^6. Using equations (VII.68–VII.71) after such an averaging we obtain (to leading order in the expansion parameter $\kappa^{2/3}$) [102]

$$\frac{\partial \gamma}{\partial x^{11}} = -\frac{\partial \sigma}{\partial x^{11}} = \frac{\sqrt{2}}{24} \frac{\int d^6x \sqrt{g} \omega^{AB} \omega^{CD} G_{ABCD}}{\int d^6x \sqrt{g}} . \qquad \text{(VII.72)}$$

Substituting the vacuum expectation value of G found in [117] we can rewrite it in the form

$$\frac{\partial\gamma}{\partial x^{11}} = -\frac{\partial\sigma}{\partial x^{11}} = \frac{2}{3}\alpha\kappa^{2/3}V^{-2/3} \tag{VII.73}$$

where

$$\alpha = \frac{\pi c}{2(4\pi)^{2/3}} \tag{VII.74}$$

and c is a constant of order unity given for the standard embedding of the spin connection by

$$c = V^{-1/3}\left|\int \frac{\omega\wedge\mathrm{tr}(R\wedge R)}{8\pi^2}\right| . \tag{VII.75}$$

Our calculations, as those of Witten, are valid only in the leading non-trivial order in the $\kappa^{2/3}$ expansion. The expression (VII.73) for the derivatives of σ and γ contains an explicit factor $\kappa^{2/3}$. This means that we should take the lowest order value for the Calabi–Yau volume in that expression. An analogous procedure has been used in obtaining all formulae presented in this paper. We always expand in $\kappa^{2/3}$ and drop all terms which are of higher order. Taking the above into account and using our units in which $M_{11} = 1$ we can rewrite eq. (VII.73) in the simple form:

$$\frac{\partial\gamma}{\partial x^{11}} = -\frac{\partial\sigma}{\partial x^{11}} = \frac{2}{3}\alpha e^{-2\sigma_0} . \tag{VII.76}$$

Eqs. (VII.72–VII.76) as derived in ref. [102] contain all the information to deduce the effective action, i.e. Kähler potential, superpotential and gauge kinetic function of the 4–dimensional effective supergravity theory.

It is the above dependence of σ and γ on x^{11} that leads to these consequences. One has to be careful in defining the fields in $d = 4$. It is obvious, that the 4–dimensional fields $\mathcal{S}$ and $\mathcal{T}$ can no longer be defined by eqs. (VII.61, VII.62) because now σ and γ are 5–dimensional fields. We have to integrate out the dependence on the 11th coordinate. In the present approximation, this procedure is quite simple: we have to replace σ and γ in the definitions of $\mathcal{S}$ and $\mathcal{T}$ with their averages over the S^1/Z_2 interval [102]. With the linear dependence of σ and γ on x^{11} their average values coincide with the values taken in the middle of the S^1/Z_2 interval

$$\bar\sigma = \sigma\left(\frac{\pi\rho}{2}\right) = \sigma_0 + \tilde\sigma(x^\mu) , \tag{VII.77}$$

$$\bar\gamma = \gamma\left(\frac{\pi\rho}{2}\right) = \gamma_0 + \tilde\gamma(x^\mu) . \tag{VII.78}$$

When we reduce the boundary part of the Lagrangian of M–theory to 4 dimensions we find exponents of the fields σ and γ evaluated at the boundaries. Using eqs. (VII.55) and (VII.76) we get

$$e^{-\gamma}|_{M_i^{10}} = e^{-\gamma_0} \pm \frac{1}{3}\alpha e^{-3\sigma_0}\,, \tag{VII.79}$$

$$e^{3\sigma}|_{M_i^{10}} = e^{3\sigma_0} \pm \alpha e^{\gamma_0}\,. \tag{VII.80}$$

The above formulae have very important consequences for the definitions of the Kähler potential and the gauge kinetic functions. For example, the coefficient in front of the $D_\mu C_i^* D^\mu C_i$ kinetic term is proportional to $e^{-\gamma}$ evaluated at the E_6 wall where the matter fields propagate. At the lowest order this was just $e^{-\gamma_0}$ or $\langle \mathcal{T} \rangle^{-1}$ up to some numerical factor. From eq. (VII.79) we see that at the next to leading order $\langle \mathcal{S} \rangle^{-1}$ is also involved with relative coefficient $\alpha/3$. Taking such corrections into account we find that at this order the Kähler potential is given by

$$K = -\log(\mathcal{S}+\mathcal{S}^*) + \frac{2\alpha C_i^* C_i}{\mathcal{S}+\mathcal{S}^*} - 3\log(\mathcal{T}+\mathcal{T}^* - 2C_i^* C_i) \tag{VII.81}$$

with $\mathcal{S}$ and $\mathcal{T}$ now defined by

$$\mathcal{S} = \frac{1}{(4\pi)^{2/3}}\left(e^{3\bar{\sigma}} + i24\sqrt{2}\bar{D} + \alpha C_i^* C_i\right)\,, \tag{VII.82}$$

$$\mathcal{T} = \frac{1}{(4\pi)^{2/3}}\left(e^{\bar{\gamma}} + i6\sqrt{2}\bar{C}_{11} + C_i^* C_i\right) \tag{VII.83}$$

where bars denote averaging over the 11th dimension. It might be of some interest to note that the combination $\mathcal{S}\mathcal{T}^3$ is independent of x^{11} even before this averaging procedure took place. The solution above is valid only for terms at most linear in α. Keeping this in mind we could write the Kähler potential also in the form

$$K = -\log(\mathcal{S}+\mathcal{S}^* - 2\alpha C_i^* C_i) - 3\log(\mathcal{T}+\mathcal{T}^* - 2C_i^* C_i). \tag{VII.84}$$

Equipped with this definition the calculation of the gauge kinetic function(s) from eqs. (VII.76, VII.80) becomes a trivial exercise [102]. In the five–dimensional theory f depends on the 11–dimensional coordinate as well, thus the gauge kinetic function takes different values at the two walls. The averaging procedure allows us to deduce these functions directly. For the simple case at hand (the so–called standard embedding) eq. (VII.80) gives [102]

$$f_6 = \mathcal{S} + \alpha\mathcal{T}\,; \qquad f_8 = \mathcal{S} - \alpha\mathcal{T}\,. \tag{VII.85}$$

It is a special property of the standard embedding that the coefficients are equal and opposite. The coefficients might vary for more general cases. This completes the discussion of the $d = 4$ effective action in next to leading order, noting that the superpotential does not receive corrections at this level.

The nontrivial dependence of σ and γ on x^{11} can also enter definitions and/or interactions of other 4–dimensional fields. Let us next consider the gravitino. After all we have to show that this field is massless to give the final proof that the given solution respects supersymmetry. Its 11–dimensional kinetic term

$$-\frac{1}{2}\sqrt{g}\bar{\psi}_I\Gamma^{IJK}D_J\psi_K \tag{VII.86}$$

remains diagonal after compactification to $d = 4$ if we define the 4–dimensional gravitino, $\psi_\mu^{(4)}$, and dilatino, $\psi_{11}^{(4)}$, by the relations

$$\psi_\mu = e^{-(\sigma-\sigma_0)/2}e^{-(\gamma-\gamma_0)/4}\left(\psi_\mu^{(4)} + \frac{1}{\sqrt{6}}\Gamma_\mu\psi_{11}^{(4)}\right), \tag{VII.87}$$

$$\psi_{11} = -\frac{2}{\sqrt{6}}e^{(\sigma-\sigma_0)/2}e^{(\gamma-\gamma_0)/4}\Gamma^{11}\psi_{11}^{(4)}. \tag{VII.88}$$

The $d = 11$ kinetic term (VII.86) gives after compactification also a mass term for the $d = 4$ gravitino of the form

$$\frac{3}{8}e^{\sigma_0}e^{-\gamma_0}\frac{\partial\gamma}{\partial x^{11}} = \frac{\sqrt{2}}{64}e^{\sigma_0}e^{-\gamma_0}\frac{\int d^6x\sqrt{g}\omega^{AB}\omega^{CD}G_{ABCD}}{\int d^6x\sqrt{g}} = \frac{1}{4}\alpha e^{-\sigma_0}e^{-\gamma_0}. \tag{VII.89}$$

The origin of such a term can be traced back to nonzero values of the spin connection components ω_μ^{a11} and ω_m^{a11} resulting from the x^{11} dependence of the metric. It is a constant mass term from the 4–dimensional point of view. This, however, does not mean that the gravitino mass is nonzero. There is another contribution from the 11–dimensional term

$$-\frac{\sqrt{2}}{384}\sqrt{g}\bar{\psi}_I\Gamma^{IJKLMN}\psi_N\left(G_{JKLM} + \hat{G}_{JKLM}\right). \tag{VII.90}$$

After redefining fields according to (VII.87,VII.88) and averaging the nontrivial vacuum expectation value of G over X^6 we get from eq. (VII.90) a mass term which exactly cancels the previous contribution (VII.89). The gravitino is massless – the result which we expect in a model with unbroken supersymmetry and vanishing cosmological constant. Thus, we find that our simple reduction and truncation method (including the correct x^{11} dependence in next to leading order) reproduces the main features of the model.

The factor $\langle \exp(3\sigma) \rangle$ represents the volume of the six–dimensional compact space in units of M_{11}^{-6}. The x^{11} dependence of σ then leads to the geometrical picture that the volume of this space varies with x^{11} and differs at the two boundaries:

$$V_{E_8} = V_{E_6} - 2\pi^2 \rho \left(\frac{\kappa}{4\pi}\right)^{2/3} \left| \int \omega \wedge \frac{\mathrm{tr}(F \wedge F) - \frac{1}{2}\mathrm{tr}(R \wedge R)}{8\pi^2} \right| \quad \text{(VII.91)}$$

where the integral is over X^6 at the E_6 boundary. In the given approximation, this variation is linear, and for growing ρ the volume on the E_8 side becomes smaller and smaller. At a critical value of ρ the volume will thus vanish and this will provide us with an upper limit on ρ:

$$\rho < \rho_{crit} = \frac{(4\pi)^{2/3}}{c\pi^2} M_{11}^3 V_{E_6}^{2/3} \quad \text{(VII.92)}$$

where c was defined in eq. (VII.75). The critical value is model dependent and we shall not discuss this in detail here.

Let us now compare the M-theory picture with that of the weakly coupled heterotic string. Inspection of (VI.45) and (VII.85) reveals a close connection between the two [7, 105]. The variation of the Calabi–Yau manifold volume as discussed above is the analogue of the one loop correction of the gauge kinetic function (VI.45) in the weakly coupled case and has the same origin, namely a Green–Schwarz anomaly cancellation counterterm. In fact, also in the strongly coupled case this leads to a correction for the gauge coupling constants at the E_6 and E_8 side. As we have seen, gauge couplings are no longer given by the (averaged) $\mathcal{S}$–field, but by that combination of the (averaged) $\mathcal{S}$ and $\mathcal{T}$ fields which corresponds to the $\mathcal{S}$–field before averaging at the given boundary leading to

$$f_{6,8} = \mathcal{S} \pm \alpha \mathcal{T} \quad \text{(VII.93)}$$

at the E_6 (E_8) side respectively. The critical value of R_{11} will correspond to infinitely strong coupling at the E_8 side: $\mathcal{S} - \alpha\mathcal{T} = 0$. Since we are here close to criticality a correct phenomenological fit of $\alpha_{\mathrm{GUT}} = 1/25$ should include this correction $\alpha_{\mathrm{GUT}}^{-1} = \mathcal{S} + \alpha\mathcal{T}$ where $\mathcal{S}$ and $\alpha\mathcal{T}$ give comparable contributions. This is a difference to the weakly coupled case, where in $f = S + \epsilon T$ the latter contribution was small compared to S. The stability of this result for the corrections to f when going from weak coupling to strong coupling is only possible by virtue of the rather special properties of f. f does not receive further perturbative corrections beyond one loop [109, 96], and the one loop corrections are determined by the anomaly considerations. The formal expressions for the corrections are identical, the only difference being that in the strongly coupled

case these corrections are to be interpreted as having an importance comparable to the classical value.

VIII. Supersymmetry breaking at the hidden wall

For the discussion of supersymmetry breakdown we should carefully examine the supersymmetry transformations of fermionic fields. Of particular importance are the fields that originate from the higher dimensional gravitino. For the $d = 11$ action, the supersymmetry transformation laws for these fields are given by

$$\begin{aligned}\delta\psi_A &= D_A\eta + \frac{\sqrt{2}}{288}G_{IJKL}\left(\Gamma_A^{IJKL} - 8\delta_A^I\Gamma^{JKL}\right)\eta - \\ &- \frac{1}{1152\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3}\delta(x^{11})\left(\bar{\chi}^a\Gamma_{BCD}\chi^a\right)\left(\Gamma_A^{BCD} - 6\delta_A^B\Gamma^{CD}\right)\eta\ldots\end{aligned} \tag{VIII.94}$$

as well as

$$\begin{aligned}\delta\psi_{11} &= D_{11}\eta + \frac{\sqrt{2}}{288}G_{IJKL}\left(\Gamma_{11}^{IJKL} - 8\delta_{11}^I\Gamma^{JKL}\right)\eta + \\ &+ \frac{1}{1152\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3}\delta(x^{11})\left(\bar{\chi}^a\Gamma_{ABC}\chi^a\right)\Gamma^{ABC}\eta + .\end{aligned} \tag{VIII.95}$$

where gaugino bilinears appear in the right hand side of both expressions. Again we consider gaugino condensation at the hidden E_8 boundary:

$$\langle\bar{\chi}^a\Gamma_{ijk}\chi^a\rangle = g_8^2\Lambda^3\epsilon_{ijk}. \tag{VIII.96}$$

The E_8 gauge coupling constant appears in this equation because the straightforward reduction and truncation leaves a non–canonical normalization for the gaugino kinetic term. An important property of the weakly coupled case ($d = 10$ Lagrangian) was the fact that the gaugino condensate and the three–index tensor field H contributed to the scalar potential in a full square. Hořava made the important observation that a similar structure appears in the M–theory Lagrangian as well [53]:

$$-\frac{1}{12\kappa^2}\int_{M^{11}} d^{11}x\sqrt{g}\left(G_{ABC11} - \frac{\sqrt{2}}{32\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3}\delta(x^{11})\bar{\chi}^a\Gamma_{ABC}\chi^a\right)^2 \tag{VIII.97}$$

with the obvious relation between H and G. Let us now have a closer look at the form of G. At the next to leading order we have

$$\begin{aligned}G_{11ABC} &= (\partial_{11}C_{ABC} + \text{permutations}) \\ &+ \frac{1}{4\pi\sqrt{2}}\left(\frac{\kappa}{4\pi}\right)^{2/3}\sum_i\delta(x^{11} - x_i^{11})(\omega_{ABC}^{YM} - \frac{1}{2}\omega_{ABC}^L).\end{aligned} \tag{VIII.98}$$

Observe, that in the bulk we have $G = dC$ with the Chern–Simons contributions confined to the boundaries. Formula (VIII.97) suggests a cancellation between the gaugino condensate and the G–field in a way that is very similar to the weakly coupled case, but the nature of the cancellation of terms becomes much more transparent now. Remember that in the former case we had argued that because of the quantization condition for $\langle dB \rangle$ the gaugino condensate is cancelled not by $\langle dB \rangle$ but by a flux of the Chern–Simons terms. *Here, this becomes obvious.* The condensate is located at the wall as are the Chern–Simons terms, so this cancellation has to happen locally at the wall and dC should vanish to ensure that G does not have a *vev* in the bulk. In any case there is a quantization condition for dC as well [118].

So this cancellation is very similar to the one in the weakly coupled case. At the minimum of the potential we obtain $G_{ABCD} = 0$ everywhere and

$$G_{ABC11} = \frac{\sqrt{2}}{32\pi} \left(\frac{\kappa}{4\pi}\right)^{2/3} \delta(x^{11}) \bar{\chi}^a \Gamma_{ABC} \chi^a \tag{VIII.99}$$

at the hidden wall. Eqs. (VIII.94) and (VIII.95) then become

$$\delta\psi_A = D_A \eta + \dots \tag{VIII.100}$$

$$\delta\psi_{11} = D_{11}\eta + \frac{1}{384\pi} \left(\frac{\kappa}{4\pi}\right)^{2/3} \delta(x^{11}) \left(\bar{\chi}^a \Gamma_{ABC} \chi^a\right) \Gamma^{ABC} \eta + \dots \tag{VIII.101}$$

An inspection of the potential shows that $\delta\psi_{11}$ is nonvanishing and that supersymmetry is spontaneously broken. Because of the cancellation in eq. (VIII.97), the cosmological constant vanishes to leading order. Recalling the supersymmetry transformation law for the elfbein,

$$\delta e_I^m = \frac{1}{2} \bar{\eta} \Gamma^m \psi_I, \tag{VIII.102}$$

one finds that the superpartner of the field $\mathcal{T}$ plays the role of the goldstino. Again we have a situation where $F_{\mathcal{S}} = 0$ (due to the cancellation in (VIII.97)) with nonvanishing $F_{\mathcal{T}}$. But here we find the novel and interesting situation that $F_{\mathcal{T}}$ differs from zero only at the hidden wall, although the field itself is a bulk field.

At that wall our discussion is completely 4–dimensional although we are still dealing effectively with a $d = 5$ theory. To reach the effective theory in $d = 4$ we have to integrate out the dependence of the x^{11} coordinate. As in the previous section this can be performed by the averaging procedure explained there. With the gaugino condensation scale Λ sufficiently small compared to the compactification scale M_{GUT}, the low–energy effective theory is well described by four dimensional $N = 1$ supergravity in which supersymmetry is spontaneously broken. In this

case, the modes which remain at low energies will be well approximated by constant modes along the x^{11} direction. This observation justifies our averaging procedure to obtain four dimensional quantities. Averaging $\delta\psi_{11}$ over x^{11}, we thus obtain the *vev* of the auxiliary field $F_{\mathcal{T}}$

$$F_{\mathcal{T}} = \frac{1}{2}\mathcal{T}\frac{\int dx^{11}\sqrt{g_{11\,11}}\delta\psi_{11}}{\int dx^{11}\sqrt{g_{11\,11}}}. \qquad \text{(VIII.103)}$$

Note that this procedure allows for a nonlocal cancellation of the *vev* of the auxiliary field in $d = 4$. A condensate with equal size and opposite sign at the observable wall could cancel the effect and restore supersymmetry. Using $\int dx^{11}\sqrt{g_{11\,11}}\delta(x^{11}) = 1$, the auxiliary field is found to be

$$F_{\mathcal{T}} = \mathcal{T}\frac{1}{32\pi(4\pi)^{2/3}}\frac{g_8^2\Lambda^3}{R_{11}M_{11}^3} \qquad \text{(VIII.104)}$$

Similarly one can easily show that $F_{\mathcal{S}}$ as well as the vacuum energy vanish. This allows us then to unambiguously determine the gravitino mass, which is related to the auxiliary field in the following way:

$$m_{3/2} = \frac{F_{\mathcal{T}}}{\mathcal{T}+\mathcal{T}^*} = \frac{1}{64\pi(4\pi)^{2/3}}\frac{g_8^2\Lambda^3}{R_{11}M_{11}^3} = \frac{\pi}{2}\frac{\Lambda^3}{M_{Planck}^2}. \qquad \text{(VIII.105)}$$

As a nontrivial check one may calculate the gravitino mass in a different way. A term in the Lagrangian

$$-\frac{\sqrt{2}}{192\kappa^2}\int dx^{11}\sqrt{g}\bar{\psi}_I\Gamma^{IJKLMN}\psi_N G_{JKLM}, \qquad \text{(VIII.106)}$$

becomes the gravitino mass term when compactified to four dimensions. Using the $vev's$ of the G_{IJK11} given by eq. (VIII.99), one can obtain the same result as eq. (VIII.105). This is a consistency check for our approach and the fact that the vacuum energy vanishes in the given approximation.

It follows from eq. (VIII.105), that the gravitino mass tends to zero when the radius of the eleventh dimension goes to infinity. When the four–dimensional Planck scale is fixed to be the measured value, however, the gravitino mass in the strongly coupled case is expressed in a standard manner, similar to the weakly coupled case. To obtain the gravitino mass of the order of 1 TeV, one has to adjust Λ to be of the order of 10^{13} GeV when one constructs a realistic model by appropriately breaking the E_8 gauge group at the hidden wall.

In the minimization of the potential, we have implicitly used the leading order approximation. As was explained in a previous section, the next to leading order correction gives the non–trivial dependence of the

background metric on x^{11}. Then the Einstein–Hilbert action in eleven dimensions gives an additional contribution to the scalar potential in the four–dimensional effective theory, which shifts the $vev's$ of the G_{IJKL}. As a consequence, F_S will no longer vanish. Though this may be significant when we discuss soft masses, it does not drastically change our estimate of the gravitino mass (VIII.105) and our main conclusion drawn here is still valid after the higher order corrections are taken into account. In any case, these questions have to be addressed if one aims at realistic models for particle physics.

IX. Summary and outlook

In these lectures we have discussed the mechanism of gaugino condensation and flux stabilization within the heterotic scenario in its simplest version. The picture could be easily generalized to type II orientifolds or other types of string constructions, leading to the notion of supersymmetry breakdown on a hidden brane.

There is, however, still a long way to go towards realistic model building. One of the obstacles is the question of moduli stabilization in string theory. Without a solution of this problem we shall usually obtain so-called runaway vacua where e.g. coupling constants run to unrealistic values. Another obstacle is the appearance of instabilities of the scalar potential once we include radiative corrections. We have discussed some aspects of this in section 6 when we included radiative (threshold) corrections to the gauge coupling constants (45). If we would consider $f = S \pm \epsilon T$ and reinsert this into the $F-$terms in (32) we would obtain new contributions to the scalar potential leading to minima with a negative vacuum energy. This, of course, is nothing else than the problem of the cosmological constant.

Recently there has been some revived interest in this discussion. One aspect concerns the consideration of compactification of the extra dimensions on non-Kähler manifolds. Moduli are stabilized with the help of fluxes of various antisymmetric tensor fields. For more details and references, see [9, 15, 22, 44, 64]. This allows the stabilization of many moduli already in the supersymmetric framework in a rather general context and avoids cosmological moduli problems.

Within the heterotic M-theory context, there have been attempts to go beyond the classical level [23]. In ref. [13] one finds a rather comprehensive discussion of moduli stabilization and supersymmetry breakdown in a general set-up of heterotic M-theory including 5-branes in the bulk. Moduli can be stabilized, but a large negative vacuum energy remains. Similar results in the heterotic string theory have been reported

in [52]. These results are, of course, also of interest in the discussion of cosmological aspects of string theory [63]. So the mechanism of supersymmetry breakdown through gaugino condensation still remains one the most promising subjects in the discussion of realistic string model building.

Acknowledgements

We would like to thank the organizers, especially Gérard Smadja, for their hospitality. We would also like to thank François Gieres for a careful reading of the manuscript. This work was partially supported by the European Community's Human Potential Programme under contracts HPRN-CT-2000-00131: Quantum Space-time, HPRN-CT-2000-00148: Physics Across the Present Energy Frontier and HPRN-CT-2000-00152: Supersymmetry and the Early Universe.

References

[1] Aldazabal, G., Font, A., Ibáñez, L. E. and Uranga, A.M., *String GUTs*, *Nucl. Phys.* **B 452** (1995) 3

[2] Amaldi, U., de Boer, W., Fürstenau, H., *Phys. Lett.* **B 281** (1992) 374; Antoniadis, I., Ellis, J., Kelley, S., and Nanopoulos, D. V., *Phys. Lett.* **B 272** (1991) 31

[3] Antoniadis, I., Ellis, J., Lacaze, R. and Nanopoulos, D. V., *Phys. Lett.* **B 268** (1991) 188;
Dolan, L. and Liu, J. T., *Nucl. Phys.* **B 387** (1992) 86

[4] Antoniadis, I., Gava, E., Narain K. S. and Taylor, T. R., *Nucl. Phys.* **B 432** (1994) 187

[5] Antoniadis, I. , Narain, K. S. and Taylor, T. R., *Phys. Lett.* **B 267** (1991) 37;
Antoniadis, I., Gava, E. and Narain, K. S., *Nucl. Phys.* **B 383** (1992) 93; *Phys. Lett.* **B 283** (1992) 209;
Mayr, P., and Stieberger, S., *Nucl. Phys.* **B 412** (1994) 502

[6] Antoniadis, I. and Quirós, M., Phys. Lett. **B392** (1997) 61; hep–th/9705037

[7] Banks, T. and Dine, M., Nucl. Phys. **B479** (1996) 173

[8] Banks, T. and Dine, M., Nucl. Phys. **B505** (1997) 445

[9] K. Becker, M. Becker, K. Dasgupta and P. S. Green, JHEP **0304**, 007 (2003) [arXiv:hep-th/0301161].

[10] Bershadsky, M., Cecotti, Ooguri, S. H. and Vafa, C., *Nucl. Phys.* **B 405** (1993) 279; *Comm. Math. Phys.* **165** (1994) 311;
Hosono, S., Klemm, A., Theisen, S. and Yau, S. T., *Nucl. Phys.* **B 433** (1995) 501

[11] Binetruy, P. and Gaillard, M.K., Phys. Lett. **232B** (1989) 83

[12] Brignole, A., Ibáñez, L. E. and Muñoz, C., Nucl. Phys. **B422** (1994) 125 and references therein

[13] E. I. Buchbinder and B. A. Ovrut, arXiv:hep-th/0310112.

[14] Candelas, P., Horowitz, G., Strominger A. and Witten, E., Nucl. Phys. **B258** (1985) 46

[15] G. L. Cardoso, G. Curio, G. Dall'Agata and D. Lust, arXiv:hep-th/0310021.

[16] Chamseddine, A. H., Phys. Rev. **D24** (1981) 3065;
Chapline, G. F. and Manton, N. S., Phys. Lett. **B120** (1983) 105

[17] Choi, K., Phys. Rev. **D56** (1997) 6588

[18] Choi, K. and Kim. J. E., Phys. Lett. **B165** (1985) 71

[19] Choi, K., Kim, H. B. and Muñoz, C., Phys.Rev. **D57** (1998) 7521, hep–th/9711158

[20] Conrad, J. O., *Phys. Lett.* **B B421** (1998) 119, hep–th/9708031

[21] Cremmer, E., Ferrara, S., Girardello, L. and van Proeyen, A., *Nucl. Phys.* **B 212** (1983) 413

[22] G. Curio and A. Krause, Nucl. Phys. B **643**, 131 (2002) [arXiv:hep-th/0108220].

[23] G. Curio and A. Krause, arXiv:hep-th/0308202.

[24] Cvetič, M., Font, A., Ibáñez, L. E., Lüst, D. and Quevedo, F., *Nucl. Phys.* **B 361** (1991) 194

[25] For an extended list of references see: Carlos, B. de, Casas, J. A., Muñoz, C., Nucl.Phys. **B399** (1993) 623

[26] Font, A., Ibáñez, L. E., Lüst, D., Quevedo, F., Phys.Lett. **249B** (1990) 35

[27] Derendinger, J. P., Ferrara, S., Kounnas, C. and Zwirner, F., *Nucl. Phys.* **B 372** (1992) 145;
Antoniadis, I., Gava, E., Narain, K. S. and Taylor, T. R., *Nucl. Phys.* **B 407** (1993) 706

[28] Derendinger, J. P., Ibáñez, L. E. and Nilles, H. P., Phys. Lett. **B155** (1985) 65

[29] Derendinger, J. P., Ibáñez, L. E. and Nilles, H. P., Nucl. Phys. **B267** (1986) 365

[30] Dienes, K. R. and Faraggi, A. E., *Making ends meet: string unification and low–energy data*, Princeton IASSNS–HEP–95/24 (hep–th/9505018); *Gauge coupling unification in realistic free–fermionic string models*, Princeton IASSNS–HEP–94/113 (hep–th/9505046)

[31] Dine, M., Fischler, W., Srednicki, M., Nucl.Phys. **B189** (1981) 575

[32] Dine, M., Rohm, R., Seiberg, N. and Witten, E., Phys. Lett. **B156** (1985) 55

[33] Dine, M., Seiberg, N., Wen, X. G. and Witten, E., Nucl. Phys. **B289** (1987) 319; Nucl. Phys. **B278** (1986) 769

[34] Dixon, L., Harvey, J., Vafa, C. and Witten, E., *Nucl. Phys.* **B 261** (1985) 678; **B 274** (1986) 285;
Ibáñez, L. E., Mas, J., Nilles, H. P. and Quevedo, F., *Nucl. Phys.* **B 301** (1988) 157

[35] Dixon, L., Kaplunovsky, V. and Louis, J. *Nucl. Phys.* **B 355** (1991) 649

[36] Dudas, E., *Phys. Lett.* **B B416** (1998) 309, hep–th/9709043

[37] Dudas, E. and Grojean, C., *Nucl. Phys.* **B B507** (1997) 553, hep–th/9704177

[38] Ellis, J., Kim. J. E. and Nanopoulos, D. V., Phys. Lett. **B145** (1984) 181

[39] Ellis, J., Kelley, S. and Nanopoulos, D. V., *Phys. Lett.* **B 249** (1990) 441;
Amaldi, U., Boer, W. de and Fürstenau, H., *Phys. Lett.* **B 260** (1991) 447;
Langacker, P. and Luo, M. X., *Phys. Rev.* **D 44** (1991) 817

[40] Ferrara, S., Girardello, L. and Nilles, H. P., Phys. Lett. **B125** (1983) 457

[41] Ferrara, S., Kounnas, C., Lüst, D. and Zwirner, F., *Nucl. Phys.* **B 365** (1991) 431

[42] Font, A., Ibáñez, L. E., D. Lüst, Quevedo, F., Phys.Lett. **245B** (1990) 401;
Ferrara, S., Magnoli, N., Taylor, T. R., Veneziano, G., Phys.Lett. **245B** (1990) 409;
Nilles, H. P., M. Olechowski, Phys.Lett. **248B** (1990) 268;
Binetruy, P., Gaillard, M. K., Phys.Lett. **253B** (1991) 119;
Cvetič, M., Font, A., Ibáñez, L. E., Lüst, D., Quevedo, F., Nucl.Phys. **B361** (1991) 194

[43] Georgi, H., Kim, Jihn E., Nilles, H. P., *Phys. Lett.* **B B437** (1998) 325, hep-ph/9805510

[44] S. B. Giddings, S. Kachru and J. Polchinski, Phys. Rev. D **66**, 106006 (2002) [arXiv:hep-th/0105097].

[45] Ginsparg, P. *Phys. Lett.* **B 197** (1987) 139

[46] M. B. Green, J. H. Schwarz and E. Witten, Superstring theory, Cambridge Monographs on Mathematical Physics, Cambridge University Press 1987

[47] D. J. Gross, J. A. Harvey, E. J. Martinec and R. Rohm, Nucl. Phys. B **256**, 253 (1985).

[48] Horne, J. H., Moore, G., Nucl.Phys. **B432** (1994) 109

[49] Ibáñez, L. E., Nilles, H. P., Phys.Lett. **169B** (1986) 354;
Dixon, L., Kaplunovsky, V., Louis, J., Nucl.Phys. **B355** (1991) 649

[50] Green, M. and Schwarz, J., Phys. Lett. **B149** (1984) 117

[51] Green, M., Schwarz, J. and Witten, E., Superstring Theory, Cambridge University Press, 1987

[52] S. Gukov, S. Kachru, X. Liu and L. McAllister, arXiv:hep-th/0310159.

[53] Hořava, P., Phys. Rev. **D54** (1996) 7561

[54] Hořava, P. and Witten, E., Nucl. Phys. **B460** (1996) 506; Nucl. Phys. **B475** (1996) 94.

[55] Ibáñez, L. E., *Phys. Lett.* **B 318** (1993) 73

[56] Ibáñez, L. E. and Lüst, D., *Nucl. Phys.* **B 382** (1992) 305

[57] Ibáñez, L. E., Lüst, D. and Ross, G. G., *Phys. Lett.* **B 272** (1991) 251

[58] Ibáñez, L. E. and Nilles, H. P., Phys. Lett. **B169** (1986) 354

[59] Ibáñez, L. E., Nilles, H. P. and Quevedo, F., *Phys. Lett.* **B 187** (1987) 25;
Ibáñez, L. E., Kim. J. E., Nilles, H. P. and Quevedo, F., *Phys. Lett.* **B 191** (1987) 283

[60] Ibáñez, L. E., Nilles, H. P. and Quevedo, F., *Phys. Lett.* **B 192** (1987) 332

[61] Jungman, G., Kamionkowski, M. and Griest, K., Phys. Rep. **267** (1996) 195

[62] Kaplunovsky, V. S., *Nucl. Phys.* **B 307** (1988) 145, Erratum: *Nucl. Phys.* **B 382** (1992) 436

[63] S. Kachru, R. Kallosh, A. Linde and S. P. Trivedi, Phys. Rev. D **68**, 046005 (2003) [arXiv:hep-th/0301240].

[64] S. Kachru, M. B. Schulz and S. Trivedi, JHEP **0310**, 007 (2003) [arXiv:hep-th/0201028].

[65] Kaplunovsky, V. S. and Louis, J., Phys. Lett. **B306** (1993) 269

[66] Kaplunovsky, V. S. and Louis, J., *Nucl. Phys.* **B 444** (1995) 191

[67] Kawamura, Y., Nilles, H. P., Olechowski, M. and Yamaguchi, M., **JHEP 9806:008** (1998), hep-ph/9805397,

[68] Kawasaki, M. and Moroi, T., Prog. Theor. Phys. **93** (1995) 879

[69] Kawasaki, M. and Yanagida, T., Prog. Theor. Phys. **97** (1997) 809

[70] Kiritsis, E. and Kounnas, C., *Nucl. Phys.* **B 41** [Proceedings Sup.] (1995) 331; *Nucl. Phys.* **B 442** (1995) 472; *Infrared–regulated string theory and loop corrections to coupling constants*, hep–th/9507051

[71] Kim, J. E. and Nilles, H. P., Phys. Lett. B **553** (2003) 1 [arXiv:hep-ph/0210402].

[72] Krasnikov, N. V., Phys.Lett. **193B** (1987) 37;
Casas, J. A., Lalak, Z., Muñoz, C., Ross, G. G., Nucl.Phys. **B347** (1990) 243

[73] Lalak, Z., Niemeyer, A., Nilles, H. P., Phys.Lett. **349B** (1995) 99

[74] Lalak, Z., Niemeyer, A., Nilles, H. P., hep-th/9503170, Nucl.Phys. **B453** (1995) 100

[75] Lalak, Z. and Thomas, S., Nucl. Phys. B **515** (1998) 55 hep–th/9707223

[76] For a review see: Langacker, P., *Grand Unification and the Standard Model*, hep–ph/9411247

[77] Langacker, P. and Polonsky, N., *Phys. Rev.* **D 47** (1993) 4028 and references therein

[78] Lauer, J., Mas, J. and Nilles, H. P., Phys. Lett. B **226** (1989) 251.

[79] Lauer, J., Mas, J. and Nilles, H. P., Nucl. Phys. **B351** (1991) 353

[80] Li, T., Lopez, J. L. and Nanopoulos, D. V., Mod. Phys. Lett. A **12** (1997) 2647, hep–ph/9702237

[81] Lopes Cardoso, G., Lüst, D. and Mohaupt, T., *Nucl. Phys.* **B 450** (1995) 115

[82] Lukas, A., Ovrut, B. A. and Waldram, D., Nucl. Phys. B **532** (1998) 43, hep–th/9710208

[83] Lukas, A., Ovrut, B. A. and Waldram, D., Phys. Rev. D **57** (1998) 7529, hep–th/9711197

[84] Macorra, A. de la, Ross, G.G., Nucl.Phys. **B404** (1993) 321

[85] Matalliatakis, D., Nilles, H. P., Theisen, S., hep-th/9710247; *Phys. Lett.* **B 421** (1998) 169

[86] Mayr, P., Nilles, H. P. and Stieberger, S., *Phys. Lett.* **B 317** (1993) 53

[87] Mayr, P. and Stieberger, S., *Nucl. Phys.* **B 407** (1993) 725;
Bailin, D. , Love, A., Sabra, W. A. and Thomas, S., *Mod. Phys. Lett.* **A9** (1994) 67; **A10** (1995) 337

[88] Mayr, P. and Stieberger, S., *Phys. Lett.* **B 355** (1995) 107

[89] Mayr, P. and Stieberger, S., TUM–HEP–212/95;
Stieberger, S., *One–loop corrections and gauge coupling unification in superstring theory*, Ph.D. thesis, TUM–HEP–220/95

[90] For a review see: Mayr, P. and Stieberger, S., Proceedings *28th International Symposium on Particle Theory*, p. 72–79, Wendisch–Rietz (1994) (hep–th/9412196, DESY 95–027

[91] Montonen, C., Olive, D., Phys.Lett. **72B** (1977) 117;
Seiberg, N., Witten, E., Nucl.Phys. **B426** (1994) 19

[92] Nilles, H. P., Is Supersymmetry Afraid Of Condensates?, Phys.Lett. **112B** (1982) 455

[93] Nilles, H. P., Dynamically Broken Supergravity And The Hierarchy Problem, Phys. Lett. **B115** (1982) 193

[94] Nilles, H. P., Supergravity Generates Hierarchies, Nucl.Phys. **B217** (1983) 366

[95] Nilles, H. P., Supersymmetry, Supergravity And Particle Physics, Physics Reports **110** (1984) 1

[96] Nilles, H. P., The Role Of Classical Symmetries In The Low-Energy Limit Of Superstring Theories, Phys. Lett. **B180** (1986) 240

[97] Nilles, H. P., Lectures at the Trieste Spring School on Supersymmetry, Supergravity and Superstrings 1986, Eds. B. de Wit et al., World Scientific (1986), page 37

[98] H. P. Nilles, Vacuum Degeneracy In Four-Dimensional String Theories, Nucl. Phys. Proc. Suppl. **5B**, 185 (1988).

[99] Nilles, H. P., Int. Journ. of Modern Physics **A5** (1990) 4199

[100] Nilles, H. P., TASI lectures 1990, Testing the Standard Model, Ed. M. Cvetic and P. Langacker, World Scientific 1991, page 633

[101] Nilles, H. P., TASI lectures 1993, The Building Blocks of Creation, Ed. S. Raby and T. Walker, World Scientific 1994, page 291

[102] Nilles, H. P., Olechowski. M. and Yamaguchi, M., hep–th/9707143, Phys. Lett. **B415** (1997) 24.

[103] Nilles, H. P., Olechowski. M. and Yamaguchi, M., Nucl. Phys. B **530** (1998) 43, hep–th/9801030

[104] Nilles, H. P. and Stieberger, S., *How to reach the correct* $\sin^2\theta_W$ *and* α_s *in string theory*, hep-th/9510009, Phys.Lett. **B367** (1996) 126

[105] Nilles, H. P. and Stieberger, S., hep–th/9702110, Nucl. Phys. **B499** (1997) 3

[106] Pagels, H. and Primack, J. R., Phys. Rev. Lett **48** (1982) 223

[107] J. Polchinski, String theory, Cambridge University Press 1998

[108] Rohm, R. and Witten, E., Ann. Physics **170** (1986) 454

[109] Shifman, M. and Vainshtein, A., Nucl. Phys. **B277** (1986) 456

[110] Taylor, T. R., Phys.Lett. **164B** (1985) 43

[111] G. 't Hooft, *'Naturalness, chiral symmetry and spontaneous chiral symmetry breaking'*, in 'Recent Developments in Gauge Theories', Cargèse 1979, G. 't Hooft et al, New York 1980, *Plenum Press*

[112] Veneziano, G., Yankielowicz, S., Phys.Lett. **113B** (1982) 231

[113] Weinberg, S., Phys. Rev. Lett. **48** (1982) 1303

[114] Weinberg, S., *Phys. Lett.* **B 91** (1980) 51

[115] Witten, E., Constraints On Supersymmetry Breaking, Nucl.Phys. **B202** (1982) 253

[116] Witten, E., Dimensional Reduction Of Superstring Models, Phys. Lett. **B155** (1985) 151

[117] Witten, E., Strong Coupling Expansion Of Calabi-Yau Compactification, Nucl. Phys. **B471** (1996) 135

[118] Witten, E., hep-th/9609122, J. Geom. Phys. **22** (1997) 1

[119] Witten, E., Nucl.Phys. **B188** (1981) 513

[120] Witten, E., Symmetry Breaking Patterns In Superstring Models, *Nucl. Phys.* **B 258** (1985) 75

[121] Witten, E., New Issues In Manifolds Of SU(3) Holonomy, *Nucl. Phys.* **B 268** (1986) 79

IS DARK MATTER SUPERSYMMETRIC?

Wim de Boer
University of Karlsruhe
Wim.de.Boer@cern.ch

Abstract From the precise relic density measurement by WMAP the WIMP annihilation cross section can be determined in a model independent way. If the WIMPS are postulated to be the neutralinos of Supersymmetry, then only a limited region of parameter space matches this annihilation cross section. It is shown that the resulting positrons, antiprotons and gamma rays from the neutralino annihilation (mainly into $b\bar{b}$ quark pairs) provide the correct shape and order of magnitude for the missing gamma and hard positron fluxes in the Galactic Models and are consistent with the antiproton fluxes.

Keywords: Cold Dark Matter, Indirect detection, Diffuse Gamma Rays, Galactic Positrons, Antiprotons, Neutralinos, Supersymmetry, Haloprofile

I. Introduction

Cold Dark Matter (CDM) makes up 23% of the energy of the universe, as deduced from the temperature anisotropies in the Cosmic Microwave Background (CMB) in combination with data on the Hubble expansion and the density fluctuations in the universe [1]-[4]. The nature of the CDM is unknown, but one of the most popular explanation for it is the neutralino, a stable neutral particle predicted by Supersymmetry [5, 6]. The neutralinos are spin 1/2 Majorana particles, which can annihilate into pairs of Standard Model (SM) particles. The stable decay and fragmentation products are neutrinos, photons, protons, antiprotons, electrons and positrons. These decay products are and have been produced everywhere in the universe, but the protons and electrons are drowned in the many matter particles in the universe. However, the antimatter particles, neutrinos and photons may be detectable above the background from nuclear interactions, especially because of the much harder spectra expected from neutralino annihilation.

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 435–460.

In this paper we consider the annihilation of Dark Matter particles as a source for gamma rays, antiprotons and positrons. Indirect detection of Dark Matter with these particles has been discussed much before; for some recent references see [10]-[23]. The results presented here differ from previous ones by performing a statistical analysis to gamma rays, antiprotons and gamma rays *simultaneously*. Also most of the previous analysis were done before the WMAP data became available. These data provide us with a model independent estimate of the annihilation cross section, which in turn very much limits the parameter space.

In the following we first describe the model independent determination of this annihilation cross section from the relic density and the corresponding constraints on the parameter space in the Minimal Supersymmetric Model using supergravity inspired symmetry breaking (mSUGRA). The predictions from Supersymmetry concerning neutralino annihilation are discussed, while in the following section the background spectra for gamma rays, antiprotons and positrons, as predicted by the Galprop model [24, 25] are discussed. In the last section the global fits are discussed. They are performed within the frame work of the DarkSusy [26] program, after modifying it to simulate the background and propagation model from Galprop in DarkSusy. The propagation is only important for antiprotons and positrons because of their energy losses. For the SUSY particle spectrum we use Suspect [27], for the Higgs masses Feynhiggsfast [28] and as a cross check for the cross sections and relic density we use Micromegas [29]. Finally, some expectations for direct detection and indirect searches for Dark Matter from solar neutrinos will be given.

II. Annihilation Cross section Constraints from WMAP

In the early universe all particles were produced abundantly and were in thermal equilibrium through annihilation and production processes. The time evolution of the number density of the particles is given by the Boltzmann equation, which can be written for neutralinos as:

$$\frac{dn_\chi}{dt} + 3Hn_\chi = - < \sigma v > (n_\chi^2 - n_\chi^{eq2}), \qquad \text{(II.1)}$$

where H is the Hubble expansion rate, n_χ is the actual number density, n_χ^{eq} is the thermal equilibrium number density (before freeze-out), $< \sigma v >$ is thermally averaged value of the total annihilation cross section times the relative velocity of the annihilating neutralinos. The Hubble term takes care of the decrease in number density because of the expansion, while the first term on the right hand side represents the decrease

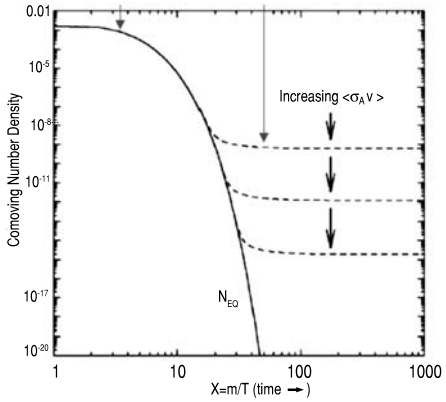

Figure 1. The comoving number density of any massive particle, which can stay in equilibrium by annihilation and creation processes with matter of the universe, until it falls out of thermal equilibrium,as indicated by the dashed lines. This happens roughly at a temperature of M/25, when the expansion rate becomes larger than the annihilation rate. Therefore the Hubble constant determines the thermally averaged annihilation cross section σv in a model independent way, if the actual density has been measured. From Ref. [6].

due to annihilation and the second term represents the increase through creation by the inverse reactions.

At temperatures below the mass of the neutralinos the number density drops exponentially. The annihilation rate $\Gamma =< \sigma v > n_\chi$ drops exponentially as well, and if it drops below the expansion rate, the neutralinos cease to annihilate. They fall out of equilibrium (freeze-out) at a temperature of about $m_\chi/25$ [30] and a relic cosmic abundance remains, as indicated in Fig. 1.

For the case that $< \sigma v >$ is energy independent, which is a good approximation in case there is no coannihilation and away from resonances, the present mass density in units of the critical density is given by [6]:

$$\Omega_\chi h^2 = \frac{m_\chi n_\chi}{\rho_c} \approx (\frac{3 \cdot 10^{-27} cm^3 s^{-1}}{< \sigma v >}). \qquad \text{(II.2)}$$

One observes that the present relic density is inversely proportional to the annihilation cross section at the time of freeze out, a result independent of the neutralino mass (except for logarithmic corrections). For the present value of $\Omega_\chi h^2 = 0.1$ the thermally averaged total cross section at the freeze-out temperature of $m_\chi/25$ must have been $3 \cdot 10^{-26} cm^3 s^{-1}$. This can be achieved only for restricted regions of parameter space in the MSSM, as will be discussed in the next section.

III. Dark Matter Predictions from Supersymmetry

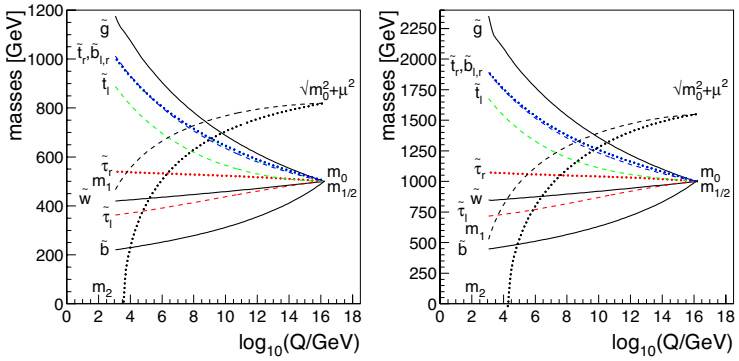

Figure 2. The running of the squark - and slepton masses starting at m_0, gaugino masses starting at $m_{1/2}$ and Higgs mass parameters starting at $\sqrt{m_0^2+\mu^2}$ for $m_0 = m_{1/2} = 500$ GeV and $\tan\beta$=51 (left) and for $m_0 = m_{1/2} = 1000$ GeV and $\tan\beta$=53 (right), which are the parameters of interest for the present analysis.

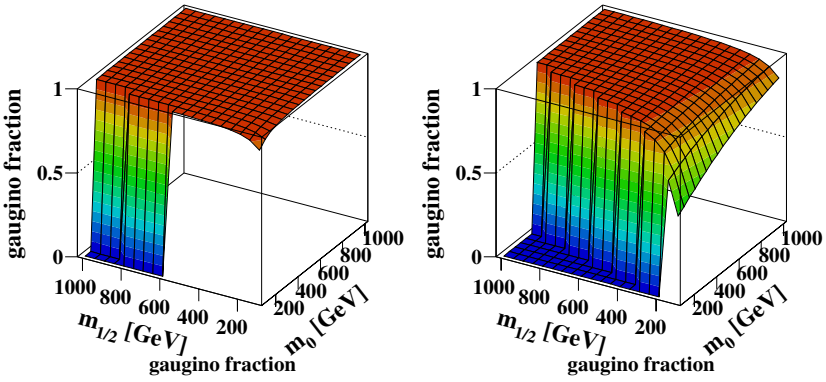

Figure 3. The gaugino fraction of the lightest neutralino as function of m_0 and $m_{1/2}$ for $\tan\beta$=10 (left) and $\tan\beta$=50 (right).

Supersymmetry [31] presupposes a symmetry between fermions and bosons, which can be realized in nature only if one assumes each particle with spin j has a supersymmetric partner with spin $|j-1/2|$ ($|j-1/2|$ for the Higgs bosons). This leads to a doubling of the particle spectrum. Unfortunately the supersymmetric particles or "sparticles" have not been observed so far, so the sparticle masses must be above the limits set by searches at present accelerators. Obviously SUSY cannot be an

exact symmetry of nature; or else the supersymmetric partners would have the same mass as the normal particles. The mSUGRA model, i.e. the Minimal Supersymmetric Standard Model (MSSM) with supergravity inspired breaking terms, is characterized by only 5 parameters: $m_0,\ m_{1/2},\ \tan\beta,\ \mathrm{sign}(\mu),\ A_0$. Here m_0 and $m_{1/2}$ are the common masses for the gauginos and scalars at the GUT scale, which is determined by the unification of the gauge couplings. Gauge unification is still possible with the precisely measured couplings at LEP [32]. The ratio of the vacuum expectation values of the neutral components of the two Higgs doublets in Supersymmetry is called $\tan\beta$ and A_0 is the trilinear coupling at the GUT scale. We only consider the dominant trilinear couplings of the third generation of quarks and leptons and assume also A_0 to be unified at the GUT scale. The constraints on the supersymmetric parameters space are practically independent of A_0 due to a coincidence from the constraints from the $b \to X_s\gamma$ rate and the lower limit on the Higgs mass of 114 GeV [32]. The absolute value of the Higgs mixing parameter μ is determined by electroweak symmetry breaking, while its sign is taken to be positive, as preferred by the anomalous magnetic moment of the muon [32].

The GUT scale masses are connected to low energy masses by the Renormalization Group Equations (RGE), as shown in Fig. 2. The running masses of the gauginos at low energy obey the simple solutions of the RGE:

$$M_i(t) = \frac{\tilde{\alpha}_i(t)}{\tilde{\alpha}_i(0)} m_{1/2}. \tag{III.3}$$

Numerically at the weak scale ($t = 2\ln(M_{GUT}/M_Z) = 66$) one finds (see fig. 2):

$$M_3(\tilde{g}) \approx 2.7 m_{1/2}, \tag{III.4}$$
$$M_2(M_Z) \approx 0.8 m_{1/2}, \tag{III.5}$$
$$M_1(M_Z) \approx 0.4 m_{1/2}. \tag{III.6}$$

The gluinos obtain corrections from the strong coupling constant α_3; therefore they grow heavier than the gauginos of the $SU(2)_L \otimes U(1)_Y$ group. Since the Higgsinos and gauginos are all spin 1/2 particles and are equal in all other quantum numbers, the mass eigenstates are in general mixtures of them, which are called generically charginos (neutralinos) for the mixture of the supersymmetric partners of the charged (neutral) gauge bosons and charged (neutral) Higgs bosons. The Majo-

rana neutralino and Dirac chargino fields can be written as:

$$\chi = \begin{pmatrix} \tilde{B} \\ \tilde{W}^3 \\ \tilde{H}_1^0 \\ \tilde{H}_2^0 \end{pmatrix}, \quad \psi = \begin{pmatrix} \tilde{W}^+ \\ \tilde{H}^+ \end{pmatrix},$$

while the mass matrices can be written as [31]:

$$M_A = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \quad Z_2 = \begin{pmatrix} -M_Z \cos\beta \sin\theta_W & M_Z \cos\beta \cos\theta_W \\ M_Z \sin\beta \sin\theta_W & -M_Z \sin\beta \cos\theta_W \end{pmatrix}$$

$$M_B = \begin{pmatrix} 0 & -\mu \\ -\mu & 0 \end{pmatrix} \quad Z_1 = \begin{pmatrix} -M_Z \cos\beta \sin\theta_W & M_Z \sin\beta \sin\theta_W \\ M_Z \cos\beta \cos\theta_W & -M_Z \sin\beta \cos\theta_W \end{pmatrix}$$

$$M^{(0)} = \begin{pmatrix} M_A & Z_1 \\ Z_2 & M_B \end{pmatrix} \tag{III.7}$$

$$M^{(c)} = \begin{pmatrix} M_2 & \sqrt{2} M_W \sin\beta \\ \sqrt{2} M_W \cos\beta & \mu \end{pmatrix} \tag{III.8}$$

The last matrix leads to two chargino eigenstates $\tilde{\chi}_{1,2}^{\pm}$. The dependence on the parameters at the GUT scale can be estimated by substituting for M_2 and μ their values at the weak scale: $M_1(M_Z) \approx 0.5 M_2(M_Z) \approx 0.4 m_{1/2}$ and $\mu(M_Z) \approx 0.63 \mu(0)$.

From Fig. 2 it can be seen that the mass parameters in the Higgs potential, m_1 and m_2, are driven negative, largely because of the large Yukawa couplings of the third generation of quarks and leptons. This leads to radiative electroweak symmetry breaking (EWSB), so the Higgs mechanism in Supersymmetry needs not to be introduced *ad hoc*, as in the Standard Model, but is caused by radiative corrections. The running is only strong enough for top masses between 140 and 200 GeV and if the starting value $\sqrt{\mu^2 + m_0^2}$ at the GUT scale is large enough, which in practice implies $\mu > M_2$. From the mass matrices III.7 and III.8 it is clear that for $M_{1,2} < \mu$ the lightest chargino is wino-like with a mass given by M_2 if the mixing is neglected and similarly the lightest neutralino is bino like with a mass given by $M_1 \approx 0.5 M_2$.

In practice, there is some mixing and the neutralino mass eigenstates are linear combinations of the weak eigenstates, i.e.

$$\chi_i^0 = N_1 |\tilde{B}> + N_2 |\tilde{W}_3> + N_3 |\tilde{H}_1^0> + N_4 |\tilde{H}_2^0> .$$

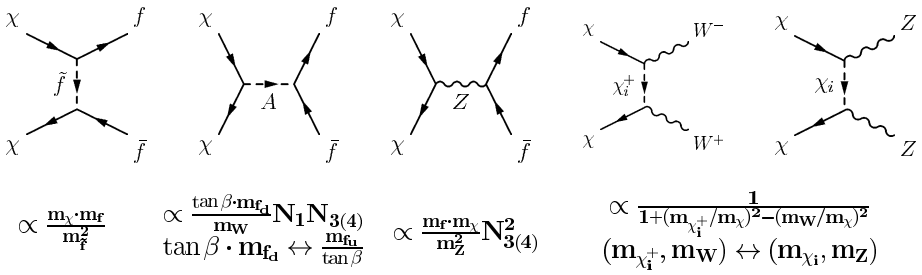

Figure 4. Dominant tree diagrams for Dark Matter annihilation. Note that the amplitudes of the graphs shown at the top are proportional to the mass of the final state fermion, while the Higgs exchange is proportional to $\tan\beta$ for d-type quarks and $1/\tan\beta$ for up-type quarks. This implies that light fermion final states can be neglected and at large $\tan\beta$ the bottom final states dominate.

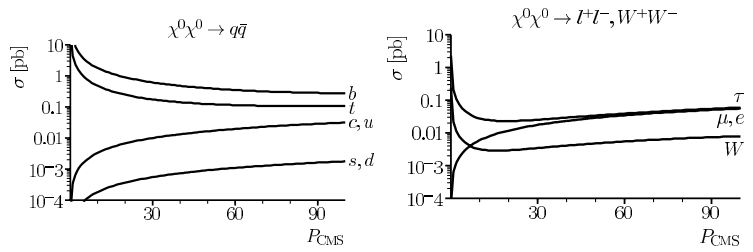

Figure 5. The neutralino annihilation total cross section for $\tan\beta$=35 as function of the center of mass momenta in GeV of the neutralinos for quark -, lepton - and W^+W^- final states, as calculated with CalcHEP [34]. Note the helicity suppression at low momenta for light fermions.

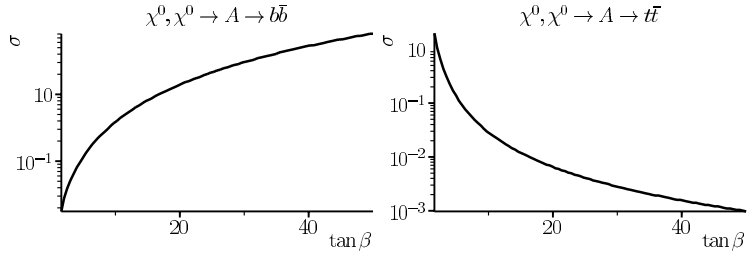

Figure 6. The neutralino annihilation cross section for pseudoscalar Higgs exchange for bottom and top final states as function of $\tan\beta$, as calculated with CalcHEP [34].

The gaugino fraction $N_1^2 + N_2^2$ is nevertheless close to one, especially if the diagonal elements are large compared with the off-diagonal elements proportional to M_Z. This is demonstrated in Fig. 3.

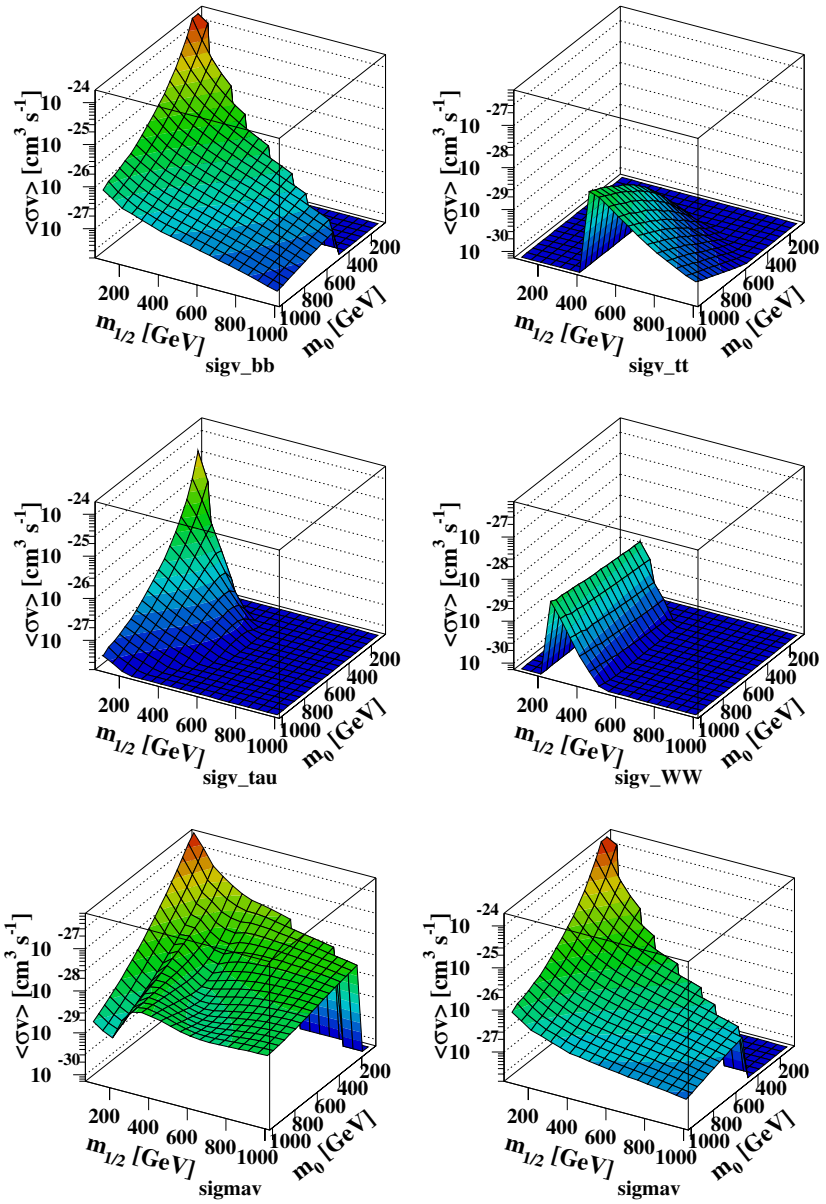

Figure 7. The first two rows show the thermally averaged annihilation cross section times velocity for neutralino annihilation as function of m_0 and $m_{1/2}$ for $\tan\beta = 50$ and $b\bar{b}$, $t\bar{t}$, W^+W^-, and $\tau\bar{\tau}$ final states (clockwise from top left). The last row shows the total cross section or $\tan\beta = 5$ (left) and 50 (right). The neutralino mass equals $\approx 0.4 m_{1/2}$ in the CMSSM, so the neutralino varies from 40 to 400 GeV along the front axis. Note the strong decrease of the cross section for heavier SUSY mass scales and the different vertical scales.

The interaction of the sparticles with normal matter is governed by a new *multiplicative* quantum number called R-parity, which is needed in order to prevent baryon- and lepton number violation. In GUT theories quarks, leptons and Higgses are all contained in the same supermultiplet, which allows couplings between quarks and leptons. Such transitions, which could lead to rapid proton decay, are not observed in nature. Therefore, the SM particles are assigned a positive R-parity and the supersymmetric partners have a negative one, which can be related to the known conserved quantum numbers of spin S, baryon number B and lepton number L by $R = (-1)^{3B+l+2S}$. Requiring R-parity conservation implies that at each vertex one needs *two* supersymmetric particles, from which it follows that:

- The rapid proton decays involving vertices with only one sparticle do not occur.
- Sparticles can be produced only in pairs, e.g. $\overline{p}p \rightarrow \tilde{q}\tilde{g}X$ or $e^+ + e^- \rightarrow \tilde{\mu}^+\tilde{\mu}^-$.
- The heavier sparticles can decay to lighter ones, like $\tilde{e} \rightarrow e\tilde{\gamma}$ or $\tilde{q} \rightarrow q\tilde{g}$, but the Lightest Supersymmetric Particle (LSP) is stable, since its decay into normal matter would change R-parity.
- The LSP has to be neutral to be a good candidate for Dark Matter.
- The interactions of particles and sparticles can be different. For example, the photon couples to electron-positron pairs, but the photino does not couple to electron-positron - or selectron-spositron pairs, since in these cases the R-parity would change from -1 to +1.
- The LSP is weakly interacting with normal matter, since the final state has to contain the LSP again, so its interaction with quarks would only be elastic scattering by e.g. Z-, Higgs or squark exchange.

Consequently, the LSP is an ideal candidate for Dark Matter, since it has all the properties of a Weakly Interacting Massive Particle (WIMP), namely it is neutral, heavy and weakly interacting, so it will form galactic haloes.

The neutralinos can annihilate through the diagrams, shown in Fig. 4. The main features of the amplitudes have been indicated below the diagrams:

- The annihilation into fermion-antifermion pairs is proportional to the fermion mass in the limit v→0, which is the important case in

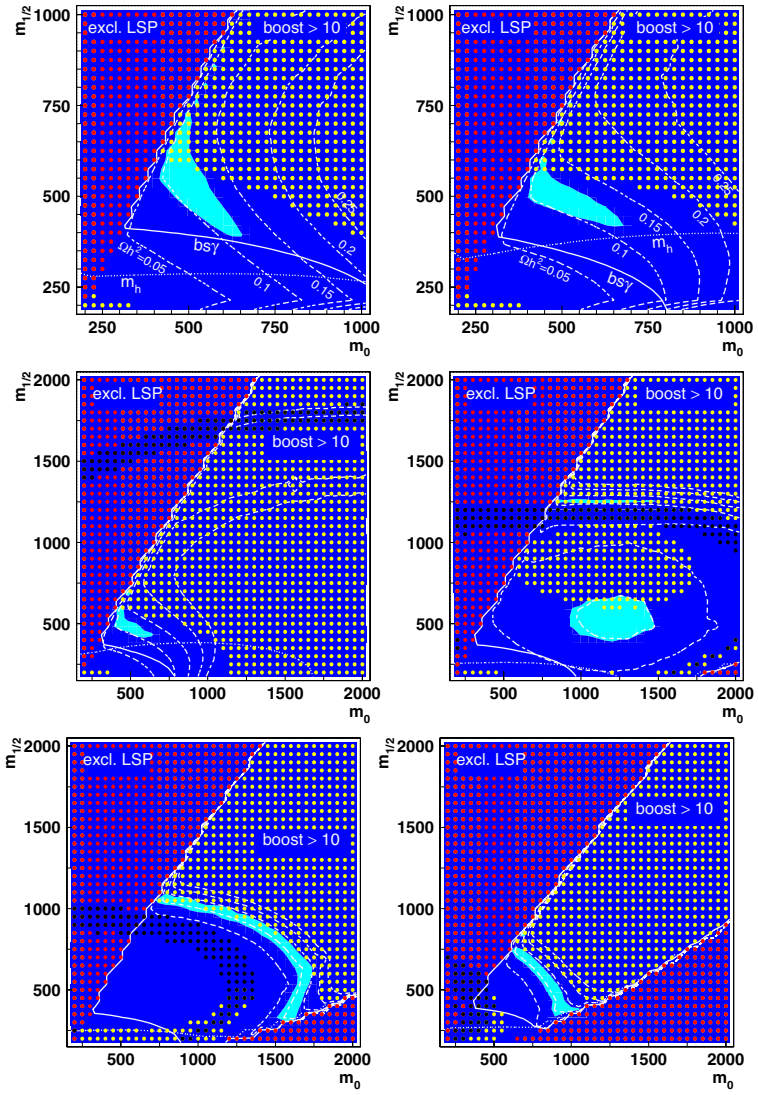

Figure 8. The light shaded (blue) area is the region in the $m_0, m_{1/2}$ plane allowed by the WMAP data ($\Omega h^2 = 0.113 \pm 0.009$) ; the contours of larger Ωh^2 are indicated by the dashed lines in steps of 0.05. The upper row is for $\tan\beta$=51 and $A_0 = 0$ (left) and $A_0 = m_0$ (right), which shows that the role of Higgs constraint (dotted line) and $b \to X_s\gamma$ constraint (solid line) are interchanged for the different values of the trilinear coupling, but the lower limit on $m_{1/2}$ is not very sensitive to A_0. The second (third) row show the same information for a larger region for $\tan\beta$=51, 52 (l. and r.) (53 and 55 (l. and r.)). The excluded regions, where the stau would be the LSP (left top corners) or EWSB fails (right bottom corners) or the boost factors are above 10 are indicated by the dots. The black dots indicate the resonance region, where $|m_A - 2m_{\chi_0}| \leq 10$ GeV. For $\tan\beta > 52$ the acceptable values for the relic densities are for $m_{1/2}$ values above the resonance region.

the present universe at a temperature of a few Kelvin. This can be understood as follows: The neutralino is a Majorana particle, so it is its own antiparticle. In addition it has spin 1/2, thus obeying Fermi statistics, which implies it cannot have identical quantum numbers. Furthermore at low velocity it annihilates into an s-wave state, which implies the spins have to be antiparallel (just like for electrons in the hydrogen s-wave). Therefore also the spins in the final state have to be antiparallel, which leads to an amplitude proportional to the fermion mass to account for the required helicity flip[33]. Consequently heavy final states are enhanced at low momenta, as demonstrated in Fig. 5. Note that at higher momenta not only s-waves contribute and the helicity suppression disappears, so during the time of freeze-out all final states were produced. Note that these arguments are only valid for the diagrams with sfermion and Z-exchange. For the Higgs exchange the proportionality to the final state fermion mass arises from the Yukawa coupling. These cross sections were calculated with the program package CalcHep[34].

- The second important point concerns the $\tan\beta$ dependence: the diagram via pseudoscalar Higgs exchange is proportional to $\tan\beta$ for down-type quarks and $1/\tan\beta$ for up-type quarks. This implies that at large $\tan\beta$ ($\tan\beta > 5$) the b-quark final states are enhanced over t-quark final states, as shown in Fig. 6. The amplitudes of the Higgs exchange and the Z-exchange have an opposite sign. Since the top final states have a large amplitude for Z-exchange (amplitude $\propto$ mass), they are additionally suppressed by the negative interference with the t-channel amplitudes. The cross sections for various final states are shown as function of $m_0, m_{1/2}$ in Fig. 7. The last row shows the strong increase of the total annihilation cross section as function of $\tan\beta$. The strong dependence on the SUSY masses is obvious.

- As shown in Fig. 4, the amplitude for pseudoscalar Higgs exchange is proportional to $N_1 N_{3,4}$, i.e. it requires that the lightest neutralino has both bino- and Higgsino components, which implies that the diagonal elements in the mass matrix III.7 should not be too large compared with the off-diagonal elements proportional to M_Z. So unless one tunes $\tan\beta$ and the SUSY masses such that one hits the resonance ($m_A \approx 2m_\chi$), in which case very small Higgsino admixtures are enough, one needs relatively light neutralino masses.

The regions of parameter space allowed by the WMAP data are plotted in Fig. 8 for different values of $\tan\beta$. It is clear that for $\tan\beta \approx 50$ only a small region is allowed, if in addition the electroweak constraints from the Higgs mass and $b \to X_s\gamma$, and the requirement that the LSP is a neutral particle have to be fulfilled. Values of $\tan\beta$ below 50 are excluded completely, if one wants to be consistent with all constraints and if one requires in addition that the boost factors are below 10. The last requirement implies that clumpiness can enhance the annihilation signal by at most a factor of 10, which is the maximum order of magnitude suggested by simulations of galaxy formation[35, 36]. However, the present simulations have a poor resolution and do not consider interactions between dark matter and normal matter, so this constraint is not a strong one. The boost factors are strongly correlated with the value of the local halo density ρ_χ. To obtain a conservative limit by the requirement that the boost factor is below 10, ρ_χ was set to its maximum allowed value of 0.8 GeV/cm^3. For values of $\tan\beta$ above 50 there are quite a range of neutralino masses allowed, since for larger values of $\tan\beta$ one hits the resonance of pseudoscalar Higgs exchange, in which case much heavier neutralinos are allowed, as shown in Fig. 8 (bottom left). Masses between 100 and 450 GeV are possible.

In summary, the annihilation cross section becomes large for large $\tan\beta$ and is dominated for $\tan\beta > 5$ by the s-channel pseudoscalar Higgs exchange into $b\bar{b}$ quark pairs. Fig. 7 shows that values of $\tan\beta$ around 50 yield the annihilation cross sections required by WMAP, given in Eq. II.2. Regions of coannihilation[37] at smaller $\tan\beta$ are allowed by WMAP data, but the coannihilation with other SUSY sparticles does not operate in the present cooled down universe, which would imply very small annihilation cross sections in the present universe and correspondingly large boost factors needed to explain the deficiencies in the positron, antiproton and gamma ray fluxes. Before discussing the global fits, the cosmic ray fluxes from nuclear interactions are discussed.

IV. Cosmic Rays generated by Nuclear Interactions

The sources of charged and neutral cosmic rays are believed to be supernovae and their remnants, pulsars, stellar winds and binary systems. Observations of X-ray and γ-ray emissions from these objects reveal the acceleration of charged particles near them. Particles accelerated near the sources propagate tens of millions of years in the interstellar medium where they can loose or gain energy and produce secondary particles and γ-rays. The spallation of primary nuclei into secondary nuclei gives rise

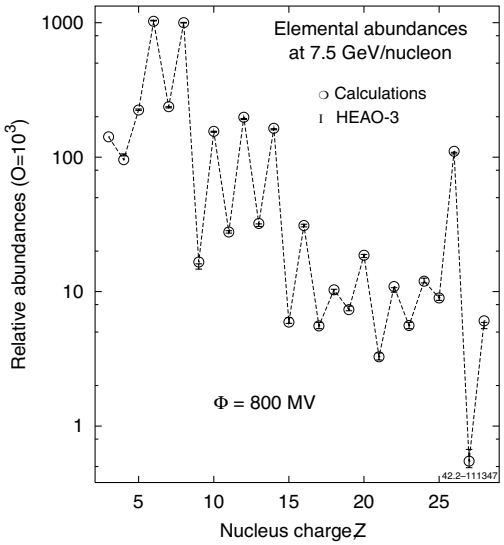

Figure 9. Propagated abundances at 7.5 GeV/nucleon, as calculated by Galprop in comparison with data. From Ref. [9].

to rare isotopes. Nuclear interactions produce not only matter, but also antimatter, like antiprotons and positrons. The latter originate mainly from the decay of charged pions and kaons.

The detailed studies of cosmic rays teach us about the production and propagation in the universe. The gammas can deliver information over intergalactic distances, while the charged particles propagate mainly on galactic distances. Secondary nuclei are produced in the galactic disc, from where they escape into the halo by diffusion and Galactic winds (convection). They may gain energy by "diffusive" reacceleration in the interstellar medium by the 2nd order Fermi acceleration mechanism, i.e. on average more interstellar scattering centers from opposite directions are hitting a given nuclei than "comoving" scatterers. In elastic collisions this leads on average to an energy gain, thus depleting the low energy part of the source spectrum. Long-living radioactive secondaries tell how long they survive in the halo before interacting in the disc again, thus determining the size of the halo. The gas density and acceleration time scale can be probed by the abundances of the K-capture isotopes, which would decay via electron K-capture in the interstellar gas.

Combining this information allows one to build a model of our galaxy. Analytical and semi-analytical models often fail when compared with all data. Therefore advanced models incorporating nuclear reaction net-

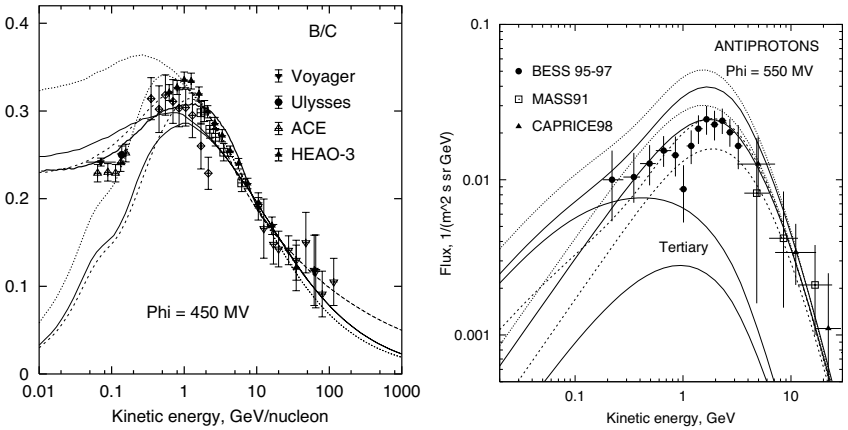

Figure 10. The B/C ratio as an example of secondary/primary nucleon ratios for various Galprop models in comparison with data. The dashed lines are the model with diffusive reacceleration, the solid ones for diffusion plus convection, and the dotted ones for plain diffusion. The lower (upper) curves of each kind are the interstellar (solar modulated) ones. The diffusive reacceleration curve (dashed) provides the best fit to the B/C ratio, but due to the large diffusion coefficient required, this leads to a deficiency in the flux of antiprotons (right). From Ref. [7].

works, cross sections for production of antiprotons, positrons, γ-rays and synchrotron radiation, energy losses, convection, diffusive reacceleration, distribution of sources, gas and radiation field etc. are needed.

In addition, the distributions of matter and antimatter in the interstellar medium (ISM) are modified locally by affects of the solar activity and magnetic fields inside our solar system, e.g. from the planets. Gleeson & Axford [38] modelled the periodically varying solar activity with a typical half cycle of 11 years[1] by a radial solar wind in which the charged particles loose kinetic energy depending on their rigidity R and on the distance r from the sun. On this time scale the incoming flux from the Galaxy does not vary and the problem reduces to an adiabatical deceleration by the solar wind with a dependence only on the radial coordinate. This can be solved analytically:

$$J(r, E, t) = \frac{R^2}{R'^2} J(\mathrm{inf}, E'), \tag{IV.9}$$

where $J(r, E, t)$ is the measured differential flux at a distance r from the sun for particles with energy E and mass E_0^2/c^2, $J(\mathrm{inf}, E')$ is the incom-

[1]A cycle can vary between 8 and 14 yr.

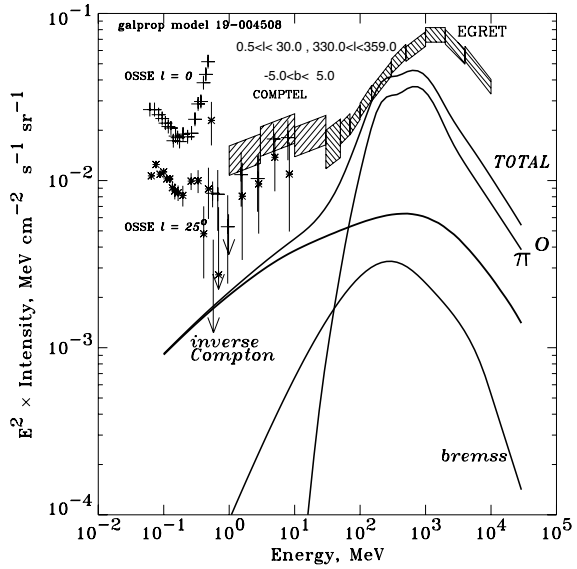

Figure 11. The gamma-ray energy spectrum of the inner Galaxy as calculated by Galprop [8] in comparison with EGRET data. Clearly, there is an access of data above 1 GeV.

ing flux to the solar system and $R^2 = (E^2 - E_0^2)/(Ze)^2$ is the rigidity with $R' = R(E')$. The energy loss can be parametrized by the solar modulation parameter $\Phi(t)$ as $E = E' - |Ze|\Phi(t)$, where $|Ze|$ is the absolute charge of the particle; $\Phi(t)$ varies between 350 and 1500 MV depending on the solar cycle. The solar modulation shifts the particle spectrum to lower energies, but the effect is only noticeable for rigidities below 10 GV. Recent determinations of the local interstellar flux (LIS) from the modulated (=measured) ones for electrons, positrons and protons can be found in Ref. [39].

The most complete and publicly available code for the production and propagation of particles in our galaxy is the GALPROP code [24, 25]. It provides a numerical solution to the transport equation including a cross section database with more than 2000 points, source functions, density distributions, etc. The cross section tables include all possible cross sections: $p+p$, $p+He$, $p+N$, $He+N$, $N+N$, where all nuclei up to the heaviest ones (Ni) are considered. Fig. 9 shows the composition of the primary and secondary nuclei, as calculated by Galprop in comparison with data. Clearly, the production of secondary nuclei is well described.

Fig. 10 shows the spectrum of the Boron over Carbon (B/C) ratio, which shows a characteristic depletion at low energies. Since Boron is a purely secondary produced nuclei, while Carbon is primarily produced, the depletion at low energy is a sensitive handle on the question of diffusive reacceleration and solar modulation. As shown, the modulation effects the spectra mainly at kinetic energies below 10 GeV/nucleon. In order to reproduce the sharp peak in the ratios of secondary to primary nuclei without any unphysical breaks in the energy dependence of the diffusion coefficients and/or the injection spectrum, the diffuse reacceleration with a rather large diffusion coefficient is needed [40, 7]. But this leads to too few antiprotons, as shown by the dashed line on the right hand side in Fig. 10. The possible way out of the discrepancy between the B/C ratio and too few antiprotons was suggested by Strong and Moskalenko: a "fresh" local unprocessed component at low energies of primary nuclei, thus decreasing the B/C ratio and allowing for a smaller diffusion coefficient. Also too few gammas are generated by Galprop, as shown in Fig. 11, which would also need either a harder nucleon spectrum or a harder electron spectrum, but this would need spatial variations which make the spectrum in our local region unrepresentative of the large scale average [8].

However, an alternative explanation may be the annihilation of neutralinos, which increases the yield of gammas, antiprotons, and high energy positrons, but does NOT effect the B/C ratio. This goes exactly in the direction of solving these discrepancies between Galprop and present data *simultaneously*, as will be shown in the next section by a global fit to all data.

V. Global fits to positrons, antiprotons and gamma rays

Trying to disentangle the contributions from nuclear interactions and neutralino annihilation to the antimatter fluxes and gamma rays is in practice not easy. Ideally one would like to implement the neutralino annihilation as a source function in the Galprop code, so the antimatter from nuclear interactions and meutralino annihilation would be transported through the galaxy in an identical way. However, this numerical code is too slow to be used in a fit program. Therefore, we used the second best possibility, namely using the publicly available code DarkSusy [26] for neutralino annihilation, which has semi-analytical solutions to the diffusion equation and includes the important energy losses for positrons. We changed the diffusion parameters and code in DarkSusy in such a way, that the energy losses of positrons and antiprotons resem-

bled as closely as possible the Galprop results and interfaced Galprop to Darksusy in order to get the shape of the background for the fit. The normalization of the background and the annihilation signal were both left free, so only the shapes were fitted.

The Dark Matter halo profile is usually assumed to be of the Navarro, Frenk & White (NFW) type [41], as supported by other studies on galaxy formation [35, 36, 42, 43]. Effectively we have chosen for the Dark Matter density distribution an isothermal spherical symmetric profile:

$$\rho(r) = \rho_0 \cdot (\frac{r}{a})^{-\gamma} \left[1 + (\frac{r}{a})^{\alpha}\right]^{\frac{\gamma-\beta}{\alpha}} ,$$

where a is a scale radius and the slopes α, β and γ can be thought of as the radial dependence at $r \approx a$, $r >> a$ and $r << a$, respectively. For definiteness we use (α, β, γ) =(1,3,1) for a scale $a = 10$ kpc, but e.g. (1.5,2,1) for a=2 kpc yields practically identical results.

With this profile the halo density increases quite steeply towards the center, so most annihilations will take place in the center of the galaxy, thus producing there gamma rays, positrons and antiprotons. The gamma rays can travel over large distances without loosing energy, but they will arrive at the detector only, if they were emitted along the line of sight. Antiprotons and positrons on the other hand change direction during the propagation along the magnetic field lines and by collisions, so they can arrive at the detector even if they were not emitted along the line of sight. This causes a larger acceptance for the antiprotons and positrons in comparison with the gamma rays. On the other hand, many of the positrons and antiprotons loose energy and/or diffuse out of the galaxy. Therefore the absolute fluxes are dependent on the halo profile and propagation parameters. However, the shape of the background spectra is much less dependent on the propagation parameters. Consequently, fitting only the shape of the background, i.e. using a free normalization factor, will be much less model dependent. Also for the annihilation signal the shape is much better known than the absolute flux, since for the dominant decay into $b\bar{b}$ quarks the fragmentation properties of these quarks have been accurately studied at the LEP accelerator in the relevant energy range of about 100 GeV. The absolute flux depends on the possible local density fluctuations around the average density. Since the annihilation rate is proportional to the square of the density, local fluctuations can boost the annihilation by a large factor, although models of galaxy formation predict the boost factors to be "of the order of a few"[35, 36]. They can be somewhat different for gammas, antiprotons and positrons, because of their different range in the galaxy.

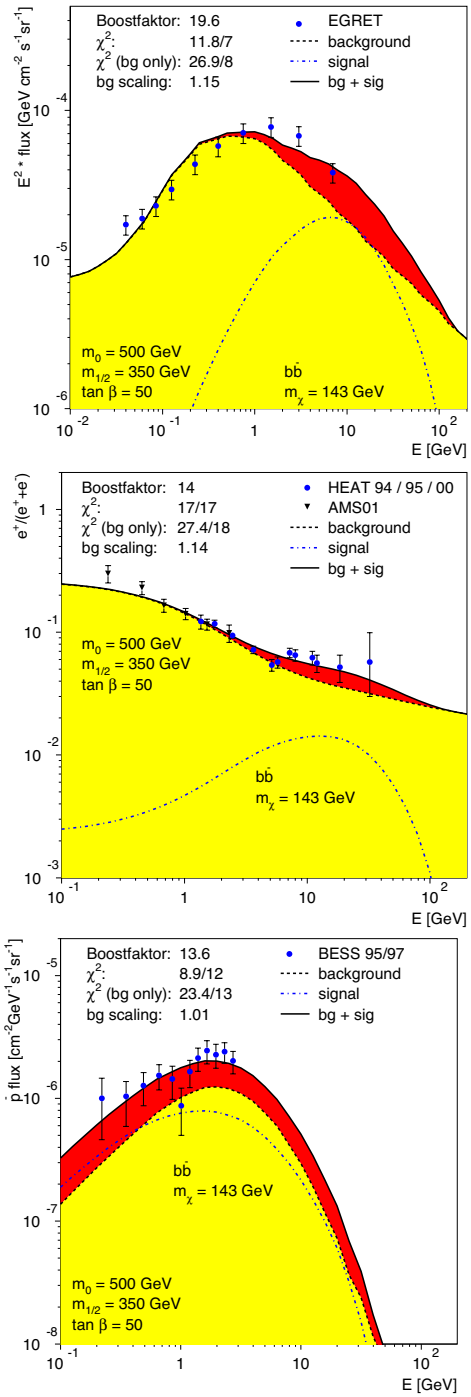

Figure 12. Spectra for gamma rays, positron fraction and antiprotons (from top) with contributions from nuclear interactions (grey/yellow) and neutralino annihilation (dark/red) for a neutralino mass of 143 GeV.

The following data were used in the fit:

- Gamma ray data from the galactic center in the angular range $330^\circ < \ell < 30^\circ$ and $-5^\circ < b < 5^\circ$ from the EGRET space telescope, which has been taking data for about 9 years on the NASA Compton Gamma Ray Observatory (CGRO). We use the data as presented in Ref. [44].
- Positron data from AMS [45] and HEAT [46].
- Antiproton data from BESS in the years 1997 and 1998 [47]

A χ^2 minimization was performed between the combined data and the sum of the annihilation signal and background with the normalizations of signal and background as free parameters. The halo parameters were fixed to a NFW (1,3,1) profile, while the mass of the neutralino was fixed to 143 GeV. Different halo profiles would not change the χ^2 of the fit, only change the boost factors. For the choosen halo profile the boost factors for all particles (gammas, positrons, and antiprotons) come out to be similar. The mass of the neutralino was choosen to be as low as possible to be compatible with the electroweak constraints. The gamma data actually prefers a somewhat lower mass, but the positron fraction a somewhat higher mass. However, the present data are not accurate enough to provide a strong constraint on the mass, so a mass of 143 GeV was choosen. This is compatible with the WMAP data for relatively large values of m_0 and $\tan\beta$, as shown in Fig. 8.

The Galprop program describes well the gamma rays below 1 GeV, but fails for the high energy part of the spectra. The shape of the spectra is better predicted than the absolute flux, since the shape is determined by the measured spectra of the protons, which determine the π_0 production, and the electron spectra, which determines the contributions from bremssthralung and inverse Compton scattering. Therefore the normalization of the background was left free in the fit, so only the shape of the background was fitted.

The χ^2 function for gamma rays was defined as $\sum_i (f_\gamma D_i - T)^2/(f_\gamma \sigma_i)^2 + (1-f_\gamma)^2/\sigma_n^2$, where the sum runs over all data bins with errors σ_i and T is the sum of the parametrized background from nuclear interactions, as calculated by Galprop and the contribution from neutralino annihilation, as calculated by DarkSusy with the modifications mentioned above. f_γ is a common normalization factor for all data points with a systematic error given by σ_n. Normally f_γ was set to one, but sometimes it was left free in the fit in order to study the effect of possible correlations between the data points. Similar χ^2 functions were defined for the antiprotons and positrons and the total χ^2 was simply the sum of the χ^2 contribu-

tions for gamma rays, antiprotons and positrons, since no correlations exist.

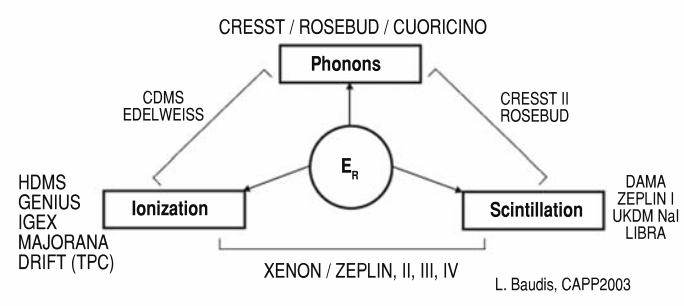

Figure 13. A summary of experiments for direct detection of Dark Matter and the signals they are looking for by the interaction of the WIMP with the detector. The smaller detectors can look for a coincidence of ionization, scintillations or phonons, while the bigger detectors can search for an annual modulation in one of the signals. From Ref. [52].

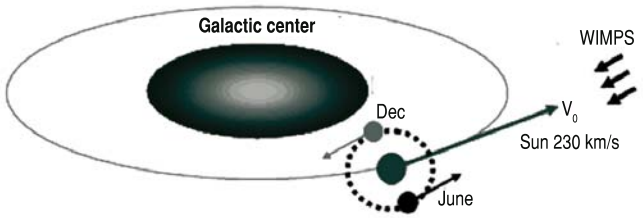

Figure 14. The flux of WIMP's in the detector is proportional to the velocity of the detector with respect to the "sea" of WIMP's, which is maximum on the second of June and minimal in December due to the movement of the earth around the sun. This causes an annual modulation of about 7% in the WIMP detection rate.

The results of the fits are shown in Fig. 12. In total six independent free parameters are fitted, namely the normalization factors for signal and background for gamma rays, positrons and antiprotons. This implies the χ^2 values are independent of the absolute normalizations of signal and background, they only depend on the shape. The latter is determined by the neutralino mass for the signal, while the shape of the

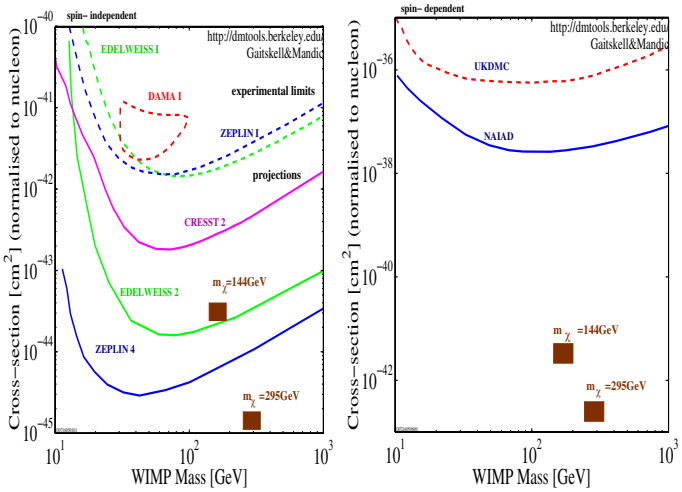

Figure 15. The spin independent (left) and spin dependent (right) cross section for direct WIMP detection on protons in comparison with existing (dashed lines) and projected (solid lines) limits and the cross section expected for neutralino masses of 144 and 295 GeV (squares) with the mSUGRA model discussed in the text.

background is dominated by the π_0 spectrum produced in nuclear interactions. As indicated in the figures the boost factors vary between 14 and 20, so they are similar for antiprotons, positrons and gamma rays for the NRW halo profile discussed before. The $\chi^2/d.o.f.$ is reduced from 90/39 for the "background-only" fit to 39/36 for the fit including neutralino annihilation. This corresponds to about a 4 σ effect, if calculated with Gaussian errors.

Improved data on the positron, antiprotons and gamma rays are expected from the "table-top" PAMELA experiment with a tracker acceptance of $\approx$0.002 m^2 sr, which is scheduled to be launched on a Russian satellite in 2004[48]. Much larger statistics is expected from the AMS-02 experiment[49] with an acceptance of $\approx$0.8 m^2 sr. This detector, which is optimized to identify antimatter, is scheduled to start three years of data taking at the International Space Station towards the end of 2006. Since it will measure the fluxes up to energies of several hundred GeV, it should be able to pin down the mass of the neutralino by measuring the end points of the spectra much more precisely. Similarly, the GLAST satellite can measure high energy gamma rays and is scheduled to start data taking in 2006[50]. Large area neutrino experiments could look for the neutrinos from neutralino annihilation in the sun[51].

V.1 Direct Searches

Dark Matter can be searched for directly by trying to detect the elastic scattering of a WIMP in a detector. The difficulty is the low scattering cross section and the low signal compared with the background. The background can be suppressed by considering two signals in coincidence, e.g. the heat from the lattice vibrations (phonons) and the ionization or scintillation induced by the displacement of the hit nucleus[6]. The heat can be measured at low temperatures of a few mK, where the heat capacity ($\propto T^3$) is sufficiently small to give a measurable rise in temperature. The ionization can be measured in semiconductor detectors by applying an electric field to the detector, which drives the electron/hole pairs to the collecting electrodes. If the detector is made from a scintillation material, e.g. NaI crystals, the scintillation light can be detected by photomultipliers. This technique of simultaneously measuring the heat and the signal from the displacement gives excellent discrimination against gamma ray background e.g. from decays of radioactive nuclei, since nuclear recoils from a heavy projectile, like WIMPS, produce relatively more phonons than ionization or scintillation, while photons produce relatively more ionization compared to phonons. Many experiments with all these different techniques are in progress, as summarized in Fig. 13.

The cross section for WIMP-nucleon scattering is proportional to the velocity between WIMP and target. The WIMP is practically at rest, but the detector flies with the speed of our solar system, modulated with the orbital speed of the earth around the sun. This allows one to search for an annular modulation in the signal with an amplitude of about 7%. A 6.3σ annual modulation signal has been observed by the DAMA experiment, which looks for a scintillation signal in extremely pure NaI crystals[53]. The experiment has been running for about 7 years and the observed period of the modulation is one year with the maximum near the expected date of June 2nd. Unfortunately, the signal has not yet been confirmed by other experiments with a similar sensitivity[54].

The various direct search results and the indirect searches are compared in Fig. 15, both for the spin-independent and spin-dependent scattering on protons. The scattering on neutrons is similar. The spin-independent cross section, mainly mediated by Higgs exchange, couples proportional to the mass, while the spin-dependent cross section couples mainly to the spin of the nucleus. The sensitivity for spin-independent scattering is usually higher, especially on heavier nuclei, because of the coherence of the scattering: a WIMP with low momenta has a wavelength of the order of the size of a nucleus, thus scattering coherently on all nuclei, which enhances the mass and therefore the amplitude by

a factor A for A nuclei and the cross section is then proportional to A^2. Complete formulae have been summarized in the nice review by Jungman, Kamionkowski and Griest[6].

The cross sections for *direct* Dark Matter detection for WIMP masses of 144 and 295 GeV were calculated with DarkSusy [26], using the SUSY parameters discussed for the indirect detection. For the heavier masses the cross section rapidly decreases, but the projected sensitivity of future direct detection experiments may be sufficient, as demonstrated in Fig. 15.

VI. Conclusion

It is shown that the discrepancies between EGRET data and the galactic models can be reduced by taking as an additional source of hard gammas the annihilation of Dark Matter, assuming Dark Matter is made of neutralinos, as predicted by Supersymmetry. In addition, it is shown that adding the positrons from neutralino annihilation in the *same* Dark Matter model to the *same* background model improves also the χ^2 fit to the positron data significantly, while the increase in antiprotons is compatible with the data. These facts, statistical significant improvement of the global fit for positrons, antiprotons and gamma rays *simultaneously* for a supersymmetric model with an annihilation cross section compatible with the model-independent WMAP value, provide strong experimental evidence for the supersymmetric nature of Dark Matter.

I thank V. Moskalenko and A. Strong for sharing with us all their knowledge about our galaxy, O. Reimers to provide us with the EGRET data and my close collaborators M. Herold, A. Gladyshev, D. Kazakov and C. Sander for help during this analysis. This work was supported by the DLR (Deutsches Zentrum für Luft- und Raumfahrt).

References

[1] The results of the first year of operation of the WMAP satellite can be found on the Web: http://map.gsfc.nasa.gov/m_mm/pub_papers/firstyear.html

[2] C.L. Bennett, et al., astro-ph/0302207.

[3] D.N. Spergel, et al., astro-ph/0302209.

[4] L. Verde et al., astro-ph/0302218

[5] J. Ellis et al., Nucl. Phys. **B238** (1984) 453.

[6] G. Jungman, M. Kamionkowski and K. Griest, Phys. Rep. **267** (1996) 195.

[7] I.V. Moskalenko, A.W. Strong, J.F. Ormes and M.S. Potgieter, Astrophys. J. **565** (2002) 280.

[8] A.W. Strong, I.V. Moskalenko and O. Reimer, Astrophys. J. **537** (2000) 763.

[9] A.W. Strong, I.V. Moskalenko, S.G. Mashnik and J.F. Ormes, Astrophys. J. **586** (2003) 1050.

[10] L. Bergström, Rept. Prog. Phys. **63** (2000) 793 [arXiv:hep-ph/0002126];

[11] J. L. Feng, K. T. Matchev and F. Wilczek, Phys. Rev. D **63** (2001) 045024 [arXiv:astro-ph/0008115].

[12] A. Morselli, Int. J. Mod. Phys. A **17** (2002) 1829.

[13] F. Donato, N. Fornengo, D. Maurin, P. Salati and R. Taillet, arXiv:astro-ph/0306207.

[14] J. Edsjö, arXiv:astro-ph/0211354.

[15] E. A. Baltz, J. Edsjö, K. Freese and P. Gondolo, Phys. Rev. D **65** (2002) 063511 [arXiv:astro-ph/0109318].

[16] L. Bergström, J. Edsjö and C. Gunnarsson, Phys. Rev. D **63** (2001) 083515 [arXiv:astro-ph/0012346].

[17] P. Ullio, L. Bergström, J. Edsjö and C. Lacey, Phys. Rev. D **66** (2002) 123502 [arXiv:astro-ph/0207125].

[18] A. Cesarini, F. Fucito, A. Lionetto, A. Morselli and P. Ullio, arXiv:astro-ph/0305075.

[19] P. Ullio, Int. J. Mod. Phys. A **17** (2002) 1777.

[20] A. Bottino, F. Donato, N. Fornengo and P. Salati, Phys. Rev. D **58** (1998) 123503 [arXiv:astro-ph/9804137].

[21] G. Jungman and M. Kamionkowski, Phys. Rev. D **51** (1995) 3121 [arXiv:hep-ph/9501365].

[22] A. Bottino, V. de Alfaro, N. Fornengo, A. Morales, J. Puimedon and S. Scopel, Mod. Phys. Lett. A **7** (1992) 733.

[23] L. Bergström, P. Ullio and J. H. Buckley, Astropart. Phys. **9** (1998) 137 [arXiv:astro-ph/9712318].

[24] A.W. Strong and I.V. Moskalenko, Astrophys. J. **509** (1998) 212.

[25] I.V. Moskalenko and A.W. Strong, Astrophys. J. **493** (1998) 694.

[26] DarkSUSY, P. Gondolo, J. Edsjö, L. Bergström, P. Ullio and E. A. Baltz, arXiv:astro-ph/0012234 and http://www.physto.se/ẽdsjo/darksusy/.

[27] A. Djouadi, J. L. Kneur and G. Moultaka, arXiv:hep-ph/0211331.

[28] S. Heinemeyer, W. Hollik and G. Weiglein, arXiv:hep-ph/0002213.

[29] G. Bélanger, F. Boudjema, A. Pukhov and A. Semenov, Comp. Phys. Commun. **149** (2002) 103 [arXiv:hep-ph/0112278] and arXiv:hep-ph/0210327 and http://wwwlap.in2p3.fr/lapth/micromegas.

[30] E. Kolb, M.S. Turner, The Early Universe, Frontiers in Physics, Addison Wesley, 1990.

[31] Reviews and original references can be found in: W. de Boer, Prog. Part. Nucl. Phys. **33** (1994) 201 [arXiv:hep-ph/9402266];
H.E. Haber, Lectures given at Theoretical Advanced Study Institute, University of Colorado, June 1992, Preprint Univ. of Sante Cruz, SCIPP 92/33; see also SCIPP 93/22;
Perspectives on Higgs Physics, G. Kane (Ed.), World Scientific, Singapore (1993);
A.B. Lahanus and D.V. Nanopoulos, Phys. Rep. **145** (1987) 1;
H.E. Haber and G.L. Kane, Phys. Rep. **117** (1985) 75;
M.F. Sohnius, Phys. Rep. **128** (1985) 39;
H.P. Nilles, Phys. Rep. **110** (1984) 1;
P. Fayet and S. Ferrara, Phys. Rep. **32** (1977) 249.

[32] W. de Boer and C. Sander, arXiv:hep-ph/0307049 and references therein.

[33] H. Goldberg, Phys. Rev. lett. **50** (1983) 1419.

[34] A. Pukhov *et al.*, arXiv:hep-ph/9908288.

[35] F. Stoehr, S. D. White, V. Springel, G. Tormen and N. Yoshida, arXiv:astro-ph/0307026;

[36] V. Berezinski, V. Dokuchaev, Y. Eroshenko, arXiv:astro-ph/0301551.

[37] K. Griest and D. Seckel, Phys. Rev. **D43** (1991) 3191

[38] L.J. Gleeson and W.I. Axford, ApJ **149** (1967) L115; ApJ **154** (1968) 1011.

[39] D. Casadei and V. Bindi, astro-ph/0302307.

[40] M. Simon, A. Molnar and S. Rösler, ApJ **499** (1998) 250;

[41] J.F. Navarro, C.S. Frank and S.D. White, ApJ **490** (1997) 493, [arXiv:astro-ph/9611107].

[42] S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn and J. Stadel, arXiv:astro-ph/9910166.

[43] D. Zhao, H. Mo, Y. Jing and G. Boerner, Mon. Not. Roy. Astron. Soc. **339** (2003) 12 [arXiv:astro-ph/0204108].

[44] A. W. Strong, I. V. Moskalenko and O. Reimer, arXiv:astro-ph/0306345.

[45] J. Alcaraz *et al.* [AMS Collaboration], Phys. Lett. B **484** (2000) 10 [Erratum-ibid. B **495** (2000) 440].

[46] S. W. Barwick *et al.* [HEAT Collaboration], Astrophys. J. **482** (1997) L191 [arXiv:astro-ph/9703192].
M. A. DuVernois *et al.*, Astrophys. J. **559** (2001) 296.

[47] BESS Coll. S. Orito et al., Phys. Rev. Lett **84** (2000) 1078. T. Maeno et al., Astrop. Phys. **16** (2001) 121; astro-ph/0010381.

[48] V. Bonvicini *et al.*, [PAMELA Collaboration], Nucl. Instrum. Meth. A **461** (2001) 262.

[49] R. Battiston, The Alpha Magnetic Spectrometer (AMS), Nucl. Instr. and Meth. **A409** (1998) 458.
The AMS Coll., B. Alpat, The Alpha Magnetic Spectrometer (AMS) on the International Space Station, Nucl. Instr. and Meth. **A461** (2001) 272.
The AMS Coll., J. Alcaraz et al., The Alpha Magnetic Spectrometer (AMS), Nucl. Instr. and Meth. **A478** (2002) 119.

[50] A. Morselli, A. Lionetto, A. Cesarini, F. Fucito and P. Ullio [GLAST Collaboration], Nucl. Phys. Proc. Suppl. **113** (2002) 213 [arXiv:astro-ph/0211327].

[51] J. Ahrens *et al.* [AMANDA Collaboration], Phys. Rev. D **66** (2002) 032006 [arXiv:astro-ph/0202370];
F. Blanc *et al.* [ANTARES Collaboration], *Presented by L. Thompson on behalf of the ANTARES Collaboration, to appear in the proceedings of 28th International Cosmic Ray Conferences (ICRC 2003), Tsukuba, Japan, 31 Jul - 7 Aug 2003;*
S. E. Tzamarias [NESTOR Collaboration], Nucl. Instrum. Meth. A **502** (2003) 150. For a recent review, see F. Halzen and D. Cooper, Rept. Prog. Phys. **65** (2002) 1025 [arXiv:astro-ph/0204527].

[52] L. Baudis, Workshop on Cosmology and Particle Physics, CAPP 2003, CERN, June, 2003.

[53] R. Bernabei *et al.*, Riv. Nuovo Cim. **26** (2003) 1 [arXiv:astro-ph/0307403].

[54] O. Martineau *et al.* [EDELWEISS Collaboration], arXiv:astro-ph/0310657.
R. W. Schnee *et al.*, Nucl. Phys. Proc. Suppl. **124** (2003) 185.

CARGESE LECTURES ON EXTRA-DIMENSIONS

Riccardo Rattazzi
Department of Physics, Theory Division
CERN, CH-1211 Geneva 23, Switzerland
Riccardo.Rattazzi@cern.ch

Abstract I give a pedagogical introduction to the concepts and the tools that are necessary to study particle physics models in higher dimensions. I then give a more detailed presentation of warped compactifications and discuss their possible relevance to the hierarchy problem.

Keywords: Extra-dimensions, branes, effective action, effective field theories, gravity, hierarchy, Kaluza-Klein fields, large extra-dimensions, The Randall-Sundrum model, radions, moduli

Introduction

Science in general and particle physics in particular thrive from conceptual puzzles and unexplained phenomena. The gauge hierarchy problem is an exemplar source for inspiration. While we haven't got yet any direct experimental evidence onto what mechanism sets the small ratio $G_N/G_F \sim 10^{-34}$ between the Newton and Fermi constants, a great deal of theoretical progress in particle physics has been triggered in trying to come up with an explanation. For instance, the great development in supersymmetric field theory of the last three decades is to a good extent motivated by the potential relevance of supersymmetry to the hierarchy problem. The last few years have also witnessed a great revival in the interest for models with extra space dimensions. On one side this revival is motivated by important theoretical developments within superstring theory, in particular by the realization that there exist in string theory solitonic membranes, D-branes, on which ordinary particles could

D. Kazakov and G. Smadja (eds.), Particle Physics and Cosmology: The Interface, 461–517.

be localized [1]. On the other side the revival is also phenomenologically motivated by the realization that extra-dimensions can shed a new light on the hierarchy problem [2].

The potential relevance of extra-dimensions to the hierarchy problem can be grasped by the following simple line of reasoning. One way to phrase the hierarchy is that the Standard Model (SM) quanta, like the Z-boson, are much softer than the quanta of a possible underlying Grand Unified Theory (GUT) or string theory: $m_Z/m_{GUT} \sim 10^{-14}$. That is to say that the minimum frequency corresponding to a travelling Z boson wave is $m_Z \sim 10^2$ GeV while the minimal frequency of a GUT wave is $\sim 10^{16}$ GeV. Supersymmetry or technicolor allow for a dynamical explanation of this huge hierarchy. Moreover in both cases the value of m_Z/m_{GUT} is determined by a quantum phenomenon, *i.e.* dimensional transmutation. However we know since long of a basic classical phenomenon that can make quanta softer: gravitational redshift. Let us briefly recall how this works in general relativity. Consider a gravitational field specified by a metric $g_{\mu\nu}(x)$. The invariant interval separating event x from event $x + dx$ is given by $(ds)^2 = g_{\mu\nu}dx^\mu dx^\nu$. By the Equivalence Principle $(ds)^2$ equals the Lorentz invariant interval measured by any freely falling observer at x: $(ds)^2 = -(\Delta X_0)^2 + (\Delta X_1)^2 + (\Delta X_2)^2 + (\Delta X_3)^2$. Consider the case of a static metric $g_{\mu\nu}$ and of an interval dx in the time direction. In this case $(ds)^2 = g_{00}(dx^0)^2 < 0$ corresponds to the proper time interval experienced by a freely falling observer with zero velocity at x

$$d\tau = \sqrt{-g_{00}(x)}dx^0 \tag{.1}$$

By the Equivalence Principle any clock at rest at x will oscillate with its proper period $\Delta T = 1/\omega$ according to the same freely falling observer. Then by eq. (.1) the frequency $\omega(x)$ observed in the original reference frame will be rescaled according to

$$\frac{1}{\omega} = \sqrt{-g_{00}(x)}\frac{1}{\omega(x)}. \tag{.2}$$

This rescaling is not yet, by itself, a physical effect: also the unit of measure of time at x, specified by some standard clock, will undergo the same rescaling. (Another way to state this is that the overall normalization of the metric, upon which eq. (.2) crucially depends, is not an observable as it depends on the units of measure.) An observable effect arises when comparing the frequencies of two copies of the same clock located at different points. For instance we can consider two hydrogen atoms located respectively at points A and B in the gravitational field of a star (see Fig. (1)), and associate the frequency ω to a given

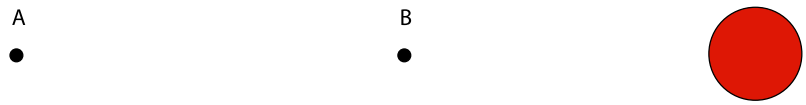

Figure 1. Two atoms A and B in the gravitational field of a star.

atomic transition. The waves emitted at A will oscillate everywhere with frequency $\omega_A = \sqrt{-g_{00}(A)}\omega$. This is because the background is time independent, and two neighbouring wavecrests, leaving A with delay $1/\omega_A$, will take the same time to reach any given point, for example point B. Similarly the wave emitted at B will oscillate everywhere with frequency $\omega_B = \sqrt{-g_{00}(B)}\omega$. The ratio

$$\frac{\omega_B}{\omega_A} = \sqrt{\frac{g_{00}(B)}{g_{00}(A)}} < 1 \qquad (.3)$$

represents a physical effect. The observer A notices that the light emitted from position B, closer to the star, is redshifted and similarly observer B notices that the light emitted at A is blue-shifted. This effect can be qualitatively understood as the photon loosing kinetic energy as it climbs up to A from deep inside the gravitational potential well at B. In everyday's life this gravitational redshift represents a tiny effect, since the gravitational field of the earth is rather weak. However gravity is a non-linear theory encompassing large gravitational fields, like the one near the horizon of a black hole. Assume indeed that the star of the previous example has collapsed to form a black hole. In this case the metric is given as a function of the radial coordinate r by $g_{00} = -1 + 2G_N M/r \equiv -1 + r_S/r$ (with M the black hole mass). As the position r_B of atom B approaches the horizon r_S we have $g_{00}(r_B) \to 0$ so that the the redshift becomes infinite! The infinite redshift of photons emitted at the horizon corresponds to the fact that light cannot escape from a black-hole. We can say that the gravitational field of a black hole creates an infinite *hierarchy* of energies of the emitted photons as the emitter is moved towards the horizon. With the gauge hierarchy in mind it is perhaps then natural to think of a wild generalization of the system we just discussed, one in which the points A or B are generalized to 3-dimensional spacelike surfaces, or 3-branes. In this process the

dynamical system living on a point, the atom, is generalized to the dynamical system that lives on a 3-surface, a 3+1 quantum field theory, for instance the Standard Model. Imagine then to place two identical copies of 3-branes hosting the Standard Model at different points inside a gravitational field, in Fig. (2), in a straightforward generalization of Fig. (1). Of course, since three space dimensions already span the membranes, the distance separating them must correspond to a new spacelike dimension, the fifth dimension. In our generalization the role of the atomic energy levels (and thus of the emitted frequencies) is played by the masses of the particles living on the branes. For instance the masses of the two identical Z bosons satisfy

$$\frac{m_Z(B)}{m_Z(A)} = \sqrt{\frac{g_{00}(B)}{g_{00}(A)}} < 1. \qquad (.4)$$

If only gravity propagates in the fifth dimension, one experimental consequence of eq. (.4) is that the gravitons emitted in radiative Z decays at point A are more energetic to observer B than those emitted in the same process at B. Also one can't stop from imagining a situation where brane B is much deeper than brane A inside a gravitational field, perhaps even very close to a horizon: then one would expect a huge hierarchy for the masses of identical particles living on the two different branes. Of course the example we are considering is not directly applicable to the gauge hierachy problem, as that does not concern two identical copies of the SM. Nevertheless the redshifting mechanism would obviously be at work even if the field theories living on the two branes where not the same, and also in more complicated situations where the SM degrees of freedom are not fully localized: it is just kinematics. Is it then possible that the weak scale hierarchy originates as a consequence of gravitational redshift in extra-dimensions? The answer to this question is affirmative and the model that proves it was proposed by Randall and Sundrum in a pioneering paper [3].

I. Part I: Extra-Dimensions

In this section we introduce the players in the game (gravity, branes and localized fields) and discuss the rules that govern their effective action.

I.1 Gravity and Branes

Gravity plays a central role in the physics of extra-dimensions. This is shown for instance by the example of the previous section. It is then important to recall the basic concepts relavant to describing the dynamics

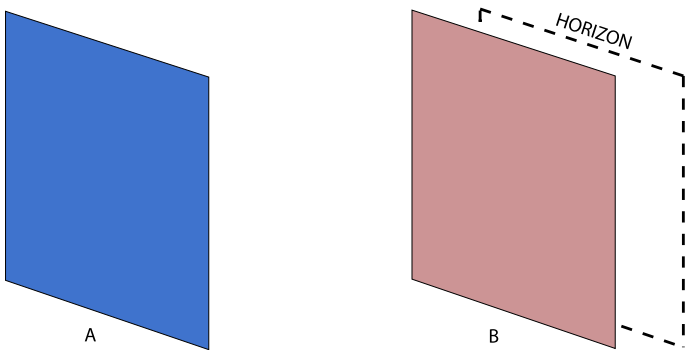

Figure 2. Two branes A and B in a 5-dimensional gravitational field.

of gravity. A basic introduction to General Relativity is taken for granted here. We will be concerned with a D-dimensional space-time with coordinates X^M and metric $g_{MN}(X)$, with $M, N = 0, \ldots, D-1$. We use the mostly plus convention wherein at any space time point a locally inertial frame can be found in which $g_{MN} = \eta_{MN} \equiv (-1, +1, \ldots, +1)$. Later on we will specialize to the phenomenologically relevant case where $D-4$ space-like dimensions are compactified. The metric, up to diffeomorphisms, contains the dynamical degrees of freedom of gravity. The affine connection $\Gamma^R_{MN} = g^{RS}(\partial_M g_{NS} + \partial_N g_{MS} - \partial_S g_{MN})/2$ defines parallel transport, by means of which the Riemann tensor, characterizing the spacetime curvature, is constructed $\mathcal{R}^S_{MNR} = \partial_N \Gamma^S_{MR} - \partial_M \Gamma^S_{NR} + \Gamma^T_{MR}\Gamma^S_{TN} - \Gamma^T_{NR}\Gamma^S_{TM}$. Using the Riemann tensor and its contracted forms, the Ricci tensor $\mathcal{R}_{MR} = \mathcal{R}^N_{MNR}$ and Ricci scalar $\mathcal{R} = \mathcal{R}_{MN} g^{MN}$, the most general invariant action can be written as

$$S = \int d^D X \sqrt{g} \left\{ a M_D^D + 2 M_D^{D-2} \mathcal{R} + b M_D^{D-4} \mathcal{R}^2 + \ldots c M_D^{D-6} \mathcal{R} \Box \mathcal{R} + \ldots \right\} \tag{I.5}$$

where M_D is D-dimensional Planck scale. In ordinary 4-dimensional Einstein gravity, according to our non-conventional normalization, we have $M_4 = (32\pi G_N)^{-1/2} \simeq 1.2 \times 10^{18}$ GeV. We have parametrized all the couplings with dimensionless coefficients $a, b, c, \ldots$ and organized the lagrangian as a derivative expansion

$$a M_D^D p^0 + M_D^{D-2} p^2 + b M_D^{D-4} p^4 + \ldots + c M_D^{D-6} p^6 + \ldots \tag{I.6}$$

where the lowest terms are those that are important at the longest distance scales. In particular the first term, the cosmological constant, influences directly the global structure of the space-time. From the effective lagrangian point of view, which we will discuss in more detail

later, we can meaningfully address only those phenomena that involve a finite number of terms, i.e. those for which p/M_D is significatively less than 1. In this respect, although one would naively expect all the coefficients $a, b, c, \ldots$ to be $O(1)$, we will assume that the cosmological term a is much smaller than 1. This is mostly for a theoretical reason: when $a = O(1)$ the background solution to Einstein equations has $\mathcal{R} = O(M_D^2)$ for which the derivative expansion breaks down. But notice also that there is a phenomenological preference to work with small cosmological constant. On one side we know from observation that the effective cosmological constant of our macroscopic 4-dimensional world is very small: $\Lambda^4 = a_4 M_4^4 \sim (10^{-3}\text{eV})^4 \ll M_4^4$. This may be an indication that also the fundamental cosmological constant aM_D^{D-2}, before compactification, is small. Moreover, as mentioned in sec. I.6, a small D-dimensional cosmological constant is also favored in the scenario of large extra-dimensions [2] by the requirement of a flat potential for the radius modulus. So, while we will assume $a \ll 1$ as a result of some tuning or, perhaps, D-dimensional supersymmetry, for all the other coefficients we will just need the perfectly natural and weak assumption that they be $\leq O(1)$. For instance in ordinary General Relativity with the above assumption the Einstein Lagrangian is a successful truncation up to the very small Planck length $\lambda_P = 1/M_4 = 10^{-33}$ cm. This is evident with the above classical Lagrangian, but as discussed later, it remains true also at the quantum level.

The second important player in the extra dimensional game is given by the so called (mem)branes. They are extended objects which span surfaces and on which excitations (particles) can be localized. An explicit physical example of a brane is given, for instance, by the surface separating two different metals, where there exist localized excitations in the charge density, the surface plasmons. Another explicit example of a brane can be provided by a domain wall. Consider a scalar field theory with a Z_2 invariant potential $V(\phi) = \frac{g}{2}(\phi^2 - v^2)^2$. In addition to the two vacuum solutions $\phi = \pm v$ this model contains domain wall solutions

$$\phi = v \tanh(m(z - z_0)) \tag{I.7}$$

where $m = \sqrt{g}v$ is the mass of the scalar field, while z indicates one of the space directions. This solution interpolates between the $\phi = -v$ vacuum at $z = -\infty$ and the $\phi = +v$ vacumm at $z = +\infty$ and is thus topologically stable [4]. With respect to the true vacuum $\phi = \pm v$ this solution has an energy density $E = (\partial_z \phi)^2 \propto \cosh^{-4}(m(z - z_0)$ localized within a distance $\sim 1/m$ from the center of the wall. Integrating E across the wall we obtain the wall tension $\tau = 4\sqrt{g}v^3/3$. Away from

the wall the scalar has mass m so that at $E < m$ there are no modes propagating through the full space. There is however a massless scalar mode localized at the wall. Indeed the domain wall solution breaks spontaneously the original D-dimensional Poincaré group down to the $(D-1)$-dimensional one, corresponding to translations and boosts in the directions parallel to the wall. As it happens for ordinary internal symmetries, we then expect the presence of Goldstone bosons associated to the broken generators. A naive application of that result to space-time symmetries is however not possible[7, 8]. The Goldstone theorem is proven by considering the local tranformations associated to the global symmetries. In the case of the Poincaré group both translations and boosts reduce, locally, to local translations. The Goldstone bosons are then in a one to one correspondence with the broken translation generators [7]. In the case at hand the translations along the z direction are broken and z_0 parametrizes the manifold of equivalent vacua. Like in the case of internal symmetries we can parametrize the Goldstone excitation by promoting z_0 to a field $z_0(x)$ depending on the $(D-1)$ longitudinal coordinates. At linearized level it corresponds to a mode

$$\delta\phi(x,z) = -\frac{vm}{\cosh^2 mz}\, z_0(x) \tag{I.8}$$

which is clearly normalizable and localized within a distance $1/m$ around the wall. Under z-translations $\phi(z,x) \to \phi(z+a,x)$ we have $z_0(x) \to z_0(x) - a$. Because of this non linear symmetry the action can depend on z_0 only through its space time derivatives. In particular there is no mass term for $z_0(x)$. The effective action for z_0, valid at momentum $\ll m$, can be carefully derived by integrating out the massive excitations of the original field ϕ. However it is intuitively clear what result to expect at lowest order. In this limit we are considering very smooth deformations of the wall, such that its position varies appreciably only over distances much bigger that its width $1/m$. At each point we then expect the field to be given approximately by eq. (I.7), but with z replaced by the direction locally orthogonal to the wall. Then the action will just be given by the integral of the original wall tension $\tau = \sqrt{\lambda} v^3$ over the volume of the deformed wall.

Other fields can be localized at a domain wall. A fermion ψ with Yukawa interaction $\lambda\psi\bar{\psi}$ wil be massive in the bulk, but will contain zero modes localized at the wall where $\phi = 0$ [5]. Similarly, ideas to localized gauge fields have been proposed [6]. Likewise, D-branes in string theory support localized modes (scalars, fermions and vectors) associated to open strings ending on them [1]. In these lectures we will be focusing on the low energy description of branes. For instance, in the case we just considered this corresponds to $E \ll m$. In this regime we

will not be concerned with the microscopic mechanism that gave origin to the brane and to the fields localized on it. We will just assume that the brane hosts a field theory of our choice and derive the consequences. The presence of some degrees of freedom, like the Goldstone z_0 above, could however just follow from symmetry considerations and not be an option.

I.2 Brane Effective Actions

We will now discuss the dynamics of branes by writing the most general effective action satisfying some basic principles. We follow closely the presentation given in ref. [7]. Let us consider an n-brane, a membrane filling n spatial dimension whose spacetime trajectory, the worldbrane, is an $n+1$-surface. We parametrize this surface with coordinates x^μ, $\mu = 0, \ldots, n$. The embedding in the full D-dimensional spacetime is described by D functions $X(x)^M$, with $M = 0, \ldots, D-1$. For instance, in the simplest case of a point particle, a 0-brane, the worldbrane is the particle trajectory, the worldline, parametrized by a time coordinate x^0: $X(x^0)$. It is physically intuitive that the distance between points on the brane, as measured by a brane observer, be the same as measured by a bulk observer,

$$\begin{aligned} ds^2|_{brane} &= G_{MN}\left(X(x)\right) dX^M|_{brane} dX^N|_{brane} \\ &= G_{MN}\left(X(x)\right) \partial_\mu X \partial_\nu X^N dx^\mu dx^\nu \end{aligned} \tag{I.9}$$

i.e. the bulk metric gives rise to an induced metric $\hat{g}_{\mu\nu}$ on the brane

$$\hat{g}_{\mu\nu}(x) = G_{MN}\left(X(x)\right) \partial_\mu X^M \partial_\nu X^N. \tag{I.10}$$

Notice that the induced metric is a scalar under the bulk diffeomorphisms (all the $M, N...$ indices are contracted) while it is a tensor under reparametrizations of the brane $x^\mu = x^\mu(x')$. As the choice of coordinates x is arbitrary, physical quantities should not depend on it. Therefore, like in ordinary gravity, starting from $\hat{g}_{\mu\nu}$ we should write an action invariant under brane reparametrizations. We will do that in a moment. Before then we want to emphasize that, like we have projected the metric, so we can do with other tensors. For instance a bulk gauge field (1-form) $A_M(X)$ leads to a brane field $\hat{A}_\mu(x) = A_M\left(X(x)\right) \partial_\mu X^M$ which under the bulk gauge tranformation $A_M \to A_M + \partial_M \alpha$ shifts as under a proper $n+1$-dimensional gauge transformation

$$\delta \hat{A}_\mu(x) = \partial_M \alpha\left(X(x)\right) \partial_\mu X^M(x) = \partial_\mu \left[\alpha\left(X(x)\right)\right]. \tag{I.11}$$

We can then use the projected field and gauge symmetry to couple the original 1-form to charged matter on the brane. A similar procedure

can be followed for the D-bein field E_M^A, necessary to couple fermions to gravity in a manifestly covariant way. Here and in what follows we indicate with $A, B, \ldots$, $A = 0, \ldots D-1$, and with $a, b \ldots$, $a = 0, n$ respectively the bulk and brane Lorentz indices. The D-bein E_M^A field represents D 1-forms ($A = 0, \ldots, D-1$) in the cotangent space, satisfying the relation $\eta_{AB} E_M^A E_N^B = G_{MN}$. The D-bein defines at each space-time point a tangent space basis corresponding to the coordinates of a free falling observer: it associates to the entries of a vector V^M in a given system of coordinates the entries $\hat{V}^A = E_M^A V^M$ in the coordinates of a free falling observer. Indeed, by the definition of E_M^A, vector products are conserved: $\hat{V}^A \hat{W}^B \eta_{AB} = V^M W^N G_{MN}$. Moreover local Lorentz rotations $E_M^A(X) \to R_B^A(X) E_M^B(X)$ are a gauge symmetry: the orientation of the D-bein at each point is not physical. This just means that the locally inertial reference frame is only defined up to a Lorentz transformation. Now, the tangent space σ to the brane at a point x is a $n+1$ subspace of the tangent space Σ at $X(x)$. A vector $v^\mu \in \sigma$ is written in free falling coordinates as $\hat{v}^A = E_M^A \partial_\mu X^M v^\mu$. By this relation, σ is represented as a $n+1$ dimensional subspace of the Lorentzian (free falling) vector space. To define the induced $n+1$-bein we have just to find an orthonormal basis of this subspace. One way to proceed is to divide the indices $\{A\}$ into two groups: $\{a\}$ for $A = 0, \ldots, n$ and $\{i\}$ for $A = n+1, \ldots, D-1$. Since σ is a time-like subspace we can always perform a Lorentz rotation $\hat{v}'^A = \bar{R}_B^A \hat{v}^B$ such that $\hat{v}'^i \equiv 0$ for $\hat{v}^A \in \sigma$. In the new basis, σ is spanned by $\hat{v}'^a$ for $a = 0, \ldots, n$, so that we have $\hat{v}'^a \hat{v}'^b \eta_{ab} \equiv \hat{v}'^A \hat{v}'^B \eta_{AB} = v^\mu v^\nu g_{\mu\nu}$. For vectors in σ, summing over a is equivalent to summing over A. The induced $n+1$-bein can then be defined as $e_\mu^a \equiv \bar{R}_B^a E_M^B \partial_\mu X^M$. It is straightforward to check that e_μ^a satisfies the basic conditions

$$e_\mu^a e_\nu^b \eta_{ab} = \hat{g}_{\mu\nu} \qquad e_\mu^a e_\nu^b \hat{g}^{\mu\nu} = \eta^{ab}. \tag{I.12}$$

One crucial remark is that the rotation $\bar{R}_B^A$ is only defined modulo the rotations in the subgroup $SO(1,n) \times SO(D-n-1)$ which leave the σ subspace and its complement invariant. In particular the induced e_μ^a is defined modulo the local Lorentz symmetry $SO(1,n)$ of the brane: the arbitrariness of our construction of e_μ^a does not affect physics provided the brane Lagrangian is written in a locally Lorentz invariant way. Using e_μ^a we can derive a brane spin connection and write a covariant Lagrangian for localized fermions.

The effective Lagrangian for a brane can then be written as

$$\mathcal{S}_{brane} = \int d^{n+1}x\sqrt{g}\Big\{ \quad -\tau + M^{n-1}\mathcal{R}(\hat{g}) + \bar{\psi}\ \not{D}\psi + D_\mu\phi^\dagger D^\mu\phi + \\ +(\text{higher derivatives})\Big\} \qquad \text{(I.13)}$$

where we have considered the example of localized fermion and scalar fields. The covariant derivatives involve the projected 1-forms, $\hat{A}_\mu$ and e^a_μ, as well as any possible localized Yang Mills field. The most relevant term, the one with the lowest number of derivatives, corresponds to the brane tension τ. In the effective theory description of the domain wall of the previous section, τ is determined by matching eq. (I.13) computed on a flat configuration, on which the intrinsic curvature terms vanish, with our calculation of the flat wall tension. This way we obtain $\tau = 4\sqrt{g}v^3/3$. The general result in eq. (I.13) also shows that the naive derivation of the effective action sketched in sect. I.1 is indeed accurate in the limit where the wall intrinsic curvature is small.

Let us consider as an explicit example a 3-brane living in D dimensional Minkowsky space, in the limit in which gravity is turned off. We can choose a gauge where the brane embedding is simply $X^\mu = x^\mu$ for $\mu = 0, 1, 2, 3$ and $X^i = Y^i(x)$ for $i = 4, \ldots, D-1$. In this parametrization the brane roughly extends along the $0, \ldots, 3$ direction of the bulk space. The functions Y^i parametrize the deformations along the orthogonal directions and are the dynamical degrees of freedom, the branons. In terms of the branons the induced metric is

$$\hat{g}_{\mu\nu} = \eta_{\mu\nu} + \partial_\mu Y^i \partial_\nu Y^j \delta_{ij}. \qquad \text{(I.14)}$$

and the tension term of eq. (I.13) expanded in powers of ∂Y becomes

$$L_{eff} = -\tau\sqrt{-g} = -\tau\left\{1 + \frac{1}{2}\partial_\mu Y^i \partial^\mu Y_i + \frac{1}{8}(\partial_\mu Y^i \partial^\mu Y_i)^2 + \ldots\right\}. \qquad \text{(I.15)}$$

This lagrangian provides a kinetic term with the right sign for Y provided $\tau > 0$. The configuration $Y^i = 0$ is stable and represents the vacuum configuration of our brane. But the most remarkable thing is that our general symmetry considerations also fix all the interaction terms involving n fields Y and a number of derivatives $\leq n$. The terms involving the curvature affect the interactions that have always at least two more derivative, and give subleading contributions to the scattering of branons at low enough energy. This is totally analogous to what happens in ordinary sigma models, like the pion lagrangian of QCD, where at lowest order in the expansion E/f_π the scattering amplitudes are fully

fixed by the group structure in terms of just one physical parameter, f_π itself. Indeed also the branon system is a σ-model whose coset space corresponds to the breakdown of the D-dimensional Poincaré group down to the 4-dimensional one [7, 8]. To conclude, notice that by indicating $\tau = f^4$ and by going to the canonical field $\hat{Y}^i = f^2 Y^i$, the branon interactions are proportional to inverse powers of f: this mass scale plays a role analogous to f_π in the pion Lagrangian.

I.3 Effective Field Theories

One general aspect of physical systems is that the dynamics at large length scales, or equivalently at low energy, does not depend too much on the microscopic details. For instance the interaction of an electromagnetic wave with an antenna of size a much smaller than the wavelength λ is described to a good accuracy by the coupling to the dipole mode of the antenna. Higher multipole moments will contribute to corrections suppressed by powers of $\lambda a \ll 1$. Another example is provided by molecules, where the *slow* vibrational modes, describing the oscillations in the distance between the various nuclei, can be accurately studied by first averaging over the *fast* motions of electrons. Averaging over the electronic states provides an effective Hamiltonian for the low frequency modes, where the higher details of the electronic structure are controlled by higher powers of the ratio $\omega_{slow}/\omega_{fast}$. Effective Field Theories technique provide a systematic way, an expansion, to treat the details of microscopic physics when discussing phenomena at low enough energy. Normally, when Quantum Field Theory is introduced as a construction to describe fundamental processes, a great emphasis is put on the requirement of renormalizability. Technically renormalizability corresponds to the possibility of sending the energy cut-off Λ of the system to infinity while keeping all the physical quantities finite (and non trivial)[1]. Physically this means that the theory can be extrapolated to infinitely small distances without encountering new microscopic structures. Renormalizable theories can be truly *fundamental* and not just an *effective* description valid in a limited energy range. Renormalizable theories are however a special case, and in practically all applications to particle physics one deals with non-renormalizable effective field theories [2] The best example of a non-renormalizable theory is given by General Relativity, which necessarily requires a new description at an energy

[1] As we will explain better below this is a somewhat stronger requirement than renormalizability: in weakly coupled theories it corresponds to asymptotic freedom.
[2] For excellent introductions to effective field theories see the papers in Ref.[9]

scale smaller or equal to $M_P = 10^{19}$ GeV. Nonetheless GR makes perfect sense as an effective field theory at energies much smaller than its cut off [10]. But also QED is an effective field theory: the interactions of electrons and photons are modified at energies much bigger than m_e by the presence of new particles and new interactions. Still in the regime $E \sim m_e$ the *small* effects of the microscopic dynamics can be accounted for by adding a suitable tower of non-renormalizable interactions. The SM, the most fundamental description of particle interactions gravity excluded, is renormalizable. Still we can only consider the SM as an effective theory. On one hand this is because the necessary inclusion of gravity makes it non-renormalizable. On the other hand, even in the absence of gravity, the SM is renormalizable but not asymptotically free. At least one of its couplings, the one associated to the hypercharge vector boson, grows logarithmically with energy and becomes infinite at a scale $M_L \sim m_W e^{b/\alpha}$, with $b = 12\pi \cos^2\theta_W/41 \sim 1$. At $E \sim M_L$ the perturbative description breaks down, very much like the effective description of G.R. breaks down at the Planck scale. The fact that $M_L \gg M_P$ makes however this second problem academic.

In order to make these general statements more concrete, let us focus on a very simple example. Let us consider a physical system which at low enough energy possesses just one scalar degree of freedom parametrized by a field ϕ. The most general local and Poincaré invariant Lagrangian can be written as an expansion in powers of ϕ and of its derivatives

$$\begin{aligned} \mathcal{L} &= \partial_\mu\phi\partial^\mu\phi - m^2\phi^2 + \lambda_4\phi^4 + \frac{\lambda_6}{M^2}\phi^6 + \frac{\lambda_8}{M^4}\phi^8 + \cdots \\ &+ \frac{\eta_4}{M^2}\phi^2\partial_\mu\phi\partial^\mu\phi + \frac{\eta_6}{M^4}\phi^4\partial_\mu\phi\partial^\mu\phi\cdots \end{aligned} \tag{I.16}$$

where for simplicity we have also assumed a symmetry $\phi \to -\phi$. We have scaled all the couplings by powers of one mass scale M and by dimensionless quantities $\lambda_i, \eta_i, \ldots$. It is reasonable to assume that $\lambda_i, \eta_i, \ldots \sim O(1)$. This corresponds to a theory that in addition to the particle mass m contains only another physical scale, M, associated to the interactions. It is easy to understand the meaning of this expansion when calculating scattering amplitudes at energies $m \ll E \ll M$. Let us focus on tree level computations first. We shall worry about quantum corrections later. Neglecting numerical factors and indicating one power of momentum generically by E, we have

$$\mathcal{A}_{2\to2}(E) \sim \lambda_4 + \eta_4\frac{E^2}{M^2}\ldots \tag{I.17}$$

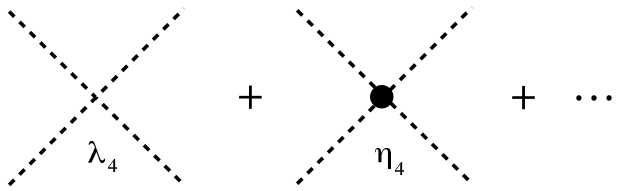

Figure 3. The diagrams contributing to the elastic process $2 \to 2$ at lowest order.

for the elastic process $2 \to 2$ corresponding to Fig. (3) and

$$\mathcal{A}_{2\to 4}(E) \sim \frac{1}{E^2}\left\{\lambda_4^2 + \lambda_4\eta_4\frac{E^2}{M^2} + \lambda_6\frac{E^2}{M^2} + \dots\right\} \qquad \text{(I.18)}$$

for the inelastic process $2 \to 4$ shown in Fig. (4). This power counting corresponds to simple dimensional analysis. Notice that for $E \ll M$ the dimensionless coupling λ_4 dominates all the amplitudes. This result is intuitively obvious. A coupling g of mass dimension $[E]^d$ can perturbatively contribute to observables via the dimensionless combination g/E^d. We can then distinguish three classes of couplings depending on whether d is positive, zero or negative. Couplings of positive dimension are called relevant, as their effect becomes more important the smaller the energy. An example is given by the mass term itself, which gives small $O(m^2/E^2)$ effects in the relativistic regime, but becomes important when $E \sim O(m)$. Couplings of vanishing dimensions, like λ_4, are termed marginal. At tree level their effects are independent of the energy scale. Finally, couplings of negative dimension are termed irrelevant, as their effects become very small in the low energy domain. Notice that while there is only a finite numer of relevant and marginal couplings, the tower of irrelevant couplings is infinite. In spite of their infinity, and as their naming suggests, irrelevant couplings do not totally eliminate the predictive power of our Lagrangian as long as we use it at low energy, $E \ll M$. At each finite order $(E/M)^n$, only a finite number of terms in the Lagrangian contributes to the amplitudes. This preserves a weak form of predictivity, which is often good enough, since we just need to match our theoretical computations to the experimental precision, which is always finite.

We can now worry about quantum corrections. These introduce some technical difficulties, but the basic conclusion is unchanged. To be fully general, let us write our Lagrangian as a sum over operators $\mathcal{O}_i$ of di-

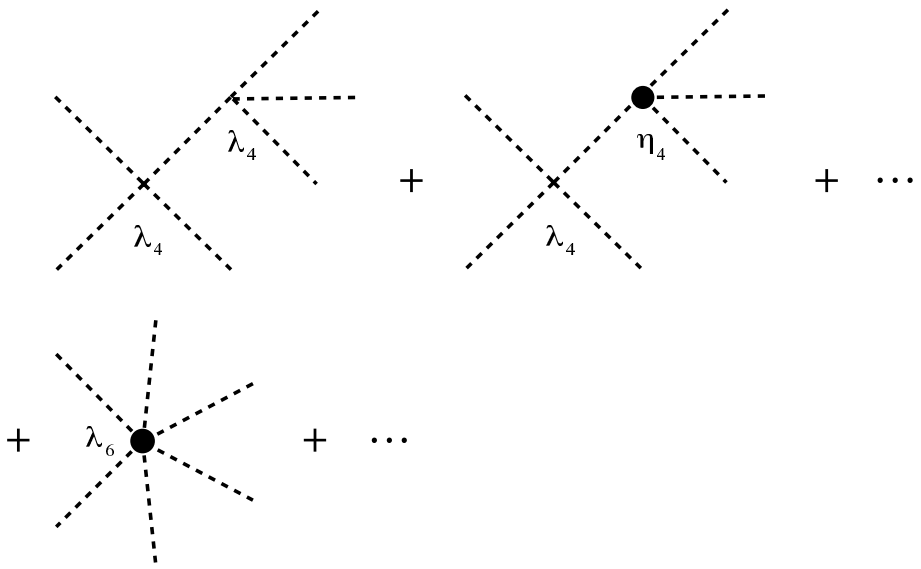

Figure 4. The leading contributions to the inelastic process $2 \to 4$.

mension $d_i + 4$

$$\mathcal{L} = \sum_i c_i \frac{\mathcal{O}_i}{M^{d_i}}. \tag{I.19}$$

Assume we want to calculate some observable at order $(E/M)^n$. Working at tree level it is enough to truncate $\mathcal{L}$ to the operators with $d_i \leq n$. The analysis at tree level is made simple by the fact that the external momenta ($\sim E$) completely fix the momenta of the internal lines and vertices. This is no longer true at loop level, where the loop momentum can be arbitrarily high. Moreover some of the loop integrals are UV divergent and must be cut-off at some scale Λ. An interaction term $c\mathcal{O}_N/M^{d_N}$, can generate quantum corrections that involve positive powers of the cut-off Λ

$$\frac{\delta\mathcal{A}}{\mathcal{A}} \sim \cdots + c\frac{\Lambda^P E^{d_N - P}}{M^{d_N}} + \cdots \tag{I.20}$$

Then an operator with $d_N > n$, which at tree level only gives corrections beyond the needed accuracy E^n, can, at loop level, generate effects that scale with a power $d_N - P \leq n$. Moreover if $\Lambda \sim O(M)$ these quantum effects are as important as the tree level contribution of operators of lower dimension. This seems very embarrassing. Fortunately it can

be proven that these effects are exactly equivalent to a renormalization of the coefficients c_i of the operators of lower dimensionality. Therefore they do not contain any new information and can be eliminated by a trivial change of renormalization scheme. Their equivalence to local operators is qualitatively understandable: loops of high virtuality are small in position space, corresponding to a region of size $1/\Lambda$, and look point-like with respect to the long wavelength $1/E$ of the external particles. Another way to understand this result is to take a Wilsonian view point where Λ is the running cut-off. After running down to a scale Λ such that $E \lesssim \Lambda \ll M$, the troublesome virtual effects becomes manifestly small: the big effect has been replaced by a local renormalization of the classical Lagrangian [3]. But there is no doubt that the most convenient method to define effective field theories at the quantum level is by Dimensional Regularization (DR). Dimensionally regulated loop integrals exhibit no powerlike divergences, only logarithmic divergences survive. The issue we just worried about does not even arise! The naive power counting we found at tree level carries over to the quantum theory up to mild logarithimic corrections.

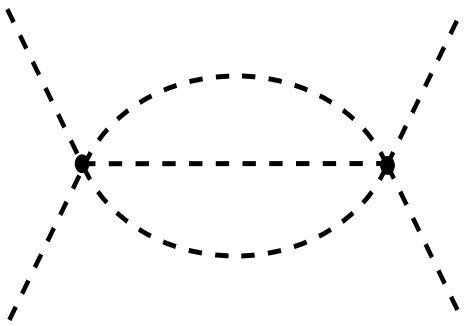

Figure 5. 2-loop contribution to elastic scattering from $\lambda_5\phi^5/M$ vertices.

Consider, for instance, the 2-loop diagram involving two insertions of the $\lambda_5\phi^5/M$ interaction shown in Fig. (5) By using a hard momentum cut-off we have

$$\delta\mathcal{A}_{2\to 2} = \lambda_5^2 \left\{ a\frac{\Lambda^2}{M^2} + b\frac{E^2}{M^2}\ln \Lambda/E \right\} \tag{I.21}$$

[3]To make this argument fully rigorous one should take into account that the Wilsonian Lagrangian at scale Λ now also contains terms that scale like inverse powers of Λ. Terms proportional to $1/(\Lambda^R M^S)$ are however fixed by the renormalized couplings associated to operators with $d_i \leq S$ [11].

while DR in $4+\epsilon$ dimensions gives

$$\delta\mathcal{A}_{2\to 2} = \lambda_5^2 \left(\frac{1}{\epsilon} + b \ln \mu/E\right) \frac{E^2}{M^2}. \tag{I.22}$$

In DR, after renormalization, this diagram gives just a logarithimic Renormalization Group (RG) evolution of the coefficient of a dimension 6 operator $(\phi^2)\Box(\phi^2)$. We emphasize that while the power divergences are totally scheme dependent, being fully saturated in the UV, the logarithmic divergence involves a physical IR singularity $\ln E$ and must be the same in both regularizations. This $\ln E$ term is associated by unitarity to the cut diagrams $|\mathcal{A}_{2\to 3}|^2$.

A by-product of this discussion is that in DR with minimal subtraction (or any other mass independent subraction scheme) the RG equation for the couplings of an effective Lagrangian follows just by dimensional analysis [12]. Using the notation of eq. (I.19) where a coupling c_i/M^{d_i} has dimension $-d_i$, the β function has the form

$$\mu\frac{dc_i}{d\mu} = \sum_{d_j+d_k=d_i} a_{j,k}c_jc_k + \sum_{d_j+d_k+d_l=d_i} a_{j,k,l}c_jc_kc_l + \cdots \tag{I.23}$$

where $a_{i,j}$, $a_{i,k,l}$, ... are numerical coefficient following from the loop integrals. Notice that the parameter submanifold where all the irrelevant couplings ($-d_i < 0$) vanish is stable under RG evolution. This interesting submanifold corresponds to what we normally call renormalizable theories. On the other hand, once we turn on an irrelevant coupling of dimension $-d < 0$ it will generate by RG evolution an infinite subset of the couplings of more negative dimension. Such theories are termed non-renormalizable, as quantum effects force the presence of infinitely many inputs, though we hope to have made it clear how to deal with them. Notice also that our original assumption to scale all the irrelevant couplings by the same mass M is stable under RG flow. Of course there can be more complicated situations and models where the higher dimensional couplings involve hierarchically different scales.

We conclude this discussion by reiterating the basic theorem. Lagrangians involving all possible non-renormalizable terms can be made sense of as effective ones. A weak form of predictivity can be preserved by working in perturbation theory in an expansion in E/M, where M is the lowest scale characterizing the non-renormalizable couplings. This works as long as $E \ll M$. When $E \sim M$ infinitely many parameters become relevant and our effective Lagrangian completely looses predictive power. A reasonable expectation is that at the scale M the theory enters a new regime where perhaps new degrees of freedom are relevant. For instance this is what happens in QCD at the scale $4\pi f_\pi \sim 1GeV$ where the

weakly coupled description of mesonic physics breaks down. At this energy the hadrons deconfine and at higher energies the dynamics is more accurately described in terms of quarks and gluons. Notice that, according to our discussion, non-asymptotically free renormalizable theories are qualitatively similar to non-renormalizable theories. In the former case, at least one coupling grows logarithmically $\lambda(E) \propto 1/\ln(M_L/E)$ with energy, while in the latter the growth of the effective dimensionless couplings is powerlike $\lambda(E) \propto (E/M)^n$. In both cases the perturbative description breaks down at some high scale, M_L or M. The only difference between the two cases is quantitative: in the renormalizable case, for a not too small a value of the running coupling at low energy, the cut-off scale M_L is exponentially far away.

I.4 Examples

We can analyze from the effective field theory viewpoint some system of interest. One instructive example is provided by pure gravity, whose Lagrangian was given in eq. (I.5). To study the interactions let us focus on the case of vanishing cosmological constant and let us expand the metric field around the flat background

$$g_{MN} = \eta_{MN} + \frac{h_{MN}}{M_D^{\frac{D}{2}-1}}. \tag{I.24}$$

Eq. (I.5) will then be written as a power series in the fluctuation h

$$\left\{(\partial h)^2 + \frac{1}{M_D^{\frac{D}{2}-1}} h(\partial h)^2 + \frac{1}{M_D^{D-2}} h^2(\partial h)^2 + \dots\right\} \cdots + b\left\{\frac{1}{M_D^{\frac{D}{2}+1}}(\partial^2 h)(\partial^2 h) + \dots\right\} + \dots \tag{I.25}$$

where we have been very schematic, suppressing all the tensor structure indices and $O(1)$ factors, but keeping the derivative expansion structure manifest. Notice that the fluctuation h has been defined in such a way that it is canonically normalized. The interactions then all scale by inverse powers of the Planck mass. The above Lagrangian, after suitable gauge fixing can be used to compute graviton scattering processes in perturbation theory. For instance the amplitude $\mathcal{A}_{2\to 2}$ scales with the energy E like

$$\mathcal{A}_{2\to 2} \sim \frac{E^{D-2}}{M_D^{D-2}}\left(1 + b\frac{E^2}{M_D^2} + \dots\right). \tag{I.26}$$

The dots also include quantum effects, which scale like positive powers of E/M_D. For instance, by simple dimensional analysis, 1-loop effects induced by the leading two derivative Lagrangian are of order $(E/M_D)^{D-2}$ with respect to the leading tree level contribution. In the ordinary purely 4 dimensional theory of Einstein gravity the Planck mass $M_4 \sim 10^{18}$ GeV is much bigger than any energy scale relevant to astrophysics or cosmology (if not for very early cosmology, even before inflation). Then the truncation of the theory to the lowest two derivative Lagrangian, the Einstein-Hilbert action, already allows a very accurate description of the dynamics.

As a second example consider the interactions of the branon excitations Y of a 3-brane

$$\mathcal{L}_{brane} = -f^4\sqrt{\hat{g}} + bf^2\sqrt{\hat{g}}\mathcal{R}(\hat{g}) + \dots \tag{I.27}$$

Substituting eq. (I.14) and writing the interactions in terms of the canonical fields $\hat{Y}_i = f^2 Y_i$ it is straighforward to power count the scaling of Feynman diagrams. For the $YY \to YY$ amplitude the Feynman diagram expansion corresponds to the series

$$\mathcal{A}_{2\to 2} = \frac{E^4}{f^4}\left(1 + C\frac{E^4}{f^4}\ln E + b\frac{E^2}{f^2} + \dots\right). \tag{I.28}$$

where the first and second term (proportional to a calculable coefficient C) are determined by the quartic interaction in eq. (I.15) respectively at tree level and at 1-loop. The tension f turns out to be the energy scale which controls the perturbative expansion. At energies $E \sim f$ the effective field theory description surely breaks down, in analogy with the case $E \sim M$ in the scalar toy model of the previous section. The quantity $(E/f)^2$ controls the strength of the interaction like $\alpha/4\pi$ does in quantum electrodynamics.

The length $L = 1/f$ can be interpreted as the quantum size of the brane, in analogy with the Compton wavelength of a particle. Indeed in the case of a 0-brane, a point particle, f coincides with the mass m and we recover the usual definition of Compton wavelength. In the case of a particle the length $1/m$ controls the domain of validity of the low energy non-relativistic effective theory. If we try to localize one electron at a distance $< 1/m$, then, by the indetermination principle, not only will its momentum p be relativistic but the production of electron-positron pairs energetically possible. In the case of the brane, we can, for instance, consider the quantum fluctuation of the linearized induced metric on the vacuum . We find (cf. eq. (I.14))

$$\langle \partial_\mu Y^i \partial^\mu Y_i \rangle \sim \frac{1}{f^4}\int k^3 dk \tag{I.29}$$

which shows that at wavelengths of order $1/f$ the fluctuation of the brane position becomes itself of order $1/f$: at these short wavelengths the brane cannot be approximated by a smooth surface. Basically it is not possible to talk about fluctuations in the position of the brane that are shorter than $1/f$ in both longitudinal and transverse directions.

With the previous considerations in mind, it is instructive to consider the field theoretic domain wall discussed in sect. I.1. For definiteness let us focus on the case of a 4-dimensional scalar theory, so that the wall is a 2-brane. The tension is $\tau = \sqrt{g}v^3$, while the cut-off of the effective description is provided by $m = \sqrt{g}v$, the energy at which extra massive modes come in. We have $m = (g\tau)^{1/3}$, so that as long as the original 4D theory was weakly coupled ($g \ll 1$), the brane theory never gets into a strong coupling regime. At the cut off scale m, the loop expansion parameter of the effective brane theory $m^3/\tau = g$ coincides with the loop expansion parameter of the original scalar field theory.

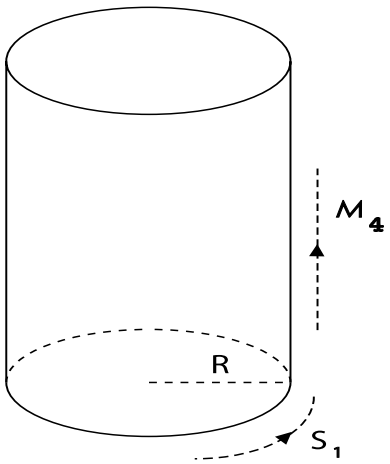

Figure 6. Cylindrical structure of 5-dimensional space-time compactified on $\mathcal{M}_4 \times S_1$.

I.5 Kaluza-Klein decomposition

So far we have been general: our discussion applies equally well to compact and to infinite extra-dimensions. However, since it is empirically very clear that we live in three macroscopic spatial dimensions, for phenomenological applications we must focus on the case in which the extra-dimensions are compactified at some small enough radius R. The dynamics at distances much bigger than R will not be able to notice the

presence of the extra compact directions. To illustrate this fact let us consider the simplest situation of a 5D scalar field ϕ with the 5th dimension compactified on a circle (see Fig. (6)) of radius R. Compactification is formally expressed by the periodicity requirement

$$\phi(x, x_5) = \phi(x, x_5 + 2\pi R) \tag{I.30}$$

Processes taking place on time scales $T \ll R$, by causality and by locality, cannot notice that the 5th dimension is compact. On the other hand to study processes happening on a time scale $T \gtrsim R$, and in particular at energies $E \lesssim 1/R$, the 5D local description is not the most adequate. In this case it is convenient to expand the field ϕ in its Fourier components with respect to the periodic coordinate x_5.

$$\phi(x, x_5) = \sum_{n=-\infty}^{n=\infty} \phi_n(x) e^{i\frac{nx_5}{R}}. \tag{I.31}$$

where the reality of ϕ implies $\phi_{-n}(x) = \phi_n(x)^*$. Notice that each different coefficient ϕ_n in this expansion corresponds to a different 4D field. The ϕ_n are called Kaluza-Klein (KK) fields. According to this expansion the 5D kinetic action integrated over x_5 becomes

$$\begin{aligned} \int \mathcal{L}_\phi dx^5 = \quad & -\tfrac{1}{2} \quad \int \left[(\partial_\mu \phi)^2 - (\partial_5 \phi)^2\right] = \\ & \tfrac{1}{2} \quad \sum_{-\infty}^{\infty} \left[-|\partial_\mu \phi_n|^2 + \frac{n^2}{R^2}|\phi_n|^2\right]. \end{aligned} \tag{I.32}$$

The original 5D massless field has been decomposed in a tower of Kaluza-Klein scalars ϕ_n with mass

$$m_n = n/R. \tag{I.33}$$

If we work at energy E, only a limited number $n \sim ER$ of KK can be produced. In particular, for $E < 1/R$ only the zero mode ϕ_0 is available. At such low energies the model looks 4-dimensional. The KK particles appear only virtually, and their effect is reproduced by a suitable set of local operators involving only the massless 4D fields. In the specific example we are considering, the full space-time symmetry is just the 4-dimensional Poincaré group times translations along the fifth direction: $P_4 \times U(1)$. The KK particle states represent just the irreducible representations of this group. In particular the index n represents the charge under the $U(1)$ group of 5D translations: 5D translational invariance shows up in 4D as the conservation of the KK indices n_i summed over the incoming and outgoing particles in a collision.

Along similar lines one can study the KK decomposition of a gauge vector field A_M. But rather than discussing it in detail we go directly to the case of the graviton: the technical issues, associated to gauge invariance, are analogous for both vector and tensor field. So let us consider the original theory of Kaluza and Klein [13]: 5D Einstein gravity compactified on $\mathcal{M}_4 \times S_1$ with the action

$$2M_5^3 \int_{\mathcal{M}_4 \times S_1} \sqrt{g}\mathcal{R}(g) \tag{I.34}$$

We can write the full metric tensor in block form

$$g_{MN}(x, x^5) = \begin{pmatrix} g_{\mu\nu} & g_{\mu 5} \\ g_{5\mu} & g_{55} \end{pmatrix} = \begin{pmatrix} \eta_{\mu\nu} + h_{\mu\nu} & h_{\mu 5} \\ h_{\mu 5} & 1 + h_{55} \end{pmatrix}. \tag{I.35}$$

To work out the spectrum we must compute the quadratic action in the linearized field h_{MN} and then use the gauge freedom provided by the linearized 5D diffeomorphisms, $x^M \to x^M + \epsilon^M(x, x^5)$

$$h_{MN} \to h_{MN} + \delta h_{MN} = h_{MN} + \partial_N \epsilon_M + \partial_M \epsilon_N. \tag{I.36}$$

to eliminate the redundant degrees of freedom. Here and in what follows, working at linear order, indices are raised and lowered using the Lorentz metric η_{MN}. We stress that the compactification of the fifth dimension implies that all our fields, including ϵ_M are periodic in x^5. Using the 5 gauge parameters ϵ_N we can essentially eliminate 5 combinations of the metric fluctuations h_{MN}. We can choose these 5 combinations to be just h_{55} and $h_{\mu 5}$. By using Fourier modes we have that $\delta h_{55} = 2\partial_5 \epsilon_5$ becomes

$$\delta h_{55}^{(n)} = 2in\epsilon_5^{(n)} \tag{I.37}$$

which explicitly shows that we can eliminate all the modes but $h_{55}^{(0)}$, which is gauge invariant. The gauge invariance of zero modes like $h_{55}^{(0)}$ follows from the periodicity of the gauge transformation

$$\oint_{S_1} \delta h_{55} = 2 \oint_{S_1} \partial_5 \epsilon_5 = 0 \tag{I.38}$$

and is a generic features of gauge theories on compact spaces. The same thing happens for $h_{\mu 5}$

$$\delta h_{\mu 5}^{(n)} = \partial_\mu \epsilon_5^{(n)} + in\epsilon_\mu^{(n)}. \tag{I.39}$$

Therefore by using the $n \neq 0$ modes of ϵ_M we can go to a gauge where

$$h_{55}(x, x_5) \equiv \phi(x) \qquad h_{\mu 5}(x, x_5) \equiv A_\mu(x) \tag{I.40}$$

while $h_{\mu\nu}(x, x_5)$ is unconstrained. However we still have the zero mode gauge freedom

$$A_\mu \to A_\mu + \partial_\mu \epsilon^{(0)} \qquad h^{(0)}_{\mu\nu} \to h^{(0)}_{\mu\nu} - \partial_\mu \epsilon^{(0)}_\nu - \partial_\nu \epsilon^{(0)}_\mu. \tag{I.41}$$

The residual scalar mode ϕ, usually called radion, is associated to fluctutations in the proper length L of the radius of compactification $\delta L = \oint h_{55}/2 = \pi R \phi$. The graviphoton A_μ, as shown in eq. (I.41), is the gauge field associated to 4D local translations of the 5th coordinate. The associated charge is just the momentum along the fifth dimension, i.e. the KK index n.

Defining the KK modes via

$$h_{\mu\nu} \equiv \sum_{n=-\infty}^{+\infty} h^{(n)}_{\mu\nu} e^{i\frac{n x_5}{R}} \tag{I.42}$$

the linearized 4D action becomes

$$\begin{aligned} \mathcal{L}^{(2)}_{4D} = M_5^3 \pi R \Bigg\{ \Big[& \sum_{-\infty}^{+\infty} h^{(n)\mu\nu} \Box h^{(-n)}_{\mu\nu} - h^{(n)\mu}_{\ \ \mu} \Box h^{(-n)\nu}_{\ \ \nu} \\ & + 2 h^{(n)}_{\mu\nu} \partial^\mu \partial^\nu h^{(-n)\rho}_{\ \ \ \rho} - 2 h^{(n)}_{\mu\nu} \partial^\mu \partial^\rho h^{(-n)\nu}_{\ \ \ \rho} \\ & + \frac{n^2}{4R^2} (h^{(n)\mu}_{\ \ \ \mu} h^{(-n)\nu}_{\ \ \ \nu} - h^{(n)\mu\nu} h^{(-n)}_{\mu\nu}) \Big] \\ & + 2\phi(\partial^\mu \partial^\nu h^{(0)}_{\mu\nu} - \Box h^{(0)\mu}_{\ \ \ \mu}) - F_{\mu\nu} F^{\mu\nu} \Bigg\} \end{aligned} \tag{I.43}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. By looking at the coefficient of the zero mode action we deduce that the effective 4-dimensional Planck scale is

$$M_4^2 \equiv M_5^3 2\pi R \tag{I.44}$$

The ∂_5 terms of the 5d Lagrangian have turned into mass terms for the $n \neq 0$ modes. As first noticed by Fierz and Pauli [14], the specific tensor structure of this mass term is the only one ensuring the absence of ghosts and tachyons in $h^{(n)}_{\mu\nu}$. The equations of motion for the massive modes reduce indeed to

$$\left(\Box + \frac{n^2}{R^2} \right) h^{(n)}_{\mu\nu} = 0 \qquad \partial^\mu h^{(n)}_{\mu\nu} = h^{(n)\mu}_{\ \ \ \mu} = 0 \tag{I.45}$$

where the second and third relations follow by taking the divergence and trace of the equation of motion. This is completely analogous to the well known case of a massive vector V_μ. There the divergence of the equation of motion gives the constraint $\partial^\mu V_\mu = 0$, implying that

only 3 out of the 4 degrees of freedom propagate, as it should be for a $J = 1$ massive particle. Here, due to the constraints, we have $10 - 5 = 5$ propagating states, corresponding to a massive $J = 2$ particle. An arbitrary symmetric two index tensor $H_{\mu\nu}$ can be decomposed in components of definite spin as

$$H_{\mu\nu} = H^{TT}_{\mu\nu} + \partial_\mu E^T_\nu + \partial_\nu E^T_\mu + \left(\eta_{\mu\nu} - \frac{\partial_\mu\partial_\nu}{\partial^2}\right)\Phi + \frac{\partial_\mu\partial_\nu}{\partial^2}\Psi \equiv 2\oplus 1\oplus 0_\Phi \oplus 0_\Psi \tag{I.46}$$

where E^T_μ and $H^{TT}_{\mu\nu}$ are respectively transverse and transverse-traceless ($\partial^\mu E^T_\mu = \partial^\mu H^{TT}_{\mu\nu} = \eta^{\mu\nu} H^{TT}_{\mu\nu} = 0$) and the spin of each component is indicated in an obvious notation. By eq. (I.45) only H^{TT} survives on-shell. An instructive exercise is to construct the projectors on H^{TT}, E^T, Φ and Ψ by writing them in a compact way in terms of the transverse and longitudinal vector projectors $\Pi^T_{\mu\nu} = \eta_{\mu\nu} - \partial_\mu\partial_\nu/\partial^2$, $\Pi^L_{\mu\nu} = \partial_\mu\partial_\nu/\partial^2$. Another instructive exercise is to write the kinetic Lagrangian in terms of the various projectors, in complete analogy with the massive $J = 1$ case, and from that to derive the massive $J = 2$ propagator

$$\langle h^{(-n)}_{\mu\nu} h^{(n)}_{\rho\sigma}\rangle = \frac{\frac{1}{2}\left(\hat\Pi^T_{\mu\rho}\hat\Pi^T_{\nu\sigma} + \hat\Pi^T_{\mu\sigma}\hat\Pi^T_{\nu\rho}\right) - \frac{1}{3}\hat\Pi^T_{\mu\nu}\hat\Pi^T_{\rho\sigma}}{p^2 + \frac{n^2}{R^2}} \equiv \frac{\Pi_{\mu\nu\rho\sigma}(m_n)}{p^2 + m_n^2} \tag{I.47}$$

where $\hat\Pi^T_{\mu\nu} = \eta_{\mu\nu} - p_\mu p_\nu/(m_n)^2$.

Let us now focus on the zero modes. Notice that the radion mixes kinetically to the graviton. It is convenient to diagonalize the kinetic term via the Weyl shift $h^{(0)}_{\mu\nu} \equiv \bar h_{\mu\nu} - \frac{1}{2}\phi\eta_{\mu\nu}$, after which ϕ acquires a self kinetic term

$$\frac{3}{2} M_5^3 \pi R \phi \Box \phi \tag{I.48}$$

while $\bar h_{\mu\nu}$ has obviously the kinetic term of massless graviton. At this point we can gauge fix the residual 4D reparametrization and gauge symmetry by using respectively the de Donder and Feynman gauges

$$2M_5^3\pi R\Big\{\left(\partial^\mu \bar h_{\mu\nu} - \frac{1}{2}\partial_\nu \bar h^\mu_\mu\right)^2 - (\partial^\mu A_\mu)^2\Big\}. \tag{I.49}$$

On shell we have 2 physical helicity states in both $\bar h$ and A. These, including ϕ, add up to 5 states: the same number we found at each excited level, but here they are shared among particles of different spin.

The presence of the radion ϕ makes this theory quite different from ordinary 4D Einstein gravity (the additional scalar is sometimes called a Brans-Dicke field). The tensor field that couples to ordinary 4D matter

and thus describes the observable fluctuations of the 4D geometry is the original metric $h^{(0)}_{\mu\nu}$ and not $\bar{h}_{\mu\nu}$. Therefore the relevant graviton propagator is

$$\begin{aligned}\langle h^{(0)}_{\mu\nu} h^{(0)}_{\rho\sigma}\rangle &= \langle \bar{h}_{\mu\nu}\bar{h}_{\rho\sigma}\rangle + \frac{1}{4}\langle\phi\phi\rangle \\ &= \frac{1}{M_4^2}\{\frac{\frac{1}{2}\left(\eta_{\mu\rho}\eta_{\sigma\nu}+\eta_{\mu\sigma}\eta_{\rho\nu}\right)-\frac{1}{2}\eta_{\mu\nu}\eta_{\rho\sigma}}{q^2} \\ &+ \frac{1}{6}\frac{\eta_{\mu\nu}\eta_{\rho\sigma}}{q^2}\}\end{aligned} \tag{I.50}$$

where the first term is just the result we would get in ordinary GR and the second contribution, proportional to 1/6, is due to the radion. In the non-relativistic regime the effects of the tensor and scalar field are indistinguishable. The Newton constant is determined by $\langle h^{(0)}_{00} h^{(0)}_{00}\rangle$ and given by

$$32\pi G_N = \frac{2}{M_4^2}\left(\frac{1}{2}+\frac{1}{6}\right) = \frac{4}{3}\frac{1}{M_4^2} \tag{I.51}$$

where we have indicated separately the contributions of the tensor and of the scalar. However in the relativistic regime the implications of the two terms are quite different. In particular ϕ does not couple to photons as they have a traceless energy momentum tensor. Now, one of the most accurate tests of GR is the measurement of the deflection of light by the gravitational field of the Sun: the experimental result agrees with the theory to about 1 part in 10^3. In the theory at hand, ϕ does not contribute to this deflection, and the scattering angle, expressed in terms of the non-relativistic G_N of eq. (I.51), is therefore only 3/4 of the GR prediction. This result is completely ruled out by the data. In order to meet consistency, ϕ should be given a mass m_ϕ, so that its contribution to the potential will decay as $e^{-m_\phi r}/r$ and become quickly irrelevant for $r > 1/m_\phi$. Notice that giving a mass to ϕ corresponds to stabilizing the size of the 5th dimension. The agreement between the value of G_N measured in post Cavendish experiments [21–23] down to distances of order 100 μm with that governing post Newtonian corrections in the solar system forces the mass of ϕ to be bigger that $\hbar c/100\mu m \sim 10^{-3}$ eV.

To gain another viewpoint on the compactification of gravity from D down to 4 dimensions, it is worth to count the physical states of gravity around ordinary D-dimensional (non-compact) Minkowsky space. The metric symmetric tensor h_{MN} corresponds to $D(D-1)/2$ fields. By using the D gauge transformations ϵ_N we can eliminate D of these fields. For instance we can go to the Gaussian normal gauge where all time components vanish $h_{00} = h_{0i} = 0$. This choice is the analogue of Coulomb

gauge in Maxwell's theory. However as in any gauge theory, even after fixing the gauge, we must still impose the equations of motion of h_{00} and h_{0i} as an initial time constraint

$$\frac{\partial \mathcal{L}}{\delta h_{00}} = \nabla^i \nabla^j h_{ij} - \nabla^i \nabla_i h^j_j |_{t=0} = 0 \qquad \frac{\partial \mathcal{L}}{\delta h_{0i}} = \left(\nabla^j \dot{h}_{ij} - \nabla_i \dot{h}^j_j \right) |_{t=0} = 0 \tag{I.52}$$

where by the dot and by ∇_i we indicate respectively detivatives with respect to the time, t, and space, x^i, coordinates. The divergence of the h_{ij} equations of motion

$$\left(\nabla^j \ddot{h}_{ij} - \nabla_i \ddot{h}^j_j \right) = 0 \tag{I.53}$$

ensures that both constraints remain valid at all times[4]. To clarify things it is convenient to decompose h_{ij} in spin components as previously done in eq. (I.46)

$$h_{ij} \equiv H^{TT}_{ij} + \nabla_i V^T_j + \nabla_j V^T_i + \left(\delta_{ij} - \frac{\nabla_i \nabla_j}{\nabla^2} \right) H + \frac{\nabla_i \nabla_j}{\nabla^2} V \tag{I.54}$$

Moreover it should be noticed that there is a residual gauge freedom preserving the Gaussian normal condition

$$\epsilon_0 \equiv \epsilon_0(x_i) \qquad \epsilon_i \equiv \epsilon^T_i(x_i) - \nabla_i(\epsilon_L(x_i) + t\epsilon_0(x_i)). \tag{I.55}$$

Notice that the ϵ's depend only on the space coordinates and that $\nabla^i \epsilon^T_i = 0$. Now, the divergence and trace of the h_{ij} equation of motion imply

$$\nabla_i \ddot{H} = \nabla^2 \ddot{V}^T_i = 0 \qquad \ddot{V} + (D-2)\ddot{H} - (D-3)\nabla^2 H = 0. \tag{I.56}$$

Assuming that our fields F vanish fast enough at spactal infinity, $\nabla^2 F = 0$ implies $F = 0$. The first two equations then imply $H = H_0(x_i) + tH_1(x_i)$ and $V^T_j = V^T_{0j}(x_i) + tV^T_{1j}(x_i)$. The initial time constraints imply however $H_0 = H_1 = V^T_{1j} = 0$. In turn eq. (I.56) implies $V = V_0(x_i) + tV_1(x)$. At this point we are left with D functions V^T_{0i}, V_0, V_1 which can be completely eliminated by the residual gauge freedom $\epsilon^T_i, \epsilon_L, \epsilon_0$. Notice that the initial time constraints eliminate 1 dynamical variable, H, plus $D-1$ "velocities" $\dot{H}, \dot{V}^T_i$. Instead the gauge freedom allows to eliminate $D-1$ variables and 1 velocity. This generalizes the situation

[4]Again this is analogous to electromagnetism. In Coulomb gauge, the A_0 equation of motion constraint gives Gauss law $\vec{\nabla} \cdot \vec{E} = 0$ at initial times. Maxwell's equation $\dot{\vec{E}} = \vec{\nabla} \times \vec{B}$ implies the validity of Gauss's law at all times.

field	original d.o.f	(−) gauge fixing	(−) eqs. motion	propagating d.o.f.
$(J=2)\ h_{\mu\nu}$	10	0	-5	5
$(J=1)\ h_{\mu i}$	$4n$	-4	$-(n-1)$	$3(n-1)$
$(J=0)\ h_{ij}$	$\frac{n(n+1)}{2}$	$-n$	0	$\frac{n(n-1)}{2}$

Table 1. Number of degrees of freedom (d.o.f.) off-shell and on-shell for each field component

in electromagnetism where the Gauss constraint eliminates 1 velocity, $\nabla^i \dot{A}_i$, while the residual gauge freedom eliminates $\nabla^i A_i$.

The result of all this is that, after going to the $h_{00} = h_{0i} = 0$ gauge, an additional D deegrees of freedom are eliminated (V_i^T, H, V) and we are left with $D(D-1)/2 - 2D = D(D-3)/2$ propagating fields, corresponding to H_{ij}^{TT}.

Before concluding this section we discuss the more general case of Einstein gravity in $D = 4 + n$ dimensions, with the n extra-dimensions compactified on a square torus T^n [15]. Indicating by i and x^i the extra indices and coordinates, T^n is defined by the equivalence relation $x^i \sim x^i + 2\pi R n^i$ with n^i a vector with integer entries. KK levels are labelled by a vector of integers $(\overrightarrow{n})_i = n_i$ associated to the momentum n_i/R along the T^n directions. The counting of physical degrees of freedom for each massive KK level is shown in Table (1). It is easy to check that gauge invariance allows to eliminate the 4D vector $n^i h_{i\mu}$ (4 fields) and the scalars $n^i h_{ij}$ (n fields). On shell $n-1$ longitudinal components from the remaining $n-1$ vectors and 5 more components from $h_{\mu\nu}$ are eliminated. The propagating degrees of freedom are correspondingly $n-1$ massive vectors and 1 massive graviton. Thus there finally result $(n+4)(n+1)/2 = D(D-3)/2$ physical states, in agreement with our previous derivation. At the zero mode level there is the same number of degrees of freedom, but they are shared among 1 graviton $h_{\mu\nu}^{(0)}$, $2n$ graviphotons A_μ^i and a symmetric matrix of $n(n+1)/2$ scalars ϕ_{ij}. The scalars ϕ_{ij} are the moduli describing the fluctuations in the shape and size of the torus. In particular the trace ϕ_i^i describes the fluctuations of the torus volume. This field mixes to the 4D graviton leading to a propagator with an extra scalar term

$$\frac{1}{M_4^2}\{\frac{\frac{1}{2}\left(\eta_{\mu\rho}\eta_{\sigma\nu} + \eta_{\mu\sigma}\eta_{\rho\nu}\right) - \frac{1}{2}\eta_{\mu\nu}\eta_{\rho\sigma}}{q^2} + \frac{n}{2n+4}\frac{\eta_{\mu\nu}\eta_{\rho\sigma}}{q^2}\}. \qquad \text{(I.57)}$$

In order to agree with observations, the volume modulus should be stabilized.

I.6 Large Extra-Dimensions

A very interesting arena where to apply the concepts that we introduced is given by the scenario of large extra-dimensions. This scenario has been advocated by Arkani-Hamed, Dimopoulos and Dvali (ADD) as an alternative viewpoint on the gauge hierarchy problem [2]. With respect to the standard picture for physics beyond the SM the ADD proposal represents a dramatic shift of paradigm. In the standard scenario, fundamental interactions are described by an ordinary quantum field theory up to energy scales larger that the Grand Unification scale 10^{16} GeV. Above this scale quantum gravity effects or string theory imply a radical revision of fundamental physics. According to the ADD proposal, instead, this radical revision is needed right above the weak scale! The proposal is specified by three main features

- There exists a number of n new spatial compact dimensions. For instance a simple manifold could be just $\mathcal{M}_4 \times T^n$.
- The fundamental Planck scale of the theory is very low $M_D \sim$ TeV.
- The SM degrees of freedom are localized on a 3D-brane stretching along the 3 non-compact space dimensions.

As we will now explain, these three requirement allow for a drastically different viewpoint on the hierarchy problem, without leading to any stark disagreement with experimental observations. Let us focus on gravity first. As we have already seen in the simple case of Kaluza-Klein's theory, the macroscopic Planck mass M_4^2 of the effective 4D theory is related to the microscopic M_D via

$$M_4^2 = M_D^{2+n} V_n \tag{I.58}$$

where V_n is the compactification volume. For a torus we have $V_n = (2\pi R)^n$ and the above result follows from a simple generalization of the analysis we previously did on S_1. We can also obtain this relation by considering directly the effective action for a purely zero mode $g_{\mu\nu}(x^\mu, x^i) \equiv \bar{g}_{\mu\nu}(x^\mu)$ fluctuation of the metric along $\mathcal{M}_4$

$$2M_D^{2+n} \int d^4x^\mu d^n x^i \sqrt{g} \mathcal{R}_D(g) \quad \leftrightarrow \quad 2M_D V_n \int d^4x^\mu \sqrt{\bar{g}} \mathcal{R}_4(\bar{g}) \tag{I.59}$$

where we have explicitly indicated the dimensionality of the Ricci tensor. The main remark of ADD is based on eq. (I.58). Provided the volume of compactification is large enough, even a low gravity scale M_D can reproduce the physical value $M_4 = 2 \times 10^{18}$ GeV. Before discussing the

n	R
1	6×10^{13} cm
2	$0.4\,\mathrm{mm} = 1/(10^{-4}\mathrm{eV})$
4	$10^{-8}\,\mathrm{mm} = 1/(20\mathrm{KeV})$
6	$2.5 \times 10^{-11}\,\mathrm{mm} = 1/(10\mathrm{MeV})$

Table 2. Radius of compactification for fixed value of $M_D^{GRW} = 1$ TeV , where $(M_D^{GRW})^{2+n} \equiv 4(2\pi)^n M_D^{2+n}$ is the Planck mass defined in the first paper of ref. [15].

needed size of R, notice that eq. (I.58) has a very simple interpretation via Gauss's theorem. Consider the Newtonian potential $\varphi \equiv h_{00}/2$ generated by a test mass M in the linearized approximation. At a distance $r \ll R$ the compactness of the extra-dimensions does not play a relevant role: the potential is to a good approximation $SO(3+n)$ symmetric and given by

$$\varphi|_{r \ll R} \simeq -\frac{\Gamma(\frac{n+3}{2})}{(2n+4)\pi^{\frac{3+n}{2}}} \frac{1}{M_D^{2+n}} \frac{M}{r^{1+n}}. \tag{I.60}$$

as dictated by Gauss's theorem. At $r \gg R$ the field lines stretch along the 3 non-compact directions, the potential is only $SO(3)$ symmetric. The surface encompassing the field flux is now the two sphere (non-compact directions) times the compactification manifold; for instance $S_2 \times T^n$. Applying Gauss's theorem we find then

$$\varphi|_{r \gg R} \simeq -\frac{n+1}{16\pi(n+2) M_D^{2+n} V_n} \frac{M}{r} \equiv \frac{n+1}{16\pi(n+2)} \frac{1}{M_4^2} \frac{M}{r} \tag{I.61}$$

from which we recover again eq. (I.58). In practice the large distance field is made weaker by the large extradimensional volume in which the field lines can spread. (The dependence of eq. (I.61) on n is due to the massless radion. For $n = 1$ it agrees with eq. (I.51), while for general n eq. (I.61) is simply the Fourier transform of eq. (I.57).).

If the ultimate cut-off M_D is of order the weak scale itself $G_F^{-1/2}$, then the expected quantum corrections to the Higgs mass are of the order of its phenomenologically favored value $m_H \sim G_F^{-1/2}$. In this respect the hierarchy problem, in its ordinary formulation, is practically eliminated when $M_D \sim 1$ TeV. With this input, and with the observed value of M_4, eq. (I.58) predicts the size of V_n. In Table 2, we give the radius of compactification in the case of a square n-torus. We stress, see eqs. (I.60,I.61), that Newton's law is reproduced only at distances large than R. The case $n = 1$ requires a radius of compactification of the size of the solar system, which is largely ruled out. However already for n greater or equal than 2 the resulting radius is not unreasonable. Indeed

experimental tests of gravity at distances shorter than a millimeter are extremely arduous. This is largely due to the presence of Van der Waals forces, which tend to swamp any interesting measurement. At present the best bound relegates $O(1)$ deviations from Newton's law (the ones we would expect is our scenario at $r \lesssim R$) to distances shorter than $200\mu m$. In this respect the case $n = 2$ is not barely inconsistent. $n = 2$ is also experimentally interesting, as it predicts deviations in the range of present sensitivities. The search for deviations from Newton's law is an active experimental field, also greatly stimulated by the ADD proposal.

Focusing on gravity only, we have shown that for $n \geq 2$ the radius of compactification is small enough. On the other hand the Standard Model has been verified down to distances much shorter than the radii shown in the table. The SM is a 3+1 dimensional quantum field theory and its predictions depend crucially on this property. LEP, SLC and Tevatron have tested the SM up to an energy of order 1 TeV, corresponding to a distance of order 10^{-16} cm. Experimentally then, the SM is a 3+1 dimensional system down to a distance much shorter that the radius of compactification. Localizing all the SM degrees of freedom on a 3-brane is an elegant way to realize this experimental fact, while keeping larger radii of compactification. Now it will be the brane size, or whatever other characteristic brane cut-off scale, perhaps $1/M_D$ itself, to characterize the length scale down to which the SM is a valid effective field theory. This scale can conceivably be $\gtrsim 1$ TeV. For instance, the ADD scenario could be realized in type I string theory [16, 17] with the SM localized on a D-brane. In this case the string scale M_S, governing the mass of Regge resonances, acts as UV cut off of the brane effective theory.

This completes the basic description of the ADD scenario. It must however be said that, as it stands, the ADD proposal is a reformulation of the hierarchy problem and not yet a solution [18]. Instead of the small Higgs vacuum expectation value (VEV) of the old formulation, we now need to explain why the compactification volume V_n is so much bigger that its most natural scale $1/M_D^n$:

$$V_n M_D^n \sim 10^{33}. \tag{I.62}$$

V_n, or equivalently the radius R, is a dynamical degree of freedom, a scalar field. We have already shown that in the case of T^n the fluctuation of $V_n \equiv (2\pi R)^n$ corresponds at linear order to the trace h_i^i. Since we want a large $\langle R \rangle$ the scalar potential $V(R)$ will have to be much flatter than naively expected at large values of R. As far as we know, the most natural way to achieve such flat potentials is by invoking supersymmetry. So, if the ADD scenario is realized in Nature it is likely to be so together with supersymmetry at some stage. Notice that in the conventional

formulation of the hierarchy problem supersymmetry is invoked to ensure a flat potential at small values of the Higgs field, *i.e.* a small Higgs mass. As a matter of fact, the ADD proposal maps a small VEV problem into a basically equivalent large VEV problem. In the new scenario the hierarchy problem has become a sort of cosmological constant problem. Indeed a vacuum energy density Λ^{4+n} would add to the radius potential a term $\sim \Lambda^{4+n} R^n$. This grows very fast at large R so we expect [18] that Λ^{4+n} should be much smaller than its natural value $(\,\mathrm{TeV})^{4+n}$. In this respect the presence of bulk supersymmetry would be a natural way to enforce a small Λ^{4+n}, thus helping to explain the large volume. Indeed [19] presents a simple mechanism which produces large radii at $n = 2$, but which works for a vanishing bulk cosmological constant Λ^6. Although the model considered is not supersymmetric it is conceivable that the same mechanism will generalize to a supersymmetry set-up and thus lead to a truly natural generation of the hierarchy.

One reason why the ADD proposal is important is theoretical. The hope is that such a drastic revision of our view of fundamental interactions may open the way to new solutions to old problems, like the cosmological constant problem for instance. Having string theory right at the weak scale may also end up being the right ingredient to build the right string model. However none of these breakthroughs has come yet. The interest in the ADD proposal is at the moment associated to its potentially dramatic phenomenological implications [20]. There are two classes of laboratory tests of large extra-dimensions. We have already commented on the first class, the search for deviations from Newton's law at short but macroscopic distances. This is done in table top experiments. These deviations could be determined by the light moduli, like the radius R [18], or by the lowest Kaluza-Klein (KK) J=2 modes. Another source of deviation could be the lowest KK mode of a bulk vector field gauging baryon number [20]. At present, $O(1)$ deviations from Newton's law have been excluded down to a length $\sim 200\,\mu$m [21], while forces that have a strength $> 10^4$ of gravity are bounded to have a range smaller than $20\,\mu$m [22, 23]. Notice that this class of effects crucially depends on the features of the compactification manifold at large "lengths", as they determine the masses of the lightest modes. For instance the presence of even a small curvature of the compactification manifold can drastically affect these prediction by lifting the lightest states. On dimensional grounds, if the typical curvature length is L the modes with mass $< 1/L$ will be affected and possibly made heavier.

The second class of tests is given by high energy collisions [15]. In this case we deal with either gravitons at virtuality $Q \gg 1/R$ or with real gravitons measured with too poor an energy resolution to distinguish

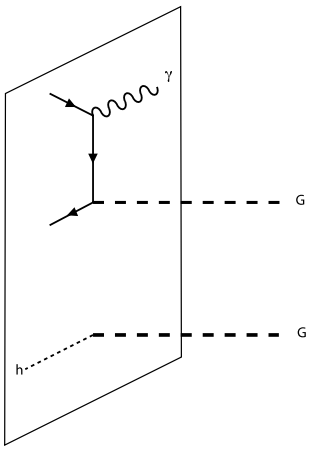

Figure 7. Example of two processes with missing energy by bulk graviton radiation.

individual KK levels. In practice, for high energy processes happening on a short time scale $\tau \sim 1/E \ll R$, causality and locality imply that we cannot notice that the extra-dimensions are compact. Therefore we can take the limit $R \to \infty$ and work as if our brane were embedded in infinite $(4+n)$-dimensional Minkowski space. (If the compactification manifold had curvature length $L \ll R$, then the same reasoning would apply for energies $E \gg 1/L$). Moreover at energy $E < M_D$, as discussed in section I.4, we can reliably compute the amplitudes in a systematic derivative expansion. The characteristic signals are then associated to the emission of gravitons (G) which escape undetected into the extra-dimensions. Interesting examples (see the figure) are given by the processes $e^+e^- \to \gamma + G = \gamma + \not{E}$ or $pp \to \text{jet} + \not{E}$ or possibly by the invisible decay of the Higgs into just one graviton [24]. The latter process does not violate the conservation of angular momentum, since there are KK gravitons of spin 0. The relevant interactions are obtained by expanding the brane action for the SM (as in eq. (I.13)) in powers of the induced metric. For processes with the emission of one graviton we have just the energy momentum term

$$\mathcal{L}_{int} = -\frac{1}{2} T^{SM}_{\mu\nu} h^{\mu\nu}(x, y^i = 0) \tag{I.63}$$

where we have assumed the brane to be located at $y^i = 0$. Emission rates can be computed by expanding $h_{\mu\nu}(x, y_i = 0)$ in KK modes. For

instance, the differential cross section for $e^+e^- \to \gamma G$

$$\frac{d^2\sigma}{dx_\gamma d\cos\theta}(e^+e^- \to \gamma G) = \frac{\alpha}{128}\ \frac{1}{(4\pi)^{\frac{n}{2}}\Gamma(\frac{n}{2})}\ \left(\frac{s^{\frac{n}{2}}}{M_D^{n+2}}\right) \left[F_n(x_\gamma, \cos\theta) + O(\tfrac{s}{M_D^2})\right] \tag{I.64}$$

$$F_n(x,y) = \frac{2(1-x)^{\frac{n}{2}-1}}{x(1-y^2)}\left[(2-x)^2(1-x+x^2) - 3y^2x^2(1-x) - y^4x^4\right]. \tag{I.65}$$

Here $x_\gamma = 2E_\gamma/\sqrt{s}$, E_γ is the photon energy, and θ is the angle between the photon and beam directions. At leading order in E/M_D this process is predicted just in terms of one new parameter, M_D itself. Higher order corrections will depend on new operator coefficients. Similar results can be found for all the other processes [15].

Notice that graviton emission violates the conservation of momentum along the y directions. This is not surprising since the presence of the brane at $y = 0$ breaks translation invariance. However one might worry that non-conservation of the brane energy momentum might lead to inconsistencies in the interaction with gravity; after all Einstein equations demand energy momentum conservation. But this is not the case. Even though the *global* momentum P_i along y_i is not conserved (or better not defined) the full energy momentum tensor T_{MN} for the brane plus matter is indeed *locally* conserved. Conservation of a local current in the absence of a globally conserved charge is the landmark of spontaneous symmetry breaking, and is precisely what happens here. As we have said before, translations in y^i are a non-linearly realized symmetry, with the branons Y^i acting as Goldstone bosons. Local conservation of momentum implies the presence of the branons and hence their production in fundamental processes. Since the brane is infinitely massive it cannot undergo a *global* recoil, but conservation of its energy momentum tensor implies the possibility of *local* recoil by branon emission. The emission of branons is another possible signature of the braneworld scenario. Branons Y are emitted in pairs. At lowest order one finds [25]

$$\frac{d^2\sigma}{dx_\gamma d\cos\theta}(e^+e^- \to \gamma YY) = \frac{\alpha}{1920\pi^5}\ \left(\frac{s^3}{f^8}\right)\left[F_6(x_\gamma, \cos\theta) + O(\frac{s}{M_D^2})\right] \tag{I.66}$$

which, up to an overal constant, is the same result one obtains for graviton emission at $D = 10$. A similar result is obtained in the case of hadronic collisions. Comparing eqs. (I.64,I.66) to experiments one obtains experimental bounds of the scales M_D and f. As the effect grows with $\sqrt{s}$ the best bounds are obtained from the higher energy experiments, LEP2 and Tevatron. In particular LEP2 implies the bound

n	2	3	4	5	6
M_D^{GRW} (TeV)	> 1.45	> 1.09	> 0.87	0.72	0.65

Table 3. Collider bounds on M_D^{GRW}

$f > 100$ GeV [25]. The combined LEP Tevatron bounds on M_D [27]are shown in Table 3. These direct bounds are not very strong. The reason is that the cross section is suppressed by a rather large power of E over the high scale. The LHC will be a better machine to test these ideas through the direct production of gravitons or string states [15, 26]. At present stronger bounds come from other effects, associated to contact 4 fermion interactions, that are expected in the ADD scenario model without being a distinctive feature. The translation of the bounds from these other effects into bounds on M_D is a model dependent procedure, but it is fair to say that they roughly imply $M_D \gtrsim 3$ TeV [27]. Finally we should mention that, with enough luck, the LHC may also study gravitational scattering at energies in excess of the Planck mass M_D, the so-called transplanckian regime. For $M_D \sim 1 - 3$ TeV, the most energetic collisions at LHC, at $\sqrt{s} = 14$ TeV, should start manifesting the transplanckian dynamics, which consists of black-hole [28] or string ball [29] production and also of the characteristic gravitational elastic scattering [30]. The common features of these processes is to give cross section at high energy, and fixed angle, that asymptotically grow like a power of energy. This would be an undisputable signal that the high energy dynamics of gravity, a force whose associated charge is energy itself, has been detected.

Computations like those we have outlined are relevant also to study the cosmological and astrophysical implications of the ADD scenario. The phenomenology of these models is now a very wide field. Unfortunately in order to cover it appropriately we would be lead outside the main goal of the present lectures, which concerns the basic physics principles and tools. Excellent pedagogical reviews of recent results with extensive references are given by refs. [31, 32].

II. Part II: the Randall-Sundrum model

In the second part of these lecture we will focus on a specific model proposed by Randall and Sundrum (RS) which is interesting both from the theoretical and the pedagogical viewpoint. On one side the RS model realizes a new way of approaching the hierarchy problem, on the other it is simple enough to allow a number of rather instructive computations.

Moreover it naturally demands the introduction of the concept of holography, the basis of the AdS/CFT correspondance [33, 34]. All these reasons make the RS model a very rich and instructive lecture subject.

II.1 The Model

Let us consider a model with a 5th dimension and let us compactify it by considering the following equivalence relations for the fifth coordinate y

$$y \sim y + 2\pi \qquad y \sim -y. \tag{II.67}$$

The first relation, alone, would define a circle S_1. The second relation, a Z_2 reflection, implies identification of opposite points on the circle, as shown in the Fig. (8). $y = 0$ and $y = \pi$ are fixed points under Z_2 on the circle and are identified with themselves. The resulting space from this identifications is called a S_1/Z_2 orbifold. S_1/Z_2 is equivalent to the $[0, \pi]$ segment, but for computational purposes it is useful work with the full S_1 covering space with Z_2 identification. Consider now the metric

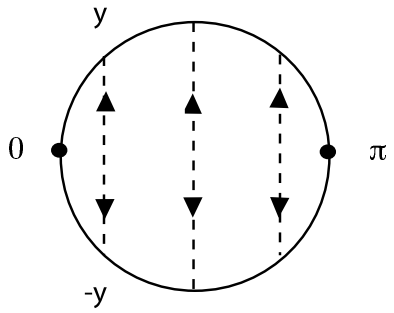

Figure 8. Graphical representation of S_1/Z_2.

on this space

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu + 2g_{\mu 5}dx^\mu dy + g_{55}dy^2. \tag{II.68}$$

The points related by $y \to -y$ are physically identical, and under Z_2 the interval ds^2 should be invariant. Since $dy \to -dy$ under Z_2, we then have

$$g_{\mu\nu}(x,y) = g_{\mu\nu}(x,-y) \quad g_{55}(x,y) = g_{55}(x,-y) \quad g_{\mu 5}(x,y) = -g_{\mu 5}(x,-y). \tag{II.69}$$

From the last identity and by the continuity of $g_{\mu 5}$ it follows that $g_{\mu 5}\,(x,0) = g_{\mu 5}(x,\pi) = 0$. Changes of coordinates

$$\bar{x}^\mu = \bar{x}^\mu(x,y) \qquad \bar{y} = \bar{y}(x,y) \tag{II.70}$$

should still parametrize an orbifold and respect eq. (II.67). Without loss of generality we can impose $\bar{y}$ to be the new orbifold coordinate satisfying

$$\bar{y}(x, y + 2\pi) = \bar{y}(x, y) + 2\pi \qquad \bar{y}(x, -y) = -\bar{y}(x, y) \tag{II.71}$$

generalizing what we did before for the circle. Notice that under reparametrizations the fixed point are mapped onto themselves $\bar{y}(x, 0) = 0$, $\bar{y}(x, \pi) = \pi$ [5] At the boundaries, eq. (II.70) reduces to 4D reparametrizations of the boundaries themselves

$$\begin{array}{ll} 0) & x^\mu \to \bar{x}^\mu(x, 0) \\ \pi) & x^\mu \to \bar{x}^\mu(x, \pi) \end{array} \tag{II.72}$$

under which the induced metrics

$$g_{0\mu\nu} \equiv g_{\mu\nu}(x, 0) \qquad g_{\pi\mu\nu} \equiv g_{\mu\nu}(x, \pi) \tag{II.73}$$

are covariant tensors. Now, using $g_{0,\pi}$ we can write the most general invariant action involving also fields and interactions localized at the boundaries

$$\begin{aligned} S &= \int d^4x \int_0^{2\pi} dy \Big\{ \sqrt{g} \left[2M_5 \mathcal{R}(g) - \tau \right] \\ &+ \delta(y)\sqrt{g_0} \left[\mathcal{L}_0 - \tau_0 \right] + \delta(y - \pi)\sqrt{g_\pi} \left[\mathcal{L}_\pi - \tau_\pi \right] \Big\} \end{aligned} \tag{II.74}$$

where τ and τ_0, τ_π are respectively the bulk cosmological constant and boundary tensions. By $\mathcal{L}_{0,\pi}$ we indicate any other interactions involving fields localized at the boundary. Neglecting the latter, the 5D Einstein equations are

$$\begin{gathered} \sqrt{g}\, G_{MN} = \\ \frac{-1}{4M_5^3} \Big[\tau \sqrt{g} g_{MN} + \left(\tau_0 \sqrt{g_0} g_{0\mu\nu} \delta(y) + \tau_\pi \sqrt{g_\pi} g_{\pi\mu\nu} \delta(y - \pi) \right) \delta^\mu_M \delta^\nu_N \Big] \end{gathered} \tag{II.75}$$

We look for a solution with the following Poincaré invariant ansatz

$$ds^2 = e^{-2\sigma(y)} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 dy^2 \tag{II.76}$$

[5] On the segment $[0, \pi]$ these are the reparametrizations that do not move the boundaries. One could allow more general reparametrizations under which the boundary points are shifted. The resulting field space would be obviously larger. The physics would however not be affected. From the point of view of this more general formulation our field space is obtained just by a partial gauge fixing.

where r_c is a parameter with dimension [length] parametrizing the proper distance between the two fixed points. The $\mu 5$ equation is identically satisfied while the 55 is

$$\left(\frac{\sigma'}{r_c}\right)^2 = -\frac{\tau}{24M_5^3} \equiv k^2. \tag{II.77}$$

implying a negative cosmological constant $\tau < 0$. Imposing the orbifold symmetry eq. (II.77) has two solutions (up to trivial coordinate changes):

$$\sigma(y) = \pm k r_c |y|. \tag{II.78}$$

Finally the $\mu\nu$ equation is

$$\sigma'' = \frac{r_c}{12M_5^3}\left[\tau_0\delta(y) + \tau_\pi\delta(y-\pi)\right] \tag{II.79}$$

which is solved by eq. (II.78) provided the following conditions hold

$$\tau_0 = -\tau_\pi = \pm 24 M_5^3 k. \tag{II.80}$$

Without loss of generality, we can choose the solution $\sigma = +kr_c|y|$. Notice that the tension τ_π is negative. We will later show that this does not lead to instabilities. Eq. (II.80) represents a tuning of two parameters, in the absence of which there would not exist a static solution with Poincaré symmetry. This does not seem at all a desired feature for a model aiming at a solution of the gauge hierarchy problem! The meaning of eq. (II.80) will become more clear below. We will then explain that in the complete model there remains just one tuning, the one associated to the 4D cosmological constant. This is a situation common to all the other solutions of the gauge hierarchy problem, like supersymmetry or technicolor.

Our orbifold with metric

$$ds^2 = e^{-2kr_c|y|}dx_\mu dx^\mu + r_c^2 dy^2 \tag{II.81}$$

corresponds to two slices of 5-dimensional anti-de Sitter space (AdS5) glued back to back at the fixed points. The full AdS space is obtained by $|y| \to y$ and for $y \in (\infty, +\infty)$. $k^2 = \mathcal{R}_5(g)$ parametrizes the curvature of this space. In order for our effective field theory to be a valid description it should be $k \ll M_5$. A region with fixed x coordinates exported along y describes a throat that gets exponentially narrow at larger y. Fig. 9 depicts the same contraction for the wavelength of infalling quanta, which we will later discuss in more detail. It is sometimes said that the space is "warped" by the y dependent factor multiplying the 4D metric.

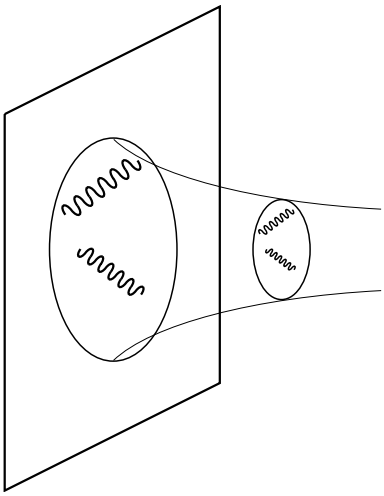

Figure 9. Contraction of wavelengths as particles fall through the AdS throat.

Our metric can be used to redshift 4-dimensional mass parameters as suggested at the beginning of these lectures. Assume we place at 0 and π two identical copies of a 4D QFT. Exactly like with two atoms in the gravitational field of a star, any direct experimental comparison of the masses of the equivalent states at each brane gives

$$\frac{m_\pi}{m_0} = e^{-kr_c\pi}. \tag{II.82}$$

For $e^{-kr_c\pi} \sim m_Z/M_4$ this effect could be relevant in explaining the weak scale hierarchy. Because of the relative shift of mass scales the 0 and π fixed points are called respectively the Planck and TeV brane. Notice also that, because of the appearance of an exponential factor, a redshift of order $m_z/M_4 \sim 10^{-16}$ is already obtained when the proper radius of compactification $r_c\pi$ is only about $35-40$ times larger than the curvature radius $1/k$. The latter could in turn be not much bigger than the 5D Planck length $1/M_5$. Therefore a large hierarchy can be obtained from a rather small fifth dimension.

II.2 Low energy effective theory

The RS model is a generalization of the Kaluza-Klein theory that we already studied. It is then straightforward for us to discuss its zero mode content and to derive the low energy effective field theory. The main difference with respect to gravity on S_1, is that $g_{\mu 5}$ is odd under orbifold parity, see eq. (II.69). Then $g_{\mu 5}$ does not have a graviphoton

zero mode and we can go to a gauge where $g_{\mu 5} \equiv 0$. On the other hand the graviton and radion zero modes are just obtained by promoting the Lorentz metric $\eta_{\mu\nu}$ and radius r_c to 4-dimensional fields [3]

$$ds^2 = G_{MN} dX^N dX^M = e^{-2kT(x)|y|} \bar{g}_{\mu\nu}(x) dx^\mu dx^\nu + T(x)^2 dy^2. \quad \text{(II.83)}$$

Notice indeed that when $\bar{g}_{\mu\nu}$ and T are constant over space-time, eq. (II.83) is a solution of the equations of motion.($T = r_c$ is not fixed by the solution, while the solutions with arbitrary constant $\bar{g}_{\mu\nu}$ is simply obtained from the original solution by applying the coordinate change, $x^\mu \to A^\mu_\nu x^\nu$, with $A^\rho_\mu A_{\rho\nu} = \bar{g}_{\mu\nu}$.) Then the 5D action is stationary over the field manifold of constant $\bar{g}_{\mu\nu}$ and T, *i.e.* there is no "potential" for $\bar{g}_{\mu\nu}$ and T, so that they must correspond to massless particles. By substituting the ansatz in eq. (II.83) in the action we find indeed

$$\mathcal{L}_4^{eff} = \int \mathcal{L}_5(\bar{g}, T) dy = \left(\frac{M_5}{k}\right)^3 \sqrt{\bar{g}} \Big\{ \left(2k^2 - 2\mu^2\right) \mathcal{R}(\bar{g}) + 12(\partial\mu)^2 \Big\} \quad \text{(II.84)}$$

where

$$\mu(x) = k e^{-kT(x)\pi}. \quad \text{(II.85)}$$

and where all the metric contractions are done with $\bar{g}_{\mu\nu}$. This Lagrangian correctly describes the interactions of the zero modes up to terms with more than two derivatives. We will explain this in more detail when deriving the KK spectrum. By substituting $\langle T \rangle = r_c$ in the above equations we can read the effective 4-dimensional Planck scale

$$M_4^2 = \frac{M_5^3}{k} \left(1 - e^{-2kr_c\pi}\right). \quad \text{(II.86)}$$

By working with $T \equiv r_c$, eq. (II.83) substituted in the 5D Einstein term gives $\sqrt{g}\mathcal{R}_5(g) = \sqrt{\bar{g}} r_c e^{-2\sigma(y)} \mathcal{R}_4(\bar{g}) + \ldots$. Eq. (II.86) then simply corresponds to the integral

$$M_4^2 = M_5^3 \int_{-\pi}^{\pi} e^{-2\sigma(y)} r_c dy. \quad \text{(II.87)}$$

The crucial aspect of this result is that M_4 is completely dominated by the region close to the Planck brane, where the warp factor is of order 1. Therefore the value of M_4 is insensitive to the redshift of mass scales that takes place in the bulk and which is maximal at the TeV brane.

Now that we have calculated the Planck mass we can discuss the issue of the gauge hierarchy in more detail. Assume we localize all the fields of the Standard Model on the TeV brane. We indicate collectively by H, ψ_α, and A_μ the scalars, fermions and vectors, and by m the mass

parameters (any mass parameter, including the Higgs mass). According to the ansatz of eq. (II.83), and keeping the radion fixed for simplicity, the induced metric at the TeV brane is

$$g_{\pi\mu\nu} = e^{-2\sigma(\pi)}\bar{g}_{\mu\nu}(x). \tag{II.88}$$

so that the low energy effective action including the $y = \pi$ boundary contribution is

$$\mathcal{L}^{eff} = \left\{ \frac{2M_5^3}{k}\left(1 - e^{-2\sigma(\pi)}\right)\sqrt{\bar{g}}\mathcal{R}(\bar{g}) + \sqrt{g_\pi}L_\pi(g_{\pi\mu\nu}, H, \psi_\alpha, \mathcal{A}_\mu, m)\right\}. \tag{II.89}$$

Notice that the metric that couples to matter is rescaled with respect to the one that appears in the Einstein term, as if different length units were used in the two actions. To make physics manifest it is useful to perform a constant Weyl transformation on the matter action. This is just a reparametrization of field variables corresponding to a change of the length unit. Indicating by w the Weyl rescaling parameter, in 4 dimension we have that the various fields transform as

$$(g_{\mu\nu},\, H,\, \psi_\alpha,\, A_\mu\,) = \left(w^{-2}g'_{\mu\nu},\, wH',\, w^{3/2}\psi'_\alpha,\, A'_\mu\right). \tag{II.90}$$

Indicating collectively the fields with Φ and the rescaled ones with Φ' a generally covariant action S satisfies the relation

$$S(\Phi, m) = S(\Phi', \frac{m}{w}). \tag{II.91}$$

One can easily check this result by considering the free scalar Lagrangian

$$\sqrt{g}\left(g^{\mu\nu}\partial_\mu H\partial_\nu H - m^2H^2\right) \tag{II.92}$$

In the case of the matter contribution in eq. (II.89) we make a rescaling with $w = e^{\sigma(\pi)}$ such that $g_{\pi\mu\nu} = w^{-2}\bar{g}_{\mu\nu}$, $H = wH'$, ... and write it as

$$\sqrt{\bar{g}}\mathcal{L}_\pi(\bar{g}_{\mu\nu}, H', \psi'_\alpha, A'_\mu, me^{-\sigma(\pi)}). \tag{II.93}$$

In these new variables it is evident that all mass parameters in the matter Lagrangian are redshifted with respect to the Planck mass $M_4 \simeq \sqrt{M_5^3/k}$. Now the importance of the RS mechanism has become very concrete. We stress once again that the basic reason for this result is that the 4-dimensional Planck mass is dominated by a contribution from the region of low redshift, while the SM lives deep inside a region of high red shift. As a matter of fact the metric $\bar{g}_{\mu\nu}$ that appears in the 4D Einstein term coincides with the induced metric on the Planck brane $g_{0\mu\nu}$. We

will later explain that all these facts corresponds to the localization of the 4D graviton near the Planck brane.

We want to emphasize that eq. (II.91) implies that only ratios of mass parameters are observable in a theory of gravity, since their overall normalization can be trivially changed by a field redefinition. This is the same situation we have in the SM, where all CP violating phases, but one, transform under field redefinitions and can thus be eliminated. Only the ratio of Planck and weak scales, the hierarchy, is an observable. Then we can have an alternative view point of our result, where the original scale in the Lagrangian is the weak scale and the Planck mass is a blue shifted derived scale. Consider indeed a rescaling with $w = e^{\sigma(\pi)}$ on our original Lagrangian and on our original solution. Now all mass parameters in the starting Lagrangian are of order TeV; for instance $M_5 \to M_5' = M_5/w$. Moreover the warp factor is 1 at $y = \pi$ and $e^{\sigma(\pi)}$ at $y = 0$. The 4D Planck mass is now written as

$$M_4^2 = \frac{(M_5')^3}{k'} \left(e^{2k' r_c' \pi} - 1 \right). \tag{II.94}$$

This picture makes more evident that the the hierarchy is generated by a "dilution" mechanism, like in the ADD scenario. In eq. (II.94), M_4^2 comes out large because of the exponential growth of the "volume" towards the Planck brane.

After the derivation of the effective Lagrangian we can better understand the meaning of the tunings imposed on the boundary tensions. As we will show below, the KK spectrum is quantized in units of the radion VEV $\langle \mu \rangle$. For $E \ll \langle \mu \rangle$ the system can be described by the zero modes $\bar{g}_{\mu\nu}$ and μ. Now, if we add to the brane tensions a perturbation which is *small* enough not to excite the KK modes we should be able to accurately describe its effects just in terms of the zero modes. Consider then the following perturbations of the tensions parametrized by $\alpha, \beta \ll 1$

$$\tau_0 = 24 M_5^3 k \, (1 + \alpha) \qquad \tau_\pi = -24 M_5^3 k \, (1 - \beta) \, . \tag{II.95}$$

At lowest order, the correction to the effective Lagrangian is simply obtained by substituting eq. (II.83) in the terms proportional to α and β in the the original Lagrangian

$$\Delta \mathcal{L}_4^{eff} = -24 \left(\frac{M_5}{k} \right)^3 \sqrt{g} \left(\alpha k^4 + \beta \mu^4 \right). \tag{II.96}$$

By integrating out the massive KK there arise extra corrections that have either derivatives in them or are of higher order in α, β. The above equation represents the leading correction to the potential. It is evident that the two tunings of the brane tensions, $\alpha = \beta = 0$, correspond to

1 vanishing radion potential

2 vanishing 4-dimensional cosmological constant.

Now, the second requirement is truly necessary since the the cosmological constant is experimentally extremely small $\Lambda_4^4 \lesssim (10^{-3}\text{eV})^4$. On the other hand, by the first requirement the radion is a massless Brans-Dicke field leading to a new unacceptable long range force. The first tuning is not only useless but experimentally ruled out. The basic RS model must be modified in order to give a mass to the radion while retaining the possibility to fine tune the 4D cosmological constant to zero. In the modified theory there would remain just one tuning.

One can ask if a modifcation of the brane tensions can lead to a realistic radion stabilization. Of course we already know from the derivation of the RS solution that with detuned tensions we would not be able to find a static solution which is also flat from the 4D viewpoint. But it is useful to study this more quantitatively. Notice that β gives rise to a simple quartic potential for μ, so it would seem that the only stationary point is $\mu = 0$, which is not interesting. The situation is however slightly more subtle since μ mixes kinetically with the graviton. In order to easily read the dynamics of μ it is useful to perform a Weyl rescaling of the metric

$$\bar{g}_{\mu\nu} \to \frac{\bar{g}_{\mu\nu}}{1 - \frac{\mu^2}{k^2}} \tag{II.97}$$

to go in the Einstein frame in which the gravitational kinetic term is exactly $2(M_5^3/k)\sqrt{\bar{g}}\mathcal{R}(\bar{g})$ with no radion contribution. At the same time the radion kinetic term receives a negligibly small modification, while the potential term becomes

$$\Delta V = -\Delta\mathcal{L}_4^{eff} = 24\left(\frac{M}{k}\right)^3 \sqrt{g}\frac{\alpha k^4 + \beta\mu^4}{(1 - \frac{\mu^2}{k^2})^2}. \tag{II.98}$$

This potential is stationary at $(\mu^2/k^2) = -\alpha/\beta$. One can easily check that this stationary point corresponds to a minimum only for $\alpha < 0$, $\beta > 0$. Around the stationary point, V is negative so that the 4D metric in turn will be $AdS4$. It turns out that both the $AdS4$ curvature k_4 and the radion mass m_μ scale in the same way $k_4^2 \sim m_\mu^2 \sim -\alpha k^2$. Then this solution does not look even approximately like the real world, and a more realistic mechanism of radius stabilization is needed. Before concluding this section we would like to notice that our simple discussion captures and explains in a physically intuitive way the results of refs. [38, 39], where the full 5D equations of motion in the presence of detuned brane tensions were studied. Our approach also clarifies the results of

early studies of RS cosmology [40, 41], where a puzzling tuning $\rho_0 = -\rho_\pi(\mu/k)^2$ between the energy momentum densities at the two different branes was found to be needed. In our particular set up we have $\rho_0 \equiv k^4\alpha$ and $\rho_\pi \equiv k^4\beta$, showing that the tuning is just the radion stationarity condition in the absence of extra contributions to the potential [42].

II.3 Radius stabilization: Goldberger-Wise mechanism

As we have already remarked, in the RS model the hierarchy between mass scales at the two boundaries depends exponentially on the proper distance πr_c between them. A moderate separation $\pi r_c/L \sim 40$ between the proper length and the AdS length of the 5th dimension is then enough to obtain a huge hierarchy. This would seem a natural way to explain the hierarchy. The implicit assumption behind this conclusion is that r_c is the natural dynamical variable to describe the stabilization of the 5th dimension, *i.e.* the radion potential is practically a polynomial in r_c. However we have so far no solid reason to believe this is the case. As a matter of fact, in the simple example at the end of the previous section, the natural variable was the warp factor μ/k itself. Now, instead of the proper distance, an observer at the Planck brane could decide to define her/his distance to the TeV brane through the time T a light signal (or a graviton!) takes to travel to the TeV brane and back. The result would be

$$T = 2\int_0^\pi \frac{dy}{\sqrt{g_{00}}} = \frac{2}{k}\left(e^{kr_c\pi} - 1\right) \sim \frac{2}{\mu} \tag{II.99}$$

which is exponentially large as a consequence of time dilation near the TeV brane. According to this perfectly acceptable definition, the size of the extra-dimension is controlled by the weak scale μ itself. Notice that the potential of the previous section is polynomial in μ. Such hugely different notions for the size of the 5th dimension arise because the global features of the RS space are controlled by curvature. We could have started with a coordinate system in which, according to the above definition of distance, the 5th coordinate is parametrized by

$$z = \frac{e^{-kr_c y}}{k}. \tag{II.100}$$

The metric would then have been

$$ds^2 = \frac{L^2}{z^2}\left(\eta_{\mu\nu}dx^\mu dx^\nu + dz^2\right). \tag{II.101}$$

where $L = 1/k$, the AdS radius, represents the AdS curvature length. In these "conformal" coordinates there is no exponential factor, but the

locations of the Planck and TeV branes, respectively z_0 and z_1, are very far apart

$$z_0 = L \ll L e^{k r_c \pi} = z_1 \tag{II.102}$$

In these new coordinates the radion field is simply associated to the position of the second brane: $\mu(x) = 1/z_1(x)$.

To solve the hierarchy problem we must then find a mechanims that stabilizes the second brane at $z_1 \sim 1/\text{TeV} \gg z_0 \sim 1/M_P$. In order to do so, Goldberger and Wise (GW) [35] have proposed a simple model involving a 5D scalar field ϕ with action

$$\begin{aligned} S_\phi = \int d^4x dz \, \{ \sqrt{g}[-(\partial\phi)^2 - m^2\phi^2] + \delta(z-z_0)\sqrt{g_0}\mathcal{L}_0(\phi) \\ + \, \delta(z-z_1)\sqrt{g_1}\mathcal{L}_1(\phi)\} \end{aligned} \tag{II.103}$$

It is assumed that the dynamics of the boundary terms $\mathcal{L}_{0,1}$ is such as to fix the values $\phi(z_0) = \tilde{v}_0$ and $\phi(z_1) = \tilde{v}_1$. For instance, one can take $\mathcal{L}_{0,1} = -\lambda_{0,1}(\phi - v_{0,1})^2$ with $\lambda_{0,1} \to \infty$. The assumption of an infinitely steep boundary potential simplifies the computations but is not essential [35, 36]. In the vacuum, the field ϕ will have a non-trivial bulk profile satisfying the 5D Klein-Gordon equation and interpolating between the two boundary values v_0 and v_1. The energy associated to this profile depends on the distance between the two branes, corresponding to a non-vanishing radion potential. Now, solving the coupled equations of motion of gravity plus ϕ exactly is in general difficult. To make our task easier we can make the simplifying assumption that ϕ only induces a small perturbation of the locally $AdS5$ Randall-Sundrum metric. Quantitatively this is equivalent to requiring the scalar energy momentum to be a small perturbation of the 5D cosmological constant

$$T^\phi_{MN} \sim (\partial_z\phi)^2 + m^2\phi^2 \sim \max(v_0^2, v_1^2) \times \max(k^2, m^2) \ll T^{vacuum}_{MN} \sim M_5^3 k^2. \tag{II.104}$$

Since, in order to generate a big hierarchy we will need $m^2 \lesssim k^2$, the above relation simply implies

$$v_{0,1}^2 \ll M_5^3 \tag{II.105}$$

(notice that a scalar field in 5D has mass dimension 3/2). When this condition holds, ϕ is determined at leading order by solving the ϕ equations of motion over the unperturbed RS background with Dirichlet boundary conditions $\phi(z_0) = v_0$ and $\phi(z_1) = v_1$. Because of this second constraint, there are no massless zero modes in the KK tower of ϕ. As long as the ϕ profile is a small perturbation of RS the only light states are those of the unperturbed model: the 4D graviton and the radion. In this respect

the procedure of solving the ϕ equations of motion and calculating the action on the solution corresponds to integrating out the tower of massive ϕ KK modes to obtain an effective Lagrangian for the light modes $\bar{g}$ and μ. This is what we will now do. For a field configuration that does not depend on the 4D coordinates the 5D Klein-Gordon equation becomes

$$\frac{z^5}{L^2}\partial_z \frac{1}{z^3}\partial_z \phi = m^2 \phi \tag{II.106}$$

and the most general solution is

$$\phi = A z^{4+\epsilon} + B z^{-\epsilon} \qquad \epsilon = \sqrt{4 + m^2 L^2} - 2 \simeq \frac{m^2 L^2}{4}\,. \tag{II.107}$$

where we are emphasizing that we will be momentarily interested in the limit $\epsilon \sim m^2 L^2/4 \ll 1$. The boundary conditions fix

$$A = z_0^{\epsilon} \frac{v_0 - v_1 (z_0/z_1)^{4+\epsilon}}{1 - (z_0/z_1)^{4+2\epsilon}} \qquad B = z_1^{-4-\epsilon} \frac{v_1 - v_0 (z_0/z_1)^{\epsilon}}{1 - (z_0/z_1)^{4+2\epsilon}} \tag{II.108}$$

and eq. (II.103) evaluated on the solution yields an effective potential for the radion μ

$$\begin{aligned} V(\mu) &= \frac{1}{1 - (\mu L)^{4+\epsilon}} \Big[(4+\epsilon)\mu^4 (\bar{v}_1 - \bar{v}_0 (\mu L)^{\epsilon})^2 \\ &+ \epsilon L^{-4} (\bar{v}_0 - \bar{v}_1 (\mu L)^{4+\epsilon})^2 \Big] \\ &= L^{-4} F(\mu L) \end{aligned} \tag{II.109}$$

where we have made the substitutions $z_0 = L$, $z_1 = 1/\mu$ and where $\bar{v}_{0,1} = L^{3/2} v_{0,1}$ are the boundary VEVs in units of the AdS curvature. For $\mu L \ll 1$ the potential becomes

$$V = \epsilon \bar{v}_0^2 L^{-4} + \Big[(4+2\epsilon)\mu^4 (\bar{v}_1 - \bar{v}_0 (\mu/\mu_0)^{\epsilon})^2 - \epsilon \bar{v}_1^2 \mu^4 \Big] + \mathcal{O}(\mu^8 L^4) \tag{II.110}$$

which for $\epsilon > 0$ is minimized at

$$\mu L \simeq \left(\frac{\bar{v}_1}{\bar{v}_0} \right)^{\frac{1}{\epsilon}}\,. \tag{II.111}$$

The hierarchy $\langle \mu \rangle L \sim M_W/M_P = 10^{-17}$ can be naturally obtained for fundamental parameters not much smaller than one (ex. $\bar{v}_1/\bar{v}_0 \sim 1/10$ and $\epsilon \sim 1/20$). The hierarchy naturally arises because of the non-analytic dependence of eq. (II.111) on ϵ . Some comments are now in order.

1 Notice that the relevant part of the potential (the term in square brackets in eq. (II.110)) has the form $\mu^4 P(\mu^\epsilon)$. This is basically a quartic potential modulated by a slow evolution of the effective coupling $\lambda \equiv P(\mu^\epsilon)$. Notice indeed that for $\epsilon \ll 1$ one has $\mu^\epsilon \sim 1 + \epsilon \ln \mu + \ldots$, so that the dependence on μ is reminiscent of a slow RG evolution. Stability of the potential at large μ corresponds to $\lim_{\mu\to\infty} P(\mu) > 0$. Because of the slow evolution of P, a minimum of the potential will exist very close to the point where P crosses zero to become negative, see eqs. (II.110,II.111). These properties are in full analogy with the Coleman-Weinberg (CW) [37] mechanism of dimensional transmutation. There, quantum corrections to the effective potential cause the quartic coupling to turn negative at some scale. Since this happens through the slow logarithmic RG evolution, broadly different mass scales can arise, making the CW mechanism very interesting to explain the weak to Planck scale hierarchy. It seems that the GW mechanism works qualitatively in the same way.

2 Expanding at second order around the minimum and using the unperturbed kinetic Lagrangian we find the radion mass

$$m_\mu^2 \simeq \frac{2}{3}\bar{v}_1^2 \epsilon^{3/2} \langle\mu\rangle^2 \ll \langle\mu\rangle^2 \qquad \text{(II.112)}$$

indicating that in the model at hand the radion is much lighter than the other KK resonances and potentially more interesting for collider phenomenology.

3 The potential at the minimum is dominated by the fist term in eq. (II.110): $V_{min} \simeq \epsilon \bar{v}_0^2 L^{-4} \sim M_4^4$, far too large! However we can now go back and slightly modify the brane tensions as we did in the previous section. For small α and β, at leading order this amounts to adding the contribution in eq. (II.96) to eq. (II.110). Around the minimum of the GW potential μ is massive so that $\beta \neq 0$ causes only a small shift in $\langle\mu\rangle$. We no longer need to tune $\beta = 0$ after the radius is stabilized. Moreover both α and β cause a shift in the potential at the minimum. By properly choosing one combination of them we can cancel the potential at the minimum. This is of course a fine tuning, but it is just the cosmological constant problem, which we never promised to solve. After radius stabilization the only fine tuning left in the model is the one associated to the vanishing cosmological constant.

4 The minimal GW potential gives rise to a stable minimum only for $\epsilon \simeq m^2 L/4 > 0$. However one could also obtain stable minima for $\epsilon < 0$ for a proper range of the detuning parameter β.

Concerning the last remark one may worry that $\epsilon \sim m^2 < 0$ will lead to instabilities in the bulk. If $|m^2|$ is small enough, however, no instability arises. Concerning this result it is instructive to consider the following exercises.

- Show that no instability is generated for $m^2 > -4k^2 + (\text{small})$ (hint: study the KK spectrum for the excitation $\Delta\phi$ around the GW background remembering that $\Delta\phi = 0$ at the boundaries).
- Study the same problem with a flat 5th dimension.

II.4 Kaluza-Klein spectrum

In this section we shall study the KK spectrum of the gravitational field. As for the flat case we studied earlier, it is convenient to use the 5D gauge freedom to eliminate as many degrees of freedom from $g_{\mu 5}$ and g_{55} as possible. It is then easy to check (for instance by working at the linearized level) that without loss of generality we can use the following parametrization

$$ds^2 = e^{-2k|y|r_c - 2f(x)e^{2k|y|r_c}} \left(\hat{g}_{\mu\nu}(x,y)dx^\mu dx^\nu\right) + \left(1 + 2f(x)e^{2k|y|r_c}\right)^2 dy^2 \tag{II.113}$$

Because of the orbifold projection we can consistently set $g_{\mu 5} \equiv 0$. On the other hand for g_{55} we can eliminate all the modes but one, the radion. This one mode remaining in g_{55} can however be parametrized at our convenience, and the choice in eq. (II.113) as an advantage that we will now explain. Since we are mostly interested in the $J = 2$ modes, consider first substituting eq. (II.113) with $f(x) = 0$ in the RS action. After changing the 5th coordinate to $z = Le^{-k|y|r_c}$, one finds

$$S(g) = 4M_5^3 L^3 \int_L^{z_1} \sqrt{\hat{g}} \frac{dz}{z^3} \left[\mathcal{R}_4(\hat{g}) + \frac{1}{4}\left((\partial_z \hat{g}^{\mu\nu})(\partial_z \hat{g}_{\mu\nu}) + (\hat{g}^{\mu\nu}\partial_z \hat{g}_{\mu\nu})^2\right)\right]. \tag{II.114}$$

Notice that all terms with no derivative acting on $\hat{g}_{\mu\nu}(x, y(z))$ have judiciously cancelled out when expanding around the RS solution. We already know that it should have been this way: any constant $\hat{g}_{\mu\nu}$ solves the equations of motion, so that the action for $\hat{g}_{\mu\nu}$ must involve only gradient terms. When $f \neq 0$ we find an extra contribution to be added to eq. (II.114) and starting at quadratic order in f

$$\Delta S = -12 \frac{M_5^3}{L}\left(z_1^2 - L^2\right) \partial_\mu f \partial^\mu f + O(f^3, f^2 h_{\mu\nu}, \ldots) \tag{II.115}$$

where $h_{\mu\nu} = \hat{g}_{\mu\nu} - \eta_{\mu\nu}$. Notice that there is no kinetic mixing between f and the metric $h_{\mu\nu}$. In other words, the graviton is automatically in the Einstein frame. This is because of our specific parametrization of the scalar mode in eq. (II.113). This should be contrasted to the parametrization in eq. (II.83) which lead to a small $O(\mu^2 L^2)$ mixing between radion and graviton. Anyway, neglecting $O(\mu L)$, terms and with the identification (cf. eqs. (II.83,II.85))

$$\mu(x) = ke^{-k\pi r_c - f(x)e^{2k\pi r_c}} = ke^{-\sigma(\pi,x)} \tag{II.116}$$

eq. (II.115) reproduces the radion kinetic term of eq. (II.84).

Let us now focus on the $J = 2$ modes. Expanding eq. (II.114) at quadratic order in $h_{\mu\nu} = \hat{g}_{\mu\nu} - \eta_{\mu\nu}$ we find

$$M_5^3 L^3 \int_L^{z_1} \frac{dz}{z^3} \left(h_{\mu\nu} K^{\mu\nu\rho\sigma} h_{\rho\sigma} - h^{\mu\nu} \partial_z^2 h_{\mu\nu} + h^\mu_\mu \partial_z^2 h^\nu_\nu \right) \tag{II.117}$$

where $K^{\mu\nu\rho\sigma}$ is the 4-dimensional $J = 2$ kinetic operator shown in the first line of eq. (I.43). Notice that the mass operator has the Fierz-Pauli form. From eq. (II.117) the equation for the eigenmodes $\psi_n(z)$ is then simply

$$-z^3 \partial_z \frac{1}{z^3} \partial \psi_n(z) = m_n^2 \psi_n(z) \tag{II.118}$$

with boundary conditions

$$\partial_z \psi_n|_{z=L} = \partial_z \psi_n|_{z=z_1} = 0, \tag{II.119}$$

as $h_{\mu\nu}$ is even under the orbifold parity and its action does not contain boundary terms.

For $m_n \neq 0$ the general solution to eq. (II.118) is written in terms of Bessel functions

$$\psi_n(z) = \frac{z^2}{L^2} \left[J_2(m_n z) + b_n Y_2(z m_n) \right] \tag{II.120}$$

and the boundary conditions enforced by using the identity

$$\partial_z \psi_n \propto z^2 \left[J_1(m_n z) + b_n Y_1(z m_n) \right]. \tag{II.121}$$

We then find

$$b_n = \frac{J_1(m_n L)}{Y_1(m_n L)} \tag{II.122}$$

while the eigenvalue equation is simply

$$J_1(m_n L) Y_1(m_n z_1) - Y_1(m_n L) J_1(m_n z_1) = 0. \tag{II.123}$$

In order to focus on the phenomenologically interesting "light" modes, satisfying $m_n L \ll 1$, we need the asymptotic behaviour of the Bessel functions

$$\begin{array}{lll} x \ll 1 & \rightarrow & \left\{ \begin{array}{l} J_2(x) = \frac{x^2}{8} + O(x^4) \\ Y_2(x) = -\frac{4}{\pi x^2} - \frac{1}{\pi} + O(x^2) \end{array} \right. \\ x \gg 1 & \rightarrow & \left\{ \begin{array}{l} J_2(x) = \sqrt{\frac{2}{\pi x}} \cos(x - \frac{5}{4}\pi) \\ Y_2(x) = -\sqrt{\frac{2}{\pi x}} \sin(x - \frac{5}{4}\pi) \end{array} \right. . \end{array} \tag{II.124}$$

In the limit $m_n L \ll 1$ eqs. (II.122) and (II.123) reduce respectively to

$$b_n \simeq (m_n L)^2 \pi/4 \ll 1 \tag{II.125}$$

and

$$J_1(m_n z_1) = 0 \tag{II.126}$$

The solutions to the last equation are quantized in units of $1/z_1 \equiv \langle \mu \rangle$

$$m_n = c_n \frac{1}{z_1} \simeq (n + \frac{1}{4}) \frac{\pi}{z_1}, \tag{II.127}$$

where the last identity is valid asymptotically for $n \gg 1$ (cf. eq. (II.124)) but works very well already for $n = 1$: $c_1 \simeq 1.21\pi$. Notice that because $b_n \ll 1$, Y_2 makes a negligible contribution to ψ_n in the region $m_n z \gg 1$ where ψ_n oscillates. In the region $z m_n \ll 1$, the Y_2 contribution is relatively important (in fact dominant), but this region contributes negligibly to the normalization of the eigenfunctions.

Indeed, by eq. (II.117) the norm of the modes is

$$\begin{aligned} ||\psi_n||^2 &= \int_{z_0=L}^{z_1} \frac{2z dz}{L^2} \left[J_2(m_n z) + b_n Y_2(m_n z) \right]^2 \simeq \int_0^{z_1} \frac{z dz}{L^2} \left[J_2(m_n z) \right]^2 \\ &= \left(\frac{z_1}{L} \right)^2 J_2(m_n z_1)^2. \end{aligned} \tag{II.128}$$

Neglecting contributions of relative size $O(L^2/z_1^2) \sim m_{weak}^2/M_P^2$, we have approximated $||\psi_n||$ by its value in the limit $z_0 = 0$, in which the Planck brane has an infinite proper distance from the TeV brane (and an infinite relative blueshift, see eq. (II.101)). $||\psi_n||$ is dominated by the region $1/m_n < z < z_1$, where the Bessel functions have an oscillatory behaviour. A mode with mass m_n is not very sensitive to the region $z < 1/m_n$. In particular the modes remain normalizable and the spectrum discrete even for $z_0 \rightarrow 0$. Notice finally that the orthonormal modes $\hat{\psi}_n(z) \equiv \psi_n(z)/||\psi_n||$ satisfy $\hat{\psi}_n(z_1) = sgn(J_2(m_n z_1)) z_1/L$, so that all modes couple with equal strength to the fields on the SM brane.

The massless mode, see eq. (II.118), simply corresponds to a constant $\psi_0 = 1$ with norm

$$||\psi_0|| = L^2 \int_{z_0=L}^{z_1} \frac{2dz}{z^3} = \left(\frac{1}{L^2} - \frac{1}{z_1^2}\right) \tag{II.129}$$

which is dominated at small z. Unlike the massive modes, ψ_0 is localized at the Planck brane: $||\psi_0||$ diverges for $z_0 \to 0$ but remains finite for $z_1 \to \infty$. From eq. (II.114) we have that the 4D Planck mass is $M_4^2 = M_5^3 L^3 ||\psi_0||^2 = M_5^3 L[1 - (L/z_1)^2]$, which is the result we obtained previously. Notice also that for $z_0 \to 0$ the 4D Planck mass diverges, corresponding to the zero mode graviton being localized infinitely far away from the TeV brane.

Using the above results, but working with the canonically normalized KK gravitons $\bar{h}_{\mu\nu}^{(n)}$ and radion $\bar{f}$ the coupling to the TeV brane energy momentum tensor is written as

$$\mathcal{L}_{int} = -\frac{1}{2}\left\{\frac{\bar{h}_{\mu\nu}^{(0)}}{M_4} + \sum_{n\geq 1}\frac{\bar{h}_{\mu\nu}^{(n)}}{\Lambda}\right\} T^{\mu\nu} + \frac{\bar{f}}{\sqrt{24}\Lambda} T_\mu^\mu \tag{II.130}$$

where $\Lambda \equiv (M_5 L)^{3/2}/z_1$. eq. (II.130) is the basic equation to study the collider implications of the RS model [43]. The interactions and spectrum of the $J = 2$ modes are fully described by two parameters, Λ and $\langle\mu\rangle = 1/z_1$. For radion phenomenology two extra parameters are needed, one is the radion mass, which depends on the stabilization mechanism, the other is a radion-Higgs mixing parameter ξ [24]. Basically the parameter ξ accounts for the fact that the energy momentum tensor for a scalar H is defined up to an "improvement term" $\Delta T_{\mu\nu} = \xi(\partial_\mu\partial_\nu - \eta_{\mu\nu}\partial^2)H^2$. A non zero ξ induces a kinetic mixing between radion and Higgs after electroweak symmetry breaking.

A little exercise one can do with the interaction Lagrangian in eq. (II.130) concerns the validity of perturbation theory. A simple quantity to calculate (estimate) is the decay width into SM particles (living at the TeV boundary). For the n-th mode we find

$$\Gamma_{(n)} \sim \frac{m_n^3}{8\pi\Lambda^2}. \tag{II.131}$$

Notice that $\Gamma_{(n)}$ grows with n, so that for n large enough the nearby resonances will start overlapping. When this takes place, by definition, perturbation theory breaks down: quantum corrections to the spectrum (the widths) make the very concept of individual KK resonances useless. Using the above equations we find that $\Gamma_{(n)} < m_{n+1} - m_n$ is satisfied

for masses below

$$m_n < M_5 \frac{L}{z_1}. \tag{II.132}$$

This shows that the UV cut-off for physics at the TeV scale is just the redshifted Planck mass. Of course by starting off in the primed Weyl frame, discussed above eq. (II.94), this is the obvious result.

II.5 Strong coupling puzzle

At the end of the previous section we have shown by a simple argument that the RS model becomes strongly coupled at a fairly low scale $M'_5 = M_5 L/z_1$. If we we want to explain the hierarchy by the ratio L/z_1, then $M'_5 \sim$TeV very much like in the ADD scenario. Although such a low-cut off limits predictivity, as long as M'_5 is somewhat bigger than the mass of the lightest KK, some control is retained: roughly a number of modes $\sim M'_5 z_1 = M_5 L$ remains weakly coupled. From the conceptual viewpoint, however, the presence of this low cut-off can be confusing, when not properly interpreted. One basic puzzle is that M'_5 depends on the location of the TeV brane. Moreover as $z_1 \to \infty$ and half of $AdS5$ is recovered, M'_5 goes to zero, as if there was no energy range where gravity on AdS makes sense as an effective field theory. The origin of this puzzle is that we are working on a curved space where particle propagation involves large or possibly infinite momentum blue-shift. As we will explain, since the puzzle arises when considering the global aspect of our spacetime, a proper resolution cannot do without accounting for the locality of the interactions.

It is instructive to go back and consider the motion of a particle in AdS space. The equations for the geodesic in conformal coordinates are

$$\ddot{x}^\mu - 2\frac{\dot{x}^\mu \dot{z}}{z} = 0 \qquad \ddot{z} - \frac{\dot{z}^2}{z} + \frac{\dot{x}^\mu \dot{x}^\nu \eta_{\mu\nu}}{z} = 0 \tag{II.133}$$

where by the dot we indicate the derivative with respect to the affine parameter λ. Let us focus on the massless case. As λ is arbitrary, we can choose it such that $dx^M/d\lambda$ coincides with the 5-momentum P^M. By this choice we have, in particular, $d/d\lambda = P^5 d/dz$ and by using the zero mass condition $P^M P^N g_{MN} = 0$ we can write eq. (II.133) and its solution as

$$z\frac{dP^M}{dz} = 2P^M \qquad \to \qquad P^M = \left(\frac{z^2}{L^2}\right)\bar{P}^M. \tag{II.134}$$

This is the momentum in the AdS coordinates, but a more physical quantity is the momentum P^M_{phys} seen in the frame of a free falling observer at

rest at z: $P^M_{phys} = P^M(L/z) = \bar{P}^M(z/L)$. Viewing the particle as a wave, the increase in momentum is simply due to the homogeneous contraction L/z of all lengths when moving down the AdS throat, see Fig. (9). Moving to larger z is like going backward in time in Friedmann-Robertson-Walker (FRW) cosmology. This analogy is not accidental as both AdS and FRW have conformally flat metrics. Now, consider two particles starting with momenta $\bar{P}^M_1$ and $\bar{P}^M_2$ at the Planck brane. If they collide after falling at a bulk position z, the center of mass energy of the collision will be $s(z) = -2P^M_1 P^N_2 g_{MN} = -2\bar{P}^M_1 \bar{P}^N_2 \eta_{MN}(z/L)^2 = s(0)(z/L)^2$. Then by starting by a sub-Planckian energy $\sqrt{s}(0) < M_5$ we can produce super-Planckian collisions if z is large enough. Of course this also means that a Planck brane observer must wait a relatively long time $T > z$ in order to observe this collision. In a moment we will see that this time delay is the central point to discuss "strong coupling" at the quantum level. Notice also that in the case of the compact RS model z is bounded by z_1, so that for $\sqrt{s(0)} < M_5 L/z_1$ the Planck mass is never exceeded in the bulk. Once again we have found that the critical energy corresponds to the redshifted Planck mass M_5'.

Consider now the limit $z_1 \to \infty$. At $z = \infty$ the metric of 4D slices vanishes, similarly to what happens to g_{00} in the Schwarzschild solution in GR. Also, at the point $z = \infty$ there is a horizon, the AdS horizon. In fact a particle falling from the Planck brane takes an infinite Planck brane time to reach $z = \infty$, but the proper time experienced by the particle is finite $\tau = \pi L/2$. This is completely analogous to the Schwarzschild case. The model so defined is named RSII [44], so that the model with two branes we have considered so far is named RSI. If we assume that the SM is instead localized at the Planck brane (and thus give up explaining the hierarchy by redshifts) the RSII model represents an "alternative to compactification". This is because, although the 5th dimension is non-compact, there is still a normalizable 4-dimensional graviton, see eq. (II.129), dominating the IR behaviour of gravity. Eq. (II.129) should be contrasted to the flat case, in which $||\psi_0||$ grows with the radius R, so that the zero mode decouples in the infinite volume limit.

The RSII model seems a very interesting way to view 4D gravity. In this model however a particle falling from the Planck brane can undergo virtually infinite redshift before colliding. Then if we blindly applied the above definition of UV cut-off we would conclude that RSII is not a viable effective field theory! And we would be wrong. The point is that the notion of "which energy is Planckian" must be a local one. As we have already stated, a Planck brane observer, while working with initial states that have $\sqrt{s} \ll M_5$, must set up an experiment that probes deep into the 5D bulk in order to see quantum gravity effects. An instructive

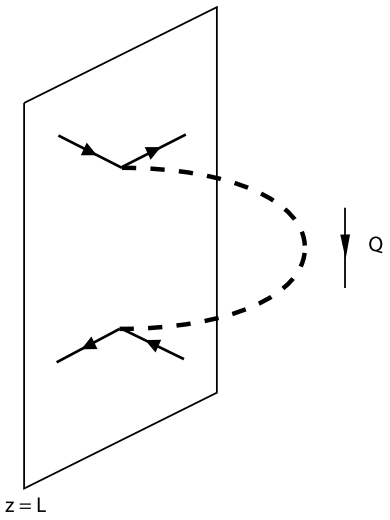

Figure 10. Amplitude for the scattering of fermions localized on the Planck brane via t-channel 5D graviton exchange.

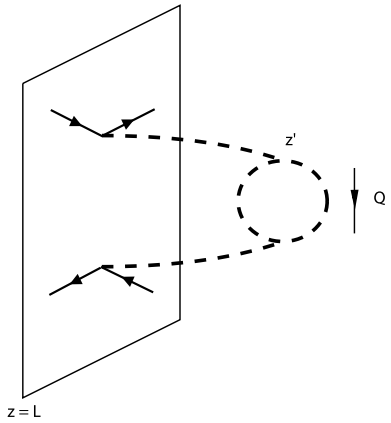

Figure 11. Leading quantum gravity correction to the process of Fig. (10) .

way to understand how things work is to consider a scattering process induced by graviton exchange among particles localized on the Planck brane. Consider first a $t-$channel process at fixed angle, for which the exchanged graviton has virtuality $Q = \sqrt{-t} \sim \sqrt{s}$. A particle with virtuality Q can exist only for a time $\sim 1/Q$, so that by causality it probes the bulk no further than $z_Q \sim 1/Q$. The virtual momentum is then only blueshifted to

$$Q(z) = Q\frac{z}{L} \lesssim Q\frac{z_Q}{L} \sim \frac{1}{L} \tag{II.135}$$

so that perturbativity of the process is maintained as long as $1/L \ll M_5$. But this was our original requirement in order to trust the RS solution. By the above simple argument we expect the leading gravitational loop corrections to the above process to be controlled by $Q(z_Q)/M_5 = 1/(M_5 L)$ independent of the scale of the kinematical parameters s and $t = Q^2$. This result is remarkable: the leading gravitational corrections are scale invariant. This does not happen by chance and is related to the possibility to interpret the RSI and RSII models according to the AdS/CFT correspondence [46, 47, 36]. As shown in Fig. (11), we can make our arguments slightly more concrete by considering the 1-loop correction to our $t-$channel process. The crucial remark [47] is that the

brane to bulk propagator (working in momentum space along the brane and in position space along the 5th dimension) behaves like

$$G(z = z_0, z', Q) \propto e^{-Q(z'-z_0)} \tag{II.136}$$

while the local loop expansion parameter is $[Q(z')]^3/M_5^3$. Taking the brane to bulk propagator into account, the 1-loop correction diagram will be of order

$$e^{-Q(z'-L)} \left(\frac{Qz'}{M_5 L}\right)^3 < \frac{1}{(M_5 L)^3} \tag{II.137}$$

consistent with our previous result. Notice that both the above inequality and the previous argument apply only for $Q < 1/L$. When $Q > 1/L$ curvature is unimportant (the relevant length scale $1/Q$ is short) so that the loop expansion parameter is just that of flat space $\sim Q^3/M_5^3$. We conclude that a Planck brane observer studying t-channel graviton exchange sees quantum gravity becoming important at the most obvious scale: M_5. By the same argument, we would also deduce that an observer sitting on a probe brane at position z sees gravity becoming strong at a local scale $M_5(z) = M_5 z/L$.

The situation for s-channel processes is somewhat different. Here with enough energy resolution one could produce, even on the Planck brane, individual KK-modes and measure their widths, being thus able to decide whether there is strong dynamics. The required energy resolution $\Delta E \lesssim 1/z_1$, corresponds, by the indetermination principle, to a time scale $T > z_1$ during which a signal can travel from the Planck to TeV brane. To be more definite consider the annihilation process $\lambda\bar{\lambda} \to \chi\bar{\chi}$ for fermions living on the Planck brane. The $s-$channel amplitude is

$$\begin{aligned} A(s) &= \frac{1}{4} T^{\mu\nu}(\text{in}) T^{\rho\sigma}(\text{out}) \langle h_{\mu\nu} h_{\rho\sigma} \rangle \\ &= \frac{1}{4} T^{\mu\nu}(\text{in}) T^{\rho\sigma}(\text{out}) \\ &\quad \sum_n \frac{|\hat{\psi}^{(n)}(z = L)|^2}{M_5^3 L} \frac{\Pi_{\mu\nu\rho\sigma}(m_n)(s)}{s - m_n^2 + i m_n \Gamma_n} \end{aligned} \tag{II.138}$$

where $\Pi(m_n)$ is the $J = 2$ projection operator defined in eq. (I.47) and where for simplicity we have neglected the radion contribution. The wave function at the Planck brane is $|\hat{\psi}^{(n)}(z = L)|^2 \sim m_n(L^2/z_1)$ for the massive modes, but for the zero mode we have $|\hat{\psi}^{(0)}(z = L)|^2 \sim 1$. The lighter massive modes, being localized away from the Planck brane, couple much more weakly than the zero mode. In this respect, the production of these modes is very suppressed. However even with this tiny coupling one could in principle study the production of individual

resonances and measure their width. In fact we do not want to stress to much the wave function suppression, as it is a specific feature of the graviton KK. For instance, in the case of a bulk vector this suppression is practically absent [45]. The point we want to stress here concerns instead the energy resolution of the experiment. In order to proceed we need the explicit expression of the brane to brane propagator as a sum over massive KK , 4D graviton and radion contributions

$$M_5^3\langle h_{\mu\nu}h_{\rho\sigma}\rangle = A(s)\Pi^{m\neq 0}_{\mu\nu\rho\sigma} + \frac{z_1^2}{z_1^2-L^2}\frac{\Pi^{m=0}_{\mu\nu\rho\sigma}}{sL} + \frac{L^2}{z_1^2-L^2}\frac{\eta_{\mu\nu}\eta_{\rho\sigma}}{6sL} \quad \text{(II.139)}$$

where

$$\Pi^{m\neq 0}_{\mu\nu\rho\sigma} = \frac{1}{2}\left(\eta_{\mu\rho}\eta_{\sigma\nu}+\eta_{\mu\sigma}\eta_{\rho\nu}\right) - \frac{1}{3}\eta_{\mu\nu}\eta_{\rho\sigma} \quad \text{(II.140)}$$

is the massive spin 2 projector of eq. (I.47) after eliminating the irrelevant longitudinal parts, while $\Pi^{m=0}_{\mu\nu\rho\sigma}$ is the massless projector given in eq. (I.50). Notice that the radion contribution, the third term, is suppressed by a power L^2/z_1^2. The massive KK amplitude is

$$\begin{aligned} A(s) &= \left\{\frac{1}{2\sqrt{s}}\frac{Y_1(\sqrt{s}z_1)J_2(\sqrt{s}L)-J_1(\sqrt{s}z_1)Y_2(\sqrt{s}L)}{Y_1(\sqrt{s}z_1)J_1(\sqrt{s}L)-J_1(\sqrt{s}z_1)Y_1(\sqrt{s}L)} - \frac{1}{sL}\frac{z_1^2}{z_1^2-L^2}\right\} \\ &\equiv \frac{F(s)}{2\sqrt{s}} - \frac{1}{sL}\frac{z_1^2}{z_1^2-L^2}. \end{aligned} \quad \text{(II.141)}$$

Notice that $A(s)$, and $F(s)$ have poles on the real positive axis in correspondence with the massive KK masses. However $A(s)$ does not have a pole at $s=0$, while $F(s)$ does. For large z_1, on the positive real s axis $A(s)$ is a complicated oscillating function with narrowly spaced poles. However when s is continued into the complex plane all this structure gives way to the much simpler euclidean behaviour. By giving $\sqrt{s}$ a positive finite imaginary part $\sqrt{s}\equiv q_R+iq_I$ and by using the asymptotic expansion for Bessel functions we find

$$\begin{aligned} F(s) &= \frac{J_2(\sqrt{s}L)+iY_2(\sqrt{s}L)}{J_1(\sqrt{s}L)+iY_1(\sqrt{s}L)}\left(1+O(e^{-2q_Iz_1})\right) \\ &= \frac{H_2^{(1)}(\sqrt{s}L)}{H_1^{(1)}(\sqrt{s}L)}\left(1+O(e^{-2q_Iz_1})\right) \end{aligned} \quad \text{(II.142)}$$

where we employed the definition of the Hankel functions. Substituting this result into eq. (II.141) we find, up to $O(e^{-2q_Iz_1})$ terms

$$\langle h_{\mu\nu}h_{\rho\sigma}\rangle(\sqrt{s}=q_R+iq_I) = -\frac{H_2^{(1)}(\sqrt{s}L)}{2\sqrt{s}H_1^{(1)}(\sqrt{s}L)}\Pi^{m\neq 0}_{\mu\nu\rho\sigma} - \frac{\eta_{\mu\nu}\eta_{\rho\sigma}}{6sL}. \quad \text{(II.143)}$$

The important point about this result is that all the dependence on powers of z_1 has disappeared. Now, the point is that if we use initial states that have an energy spread $(\Delta s)/\sqrt{s} >> 1/z_1$ (i.e. smearing this amplitude with a wave function spread over Δs) is practically equivalent to considering the amplitude at a complex point with $\mathrm{Im}(\sqrt{s}) \sim i\Delta s/\sqrt{s}$. For such spread states all the physics of the KK-modes, including their detailed strong dynamics is exponentially suppressed, and thus practically inaccessible. This result is quite analogous to the well known example of $e^+e^- \to hadrons$. In that case, the behaviour of the cross section as a function of the energy can be very complicated by the presence of the resonances, indicating that a perturbative QCD computation is not adequate. For instance this is the case near the bottom quark threshold. However by averaging the cross section over $\Delta s/\sqrt{s} >> \Lambda_{QCD}$ one obtains an observable which can be reliably computed in perturbation theory in terms of the production of quarks and gluons. This is the so-called parton-hadron duality.

To conclude we want to apply eq. (II.143) in the fully euclidean region $\sqrt{s} = iq$ to compute the gravitational potential induced on the Planck brane by a source on it. Notice that this is the exact propagator for RSII, while for RSI it applies only for $q > 1/z_1$. We are interested in the long distance behaviour $qL \ll 1$, so we use the Bessel function expansion at small argument.

We find

$$\langle h_{\mu\nu} h_{\rho\sigma} \rangle (\sqrt{s} = +iq) = \frac{1}{M_5^3 L} \left\{ \frac{\Pi^{m=0}_{\mu\nu\rho\sigma}}{q^2} + \Pi^{m\neq 0}_{\mu\nu\rho\sigma} \frac{L^2}{2} \ln q^2 + O(q^2) \right\} \tag{II.144}$$

so that the Newtonian potential induced by a body of mass m is

$$V(r) = \frac{G_N m}{r} \left(1 + \frac{2L^2}{3r^2} + \dots \right). \tag{II.145}$$

This formula applies in the long distance regime $r \gg L$. A good exercise is to check that at a distance $r \ll L$, the potential goes back to the 5D behaviour $V \sim 1/r^2$. Then RSII can be a viable alternative to compactification for $L \lesssim 100\mu$m.

References

[1] J. Polchinski, "TASI lectures on D-branes," arXiv:hep-th/9611050.

[2] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B429**, 263 (1998) [hep-ph/9803315].

[3] L. Randall and R. Sundrum, Phys. Rev. Lett. 83 (1999) 3370 [hep-ph/9905221].

[4] see for instance R. Rajaraman, *Solitons and instantons : an introduction to solitons and instantons in quantum field theory*, North-Holland, 1982.

[5] V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **125** (1983) 136.

[6] G. R. Dvali and M. A. Shifman, Phys. Lett. B **396**, 64 (1997) [Erratum-ibid. B **407**, 452 (1997)] [arXiv:hep-th/9612128].

[7] R. Sundrum, Phys. Rev. D **59**, 085009 (1999) [arXiv:hep-ph/9805471].

[8] I. Low and A. V. Manohar, Phys. Rev. Lett. **88**, 101602 (2002) [arXiv:hep-th/0110285].

[9] D. B. Kaplan, arXiv:nucl-th/9506035; A. V. Manohar, arXiv:hep-ph/9606222; I. Z. Rothstein, arXiv:hep-ph/0308266.

[10] J. F. Donoghue, Phys. Rev. D **50**, 3874 (1994) [arXiv:gr-qc/9405057].

[11] J. Polchinski, Nucl. Phys. B **231** 269 (1984).

[12] S. Weinberg, Phys. Lett. B **91**, 51 (1980).

[13] T. Kaluza, Sitzungsber. Preuss Akad. Wiss. Berlin (Math. Phys.) **1921**, 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926).

[14] M. Fierz and W. Pauli, Proc. Roy. Soc. Lond. A **173** (1939) 211.

[15] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **544**, 3 (1999); E. A. Mirabelli, M. Perelstein and M. E. Peskin, Phys. Rev. Lett. **82**, 2236 (1999);T. Han, J. D. Lykken and R. J. Zhang, Phys. Rev. D **59**, 105006 (1999); J. L. Hewett, Phys. Rev. Lett. **82**, 4765 (1999).

[16] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B436**, 257 (1998) [hep-ph/9804398].

[17] see for instance G. Aldazabal, S. Franco, L. E. Ibanez, R. Rabadan and A. M. Uranga, JHEP **0102**, 047 (2001) [arXiv:hep-ph/0011132];G. Aldazabal, L. E. Ibanez, F. Quevedo and A. M. Uranga, JHEP **0008**, 002 (2000) [arXiv:hep-th/0005067]; I. Antoniadis, arXiv:hep-th/0102202.

[18] N. Arkani-Hamed, S. Dimopoulos and J. March-Russell, Phys. Rev. D **63**, 064020 (2001)

[19] N. Arkani-Hamed, L. J. Hall, D. R. Smith and N. Weiner, Phys. Rev. D **62**, 105002 (2000) [arXiv:hep-ph/9912453].

[20] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Rev. **D59**, 086004 (1999) [hep-ph/9807344].

[21] E. G. Adelberger [EOT-WASH Group Collaboration], arXiv:hep-ex/0202008.

[22] J. Chiaverini, S. J. Smullin, A. A. Geraci, D. M. Weld and A. Kapitulnik, Phys. Rev. Lett. **90**, 151101 (2003)

[23] J. C. Long, H. W. Chan, A. B. Churnside, E. A. Gulbis, M. C. Varney and J. C. Price, Nature **421** (2003) 922.

[24] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **595**, 250 (2001) [arXiv:hep-ph/0002178].

[25] P. Creminelli and A. Strumia, Nucl. Phys. B **596**, 125 (2001) [arXiv:hep-ph/0007267].

[26] S. Cullen, M. Perelstein and M. E. Peskin, Phys. Rev. D **62**, 055012 (2000) [arXiv:hep-ph/0001166].

[27] G. F. Giudice and A. Strumia, Nucl. Phys. B **663**, 377 (2003) [arXiv:hep-ph/0301232].

[28] S. B. Giddings and S. Thomas, Phys. Rev. D **65**, 056010 (2002) [arXiv:hep-ph/0106219]; S. Dimopoulos and G. Landsberg, Phys. Rev. Lett. **87**, 161602 (2001) [arXiv:hep-ph/0106295].

[29] S. Dimopoulos and R. Emparan, Phys. Lett. B **526**, 393 (2002) [arXiv:hep-ph/0108060].

[30] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **630**, 293 (2002) [arXiv:hep-ph/0112161].

[31] V. A. Rubakov, Phys. Usp. **44**, 871 (2001) [Usp. Fiz. Nauk **171**, 913 (2001)] [arXiv:hep-ph/0104152].

[32] C. Csaki, arXiv:hep-ph/0404096.

[33] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200].

[34] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, Phys. Lett. **B428** (1998) 105 [hep-th/9802109]; E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998) [hep-th/9802150].

[35] W. D. Goldberger and M. B. Wise, Phys. Rev. Lett. 83 (1999) 4922 [hep-ph/9907447].

[36] R. Rattazzi and A. Zaffaroni, JHEP **0104**, 021 (2001) [arXiv:hep-th/0012248].

[37] S. Coleman and E. Weinberg, Phys. Rev. D **7** (1973) 1888.

[38] N. Kaloper, Phys. Rev. D **60**, 123506 (1999) [arXiv:hep-th/9905210].

[39] Z. Chacko and P. J. Fox, Phys. Rev. D **64**, 024015 (2001) [arXiv:hep-th/0102023].

[40] C. Csaki, M. Graesser, C. F. Kolda and J. Terning, Phys. Lett. B **462**, 34 (1999)

[41] J. M. Cline, C. Grojean and G. Servant, Phys. Rev. Lett. **83**, 4245 (1999)

[42] C. Csaki, M. Graesser, L. Randall and J. Terning, Phys. Rev. D **62**, 045015 (2000) [arXiv:hep-ph/9911406].

[43] H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Rev. Lett. **84**, 2080 (2000) [arXiv:hep-ph/9909255].

[44] L. Randall and R. Sundrum, Phys. Rev. Lett. 83 (1999) 4690 [hep-th/9906064].

[45] A. Pomarol, Phys. Lett. B **486**, 153 (2000) [arXiv:hep-ph/9911294].

[46] J. Maldacena, unpublished; E. Witten, Talk at ITP conference *New Dimensions in Field Theory and String Theory*, Santa Barbara http://www.itp.ucsb.edu/online/susy_c99/discussion; S. S. Gubser, "AdS/CFT and gravity," Phys. Rev. D **63**, 084017 (2001) [hep-th/9912001]; H. Verlinde, "Holography and compactification," Nucl. Phys. **B580**, 264 (2000) [hep-th/9906182].

[47] N. Arkani-Hamed, M. Porrati and L. Randall, [hep-th/0012148].