The Subnuclear Series • Volume 40

Proceedings of the International School of Subnuclear Physics

FROM QUARKS AND GLUONS TO QUANTUM GRAVITY

Edited by

Antonino Zichichi

World Scientific

The Subnuclear Series • Volume 40

Proceedings of the International School of Subnuclear Physics

FROM QUARKS AND GLUONS TO QUANTUM GRAVITY

THE SUBNUCLEAR SERIES

Series Editor: ANTONINO ZICHICHI, *European Physical Society, Geneva, Switzerland*

1. 1963 STRONG, ELECTROMAGNETIC, AND WEAK INTERACTIONS
2. 1964 SYMMETRIES IN ELEMENTARY PARTICLE PHYSICS
3. 1965 RECENT DEVELOPMENTS IN PARTICLE SYMMETRIES
4. 1966 STRONG AND WEAK INTERACTIONS
5. 1967 HADRONS AND THEIR INTERACTIONS
6. 1968 THEORY AND PHENOMENOLOGY IN PARTICLE PHYSICS
7. 1969 SUBNUCLEAR PHENOMENA
8. 1970 ELEMENTARY PROCESSES AT HIGH ENERGY
9. 1971 PROPERTIES OF THE FUNDAMENTAL INTERACTIONS
10. 1972 HIGHLIGHTS IN PARTICLE PHYSICS
11. 1973 LAWS OF HADRONIC MATTER
12. 1974 LEPTON AND HADRON STRUCTURE
13. 1975 NEW PHENOMENA IN SUBNUCLEAR PHYSICS
14. 1976 UNDERSTANDING THE FUNDAMENTAL CONSTITUENTS OF MATTER
15. 1977 THE WHYS OF SUBNUCLEAR PHYSICS
16. 1978 THE NEW ASPECTS OF SUBNUCLEAR PHYSICS
17. 1979 POINTLIKE STRUCTURES INSIDE AND OUTSIDE HADRONS
18. 1980 THE HIGH-ENERGY LIMIT
19. 1981 THE UNITY OF THE FUNDAMENTAL INTERACTIONS
20. 1982 GAUGE INTERACTIONS: Theory and Experiment
21. 1983 HOW FAR ARE WE FROM THE GAUGE FORCES?
22. 1984 QUARKS, LEPTONS, AND THEIR CONSTITUENTS
23. 1985 OLD AND NEW FORCES OF NATURE
24. 1986 THE SUPERWORLD I
25. 1987 THE SUPERWORLD II
26. 1988 THE SUPERWORLD III
27. 1989 THE CHALLENGING QUESTIONS
28. 1990 PHYSICS UP TO 200 TeV
29. 1991 PHYSICS AT THE HIGHEST ENERGY AND LUMINOSITY: To Understand the Origin of Mass
30. 1992 FROM SUPERSTRINGS TO THE REAL SUPERWORLD
31. 1993 FROM SUPERSYMMETRY TO THE ORIGIN OF SPACE–TIME
32. 1994 FROM SUPERSTRING TO PRESENT-DAY PHYSICS
33. 1995 VACUUM AND VACUA: The Physics of Nothing
34. 1996 EFFECTIVE THEORIES AND FUNDAMENTAL INTERACTIONS
35. 1997 HIGHLIGHTS OF SUBNUCLEAR PHYSICS: 50 Years Later
36. 1998 FROM THE PLANCK LENGTH TO THE HUBBLE RADIUS
37. 1999 BASICS AND HIGHLIGHTS IN FUNDAMENTAL PHYSICS
38. 2000 THEORY AND EXPERIMENT HEADING FOR NEW PHYSICS
39. 2001 NEW FIELDS AND STRINGS IN SUBNUCLEAR PHYSICS
40. 2002 FROM QUARKS AND GLUONS TO QUANTUM GRAVITY

Volume 1 was published by W. A. Benjamin, Inc., New York; 2–8 and 11–12 by Academic Press, New York and London; 9–10 by Editrice Compositori, Bologna; 13–29 by Plenum Press, New York and London; 30–40 by World Scientific, Singapore.

The Subnuclear Series • Volume 40

Proceedings of the International School of Subnuclear Physics

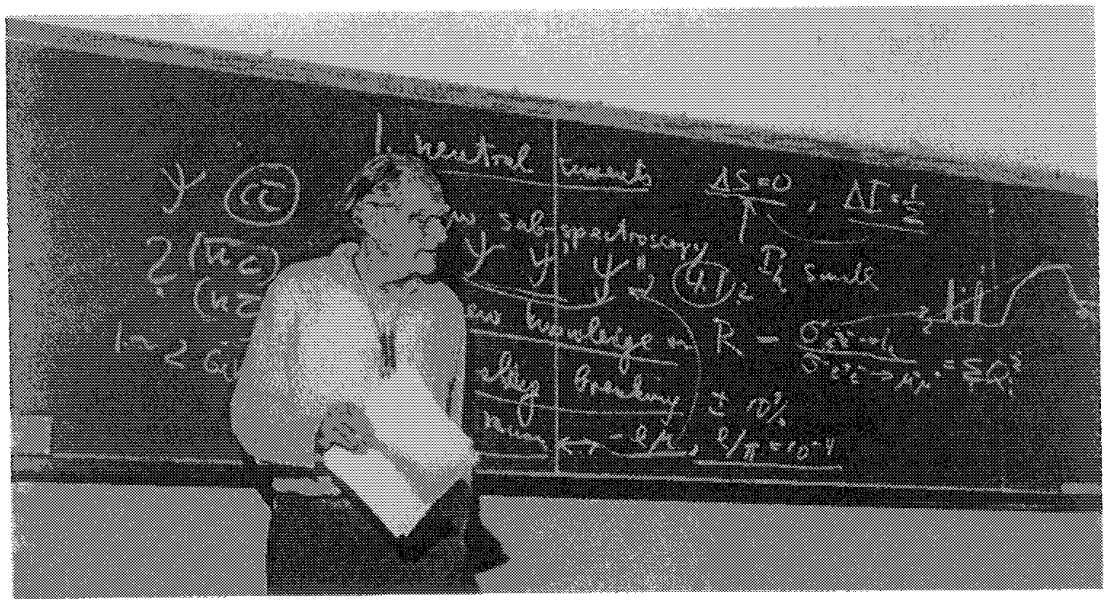
FROM QUARKS AND GLUONS TO QUANTUM GRAVITY

Edited by

Antonino Zichichi

European Physical Society
Geneva, Switzerland

World Scientific
New Jersey • London • Singapore • Hong Kong



Victor Frederick Weisskopf

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

International School of Subnuclear Physics (40th : 2002 : Erice, Italy)

From quarks and gluons to quantum gravity : proceedings of the International School of Subnuclear Physics / edited by Antonino Zichichi

p. cm. -- (The subnuclear series ; v. 40)

ISBN 9812386130

1. Quantum gravity--Congresses. 2. Gluons--Congresses. 3. Quarks -- Congresses. 4. Gauge fields (Physics)--Congresses. I. Zichichi, Antonino. II. Title. III. Series

QC178 .I575 2002

2003065685

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2003 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Printed in Singapore.

PREFACE

This volume is dedicated to the memory of Victor Frederick Weisskopf, who is a founder, together with John Stewart Bell, Patrick Maynard Stuart Blackett and Isidor Isaac Rabi of the «Ettore Majorana» Centre for Scientific Culture, this School being the first of its 114 Schools now in existence.

I was a young fellow when Europe was launching its first attempts towards the new era after the Second World War. This new era needed Science as a driving force to link together the new generations of physicists in an unprecedented challenge: to build the basic roots of scientific research in a new spirit. The spirit of European unity. For this to be accomplished, an Institution was needed where the young generations of scientists could work together. The Institution was a necessary, but not a sufficient, condition to accomplish this unprecedented task. Europe was very fortunate when Weisskopf accepted the task to take care of its first attempt towards excellency in Physics: CERN.

What makes Victor F. Weisskopf unique in the 20th century is his being a great scientist and an exceptional mentor who was endowed with a nearly unmatchable humaneness. Scientific Europe owes him an enormous debt. CERN (European Subnuclear Research Centre) had him as a scientific, moral and effective leader in the crucial years of its younger existence, from 1961 to 1965. During those years, CERN was the first European scientific enterprise to find itself in competition with the colossal USA. His responsibility as Director General of CERN was decisive for creating that which today is known throughout the world as the "spirit of CERN", which means scientific excellence. He has encouraged, inspired and guided many fellows all over Europe. In Subnuclear physics Europe is top rank, thanks to the great Weisskopf.

During August/September 2002, a group of 78 physicists from 50 laboratories in 17 countries met in Erice to participate in the 40th Course of the International School of Subnuclear Physics. The countries represented by the participants were: Austria, Belgium, Czechoslovakia, France, Germany, India, Israel, Italy, Japan, the Netherlands, Norway, Poland, Romania, Russia, Spain, Sweden, Switzerland, United Kingdom, and the United States of America.

The School was sponsored by the Academies of Sciences of Estonia, Georgia, Lithuania, Russia and Ukraine; the Chinese Academy of Sciences; the Commission of the European Communities; the European

Physical Society; the Italian Ministry of Education, University and Scientific Research; the Sicilian Regional Government; the Weizmann Institute of Science; the World Federation of Scientists and the World Laboratory.

The purpose of the School was to focus attention on the theoretical and phenomenological developments in Gauge Theories, as well as in all the other sectors of Subnuclear Physics. Experimental highlights from the most relevant sources of new data were presented and discussed, including the latest news form theoretical developments in quantizing the gravitational forces, as reported in the contents.

An original feature of the School, introduced in 1996, is a series of special sessions devoted to "New Talents". This is a serious problem in Experimental Physics where collaborations count several hundreds of participants and it is almost impossible for young fellows to be known. Even if with much less emphasis the problem exists also in Theoretical Physics. So we decided to offer the young fellows a possibility to let them be known. 26 "new talents" were invited to present a paper, followed by a discussion. Three were given an award: one for the best presentation; one for an original theoretical work; and one for an original experimental work. These special sessions devoted to New Talents represent the projection of Subnuclear Physics on the axis of the young generation.

As every year, the discussion sessions have been the focal point of the School's activity.

During the organization and the running of this year's Course, I enjoyed the collaboration of my colleague and friend, Gerardus 't Hooft, who shared with me the Directorship of the Course. I would like to thank him, together with the group of invited scientists and all the people who contributed to the success of this year's Course.

I hope the reader will enjoy the book as much as the students attending the lectures and discussion sessions. Thanks to the work of the Scientific Secretaries, the discussions have been reproduced as faithfully as possible. At various stages of my work I have enjoyed the collaboration of many friends whose contributions have been extremely important for the School and are highly appreciated. I thank them most warmly. A final acknowledgement to all those in Erice, Bologna and Geneva, who have helped me on so many occasions and to whom I feel very indebted.

Antonino Zichichi
Geneva, October 2002

CONTENTS

Mini-Courses on Basics

Lattice Field Theory and SU(N) Gauge Theories 1
M.J. Teper

Symmetries and Quasi-Particles in Hot QCD 44
C.P. Korthals Altes

Physics of QCD Instantons 105
E.V. Shuryak

Confinement and Duality 154
M.J. Strassler

Probing Grand Unification Through Neutrino Oscillations, Leptogenesis and Proton Decay 194
J.C. Pati

Status of Super String Theory 237
E. Verlinde

Perturbative Quantum Gravity 249
G. 't Hooft

Proton Structure and Its Flavor Decomposition 270
H. Abramowicz

Experimental Highlights

Selected Highlights from the First Heavy Ion Runs at RHIC 304
T.J. Hallman

Highlights from Gran Sasso 313
A. Bettini

Experimental Highlights from Super-Kamiokande 348
Y. Totsuka

The Fermilab Experimental Physics Program 384
R. Tschirhart

Special Sessions for New Talents

Interpretation of the Search for Neutral Higgs Bosons at OPAL in a CP-Violating MSSM Scenario 391
P. Bechtle

Application of the Large-N_c Limit to a Chiral Lagrangian with Resonances 402
O. Catà

Towards the Finite Temperature Gluon Propagator in Landau Gauge Yang–Mills Theory 411
A. Maas

Hermes Measurements of the Nucleon Spin Structure 420
J. Wendland

Closing Ceremony

Prizes and Scholarships 429

Participants 432

Lattice Field Theory and SU(N) Gauge Theories

Michael J. Teper

Theoretical Physics, University of Oxford, 1 Keble Road,
Oxford OX1 3NP, UK

1 Introduction

In these two lectures I will introduce lattice field theory by applying it to some specific problems that are of physical interest. My approach to the lattice is as a tool for calculating the continuum properties of physically relevant field theories. I will not dwell on what one can learn about the statistical mechanics of lattice systems from lattice Monte Carlo calculations, or indeed what one can learn about the structure of field theories through their lattice regularisation. There is much of interest down that route, and for that I refer you to some of the very good books available [1, 2, 3, 4].

The physics I will focus upon is that of QCD and, more specifically, that of 'glueballs'. These are bound states composed entirely of the gluons. If one removes quarks from QCD, one is left with an SU(3) non-Abelian gauge theory whose mass spectrum is composed of glueballs (since all we have are the gauge fields). Much of the interesting non-perturbative physics of QCD – e.g. asymptotic freedom, infrared slavery, confinement – is driven by the gauge fields and is present in the pure gauge theory. The quarks of QCD seem to play a spectator role for many purposes. Thus one might expect the glueball mass spectrum of the SU(3) gauge theory to tell us something about 'glueballs' in QCD and about the place of such states in the experimentally determined spectrum of the strong interactions. I will describe how one calculates the mass spectrum of the continuum SU(3) gauge theory and what one finds for that spectrum. These are (mostly) not new results and they have had time to feed into the phenomenology of glueball searches and we are now in the very interesting situation that we have quite convincing evidence for the place of the lightest glueball in the spectrum of hadrons. I

will briefly describe this very interesting development.

Having illustrated the use of lattice calculations in the above context, I will then move on to the major topic of these lectures: what is the physics of SU(N) gauge theories in the limit $N \to \infty$? In particular I will focus upon the crucial question: is

$$SU(3) \simeq SU(\infty) \ ? \tag{1}$$

Having answered 'yes' to the above question (primarily by a comparison of the low-lieing glueball mass spectrum) I will, somewhat paradoxically, focus on some quantities that are very different in SU(3) and SU($N \geq 4$) gauge theories. For $N = 3$ there is only one confining string that is stable – the one that joins distant charges in the fundamental representation, such as (anti)quarks. For $N = 4$ and larger there are new stable strings, called k-strings, which are (essentially) strings that connect a cluster of k quarks to k antiquarks. How their string tensions vary with k and N has been a topical question in string theory approaches to QCD, as well as (potentially) shedding light on the dynamics of confinement. Recent lattice calculations have provided accurate values for some of these string tensions and I will review what we have learned. Of course it will be quite impossible for me to go into the details of all the other interesting things that lattice field theorists have been learning about SU(N) gauge theories in the last 2 or 3 years, so I will briefly summarise the state of play in some final remarks.

Before all this however, a few very elementary remarks about why lattice calculations are needed and what makes the $N \to \infty$ limit so physically interesting.

1.1 Why the lattice?

The main analytic method for calculating physics in complex field theories such as QCD is, as you all know, perturbation theory. The experimental students amonst you in particular are used to the extensive and successful use of perturbation theory, so much so that the set of non-perturbative problems must sometimes look to you to be almost of measure zero!

The problem is that perturbation theory works well for problems where the first approximation is a more-or-less sensible description of what is going on. QED is a good example. The first approximation is a free field theory of photons and electrons (say). Well, that is indeed something like the real world (rays of light, electric current, ...). Add the interaction as a perturbation and you get accurate results. QCD on the other hand is intended as the theory of the strong interactions; a world of protons, mesons and other colourless objects. The first approximation to QCD in perturbation theory is a world of free quarks and gluons, which is nothing at all like the real world of colourless hadrons. So perturbation theory is not going to get you very far in trying to understand this world. This is a non-perturbative problem. Lattice field theory provides, as we shall see below, a framework within which a 'brute force' numerical solution of the problem can be attempted.

Of course, as you know, because of asymptotic freedom there are contexts in which QCD looks roughly like a world of free quarks and gluons. In particular this is so if we look at the theory on distance scales very small compared to the natural scale in the theory which is $O(1\text{fm})$. The experimentalists in the audience are mainly interested in 'new' physics and look for signals of that at very short distances. Here QCD can be treated perturbatively, and so sometimes you can be left with the impression that perturbation theory is all you need. But if what you are interested in is QCD *per se* then this impression is very misleading; nearly all the interesting problems are non-perturbative.

1.2 Why SU($N \to \infty$)?

When we try to solve field theories it is useful to try and find a 'neighbouring' field theory that one can analyse more simply. If the concept of 'neighbouring' can be expressed in the smallness of some parameter that characterises the difference between the theories, then one can imagine improving the result by expanding to some low order in this parameter. The classic example of this is perturbation theory where the 'neighbouring' field theory is just the free field theory, which is of course much easier to solve than the interacting field theory that we are actually interested in.

It was realised by 't Hooft a long time ago [5] that in the case of QCD it might be useful to think of it as a perturbation around the theory with $N = \infty$ colours, so that $1/N = 1/3$ is the expansion parameter that takes us from SU(∞) back to SU(3). It is not immediately obvious that SU(∞) is much simpler than SU(3) to solve. However if one makes the plausible assumption that the $N = \infty$ theory is colour confining, just like SU(3), then the fact that the number of coloured gluons is $O(N^2)$ and the number of quarks is $O(N)$, tells us a large number of interesting facts about the SU(∞) theory. For example, consider the decay of a colour singlet $q\bar{q}$ bound state decaying into two lighter $q\bar{q}$ states. If we did not have the particular constraint of confinement then there would be $O(N)$ final states of different colours. (Not $O(N^2)$ because the colour of the initial state provides a constraint.) Presumably the decay width would be O(1) in units of the mass of the bound state. If now we say that we have confinement, that means that only 1 of the N final states is in fact possible: the one where the two decay products are both colour singlets. So the decay is suppressed by $1/N$: and in the $N = \infty$ limit particles do not decay. Similarly mixing is suppressed and one has a perfect OZI rule. Now while this is different from the real world of the strong interactions, it is not so different. Decay widths are in fact modest compared to masses; there is an approximate (and mysterious) OZI rule etc. In fact many features of the real SU(3) world look like approximate reflections of exact properties of QCD with $N = \infty$ [6]. If we could demonstrate that SU(∞) is confining and that SU(3) is indeed 'close' to SU(∞), then we would have made a substantial step in understanding QCD and the world of strong interactions.

Numerical lattice methods enable us to address these questions in a direct way, and

one of my main goals in these lectures is to describe the progress that has recently been made. The questions we might wish to address include:

• Is the SU(∞) theory confining?

• Is there a smooth $N \to \infty$ limit and what is the mass spectrum in that limit?

• Is SU(3)$\simeq$SU(∞) ?

• Is the limit achieved by keeping fixed the 't Hooft coupling, $\lambda = g^2 N$, as one expects [5] from an analysis of perturbation theory to all orders?

• Are corrections $O(1/N^2)$ in the pure gauge theory – again as expected from the diagrammatic analysis [5]?

• Are there new confining strings when we go to larger N and if so, what are their string tensions?

• Are instantons suppressed exponentially in N and if so, what is the nature of any surviving topological fluctuations?

• Does chiral symmetry remainspontaneously broken as $N \to \infty$ and, if so, is the breaking driven by topology? What does this suggest about the role of instantons in chiral symmetry breaking in the real SU(3) world?

These are only some of the interesting questions one can pose. I do not have the time here to dwell upon all of them in detail, but instead will refer you to some reviews [6]. In particular the classic and influential Erice lectures by Sidney Coleman [6] should not be missed. For a collection of up-to-date reviews the recent Large-N Conference Proceedings [6] are particularly useful.

2 Lattice calculations

In this Section I shall begin with a summary of how one can use numerical methods to calculate properties of a field theory. In particular, I shall focus on the mass spectrum of SU(N) gauge theories. The method involves discretising space-time to a lattice and so I also need to discuss how one recovers the desired continuum physics of a theory from such lattice calculations. After these generalities I give an example of a mass spectrum calculation in the case of SU(3), which is the gauge group underlying the strong interactions. The results I show were performed specifically for these lectures and it is of interest to note that these physically interesting quantities were obtained in a time that translates to less than two days on an upper-end Pentium IV powered PC! This illustrates how lattice field theory now has the potential to become an everyday tool for field theorists.

2.1 Overview

We wish to calculate the spectrum of the Hamiltonian H:

$$H|n\rangle = E_n|n\rangle. \tag{2}$$

From these energies we obtain the mass spectrum. We can obtain these energies by calculating correlation functions. Let $\Phi(t)$ be a functional of the fields at time t. Then we have the standard decomposition

$$\langle\Phi^\dagger(t)\Phi(0)\rangle = \langle e^{iHt}\Phi^\dagger e^{-iHt}\Phi\rangle = \langle\Phi^\dagger e^{-iHt}\Phi\rangle = \sum_n |\langle\Omega|\Phi^\dagger|n\rangle|^2 e^{-iE_n t} \tag{3}$$

obtained by inserting a complete set of energy eigenstates, $\sum_n\langle n|n\rangle$ =1, on each side of e^{-iHt}, and then using the fact that the latter operator is diagonal in energy. (Usually we do not explicitly show the vacuum state, but where we do we denote it by $|\Omega\rangle$ and for convenience, as here, we take its energy to be zero.) So if we can calculate such correlators we can obtain the energy spectrum. Such a correlator can be expressed (schematically) as a Feynman Path Integral

$$\langle\Phi^\dagger(t)\Phi(0)\rangle = \frac{1}{Z}\int\prod_{x,\mu} dA_\mu(x)\Phi^\dagger(t)\Phi(0)e^{iS}, \tag{4}$$

with $S \equiv \int dt \int d^3x\mathcal{L}$ where S and $\mathcal{L}$ are the action and Lagrangian. We do not know how to calculate such integrals analytically for non-perturbative quantities such as the mass spectrum, and we shall therefore attempt to calculate the integral numerically. However the presence of an oscillating phase factor, e^{iS}, in the integrand will destabilise a numerical approximation. Thus we choose to calculate the correlation function in Euclidean rather than Minkowski space-time i.e. $t \to -it$. Under this transformation

$$e^{iS} = e^{i\int dt\int d^3x\mathcal{L}} \to e^{-\int dt\int d^3x\mathcal{L}_\mathcal{E}} = e^{-S_E} \tag{5}$$

where the subscript E indicates that the Euclidean Lagrangian (and action) will differ from the Minkowski one by the change in metric. If the Lagrangian is even in the time components of four-vectors, then the Euclidean action will be real as suggested by the notation in eqn(5). In general this will be the case if the theory is time-reversal invariant. Thus a gauge theory with the usual Maxwell action will have S_E real. If, however, we add a θ-term then that will contribute an imaginary piece to S_E, undermining any straightforward numerical evaluation. Similar problems arise with a finite baryon number density. Here we will confine ourselves to the case where S_E is real, although gradual progress is being made in cases where this is not so.

If we calculate the correlation function in the Euclidean field theory we immediately see by substituting $t \to -it$ in eqn(3) that

$$\langle\Phi^\dagger(t)\Phi(0)\rangle_E = \langle\Phi^\dagger e^{-Ht}\Phi\rangle = \sum_n |\langle\Omega|\Phi^\dagger|n\rangle|^2 e^{-E_n t}. \tag{6}$$

Note that any state contributing to this sum must have the quantum numbers of Φ, since otherwise $\langle\Omega|\Phi^\dagger|n\rangle = 0$. So if we choose Φ to be translation invariant, i.e. $\vec{p} = 0$, and if we choose it to have specific J^{PC} quantum numbers, then the lightest non-zero

energy in eqn(6) will typically be the mass of the lightest particle with those quantum numbers. If we use $\Phi - \langle\Phi\rangle$ as our operator, so that the vacuum state does not contribute to the correlator, then it is clear from eqn(6) that at large t the correlation function decays as a simple exponential and the decay rate gives us the mass of the lightest particle with these J^{PC} quantum numbers. From now on I will assume all operators are vacuum subtracted in this way.

In an SU(N) gauge theory the Euclidean correlator can be calculated from the corresponding Euclidean Feynman Path Integral (EFPI):

$$\langle\Phi^\dagger(t)\Phi(0)\rangle_E = \frac{1}{Z}\int \prod_{x,\mu} dA_\mu(x)\Phi^\dagger(t)\Phi(0)e^{-\frac{1}{g^2}\int Tr\{F_{\mu\nu}^2\}d^4x} \tag{7}$$

Here Z is the same integral with the replacement $\Phi^\dagger(t)\Phi(0) \to 1$, the gauge fields $A_\mu(x)$ belong to the SU(N) Lie algebra, the $F_{\mu\nu}$ are the usual non-Abelian field strengths and we have scaled the fields so that the coupling g^2 appears as an overall multiplicative factor. (And we have not imposed any gauge fixing.)

[Aside: the expression in eqn(6) involves the same eigenvalues of the Hamiltonian and exactly the same matrix elements as appear in the expression for the Minkowski space-time correlator in eqn(3). That one can express eqn(6) as a Euclidean Feynman Path Integral is straightforward to derive (see e.g. Appendix A of [7].) A more subtle question is whether the Euclidean correlation functions produce the time-ordered Minkowski Greens functions via a straightforward analytic continuation. The answer is yes to all orders in perturbation theory but this does not exclude extra non-perturbative singularities obstructing such a continuation. (See e.g. [8].) A further interesting question is whether the lattice regularision of the Euclidean field theory, that we actually work with, does indeed correspond to a theory with a Hamiltonian. Often this is not the case but insofar as we are only interested in the continuum limit this does not matter as long as the violations of positivity occur at the ultraviolet energy scale. (See e.g. [8, 1].)]

The Path Integral involves an infinite number of degrees of freedom. For a numerical evaluation we must make the problem finite, which requires introducing ultraviolet and infrared cut-offs. The former is achieved by replacing continuous space-time by a hypercubic lattice of points, of lattice spacing a. The latter is achieved by placing this lattice on a finite (hyper)torus of volume V. We expect the theory to have a finite mass gap (i.e. the lightest mass in the spectrum to be non-zero) so once V is large compared to the corresponding Compton wavelength we would expect exponentially small corrections to the physical $V = \infty$ limit. At the same time, because the theory is renormalisible, we expect that once a is much less than the dynamically generated physical length scale, Λ, of the gauge theory, the lattice corrections to physical quantities will decrease rapidly as some power of a (provided the coupling is made to run appropriately). Thus a calculation with $V = (aL)^4$, and with $a \ll 1/\Lambda$ and $La \gg 1/\Lambda$, should give a good approximation to continuum physics. One might imagine that a 16^4 lattice with an appropriate a, might begin to satisfy these conditions.

Suppose that we wish to calculate some expectation value $\langle\Psi(A)\rangle$ where $\Psi(A)$ is some functional of the gauge fields e.g. the correlator in eqn(6). On our lattice the gauge fields will be replaced by some degrees of freedom that I will call $U_\mu(n)$ where n labels the lattice site. The lattice action I call $S_L[U]$ and the lattice version of $\Psi(A)$ I call $\Psi_L(U)$. Then

$$\langle\Psi_L(U)\rangle = \frac{1}{Z}\int\prod_{n,\mu} dU_\mu(n)\Psi_L(U)e^{-S_L[U]} \xrightarrow{a\to 0} \langle\Psi(A)\rangle \tag{8}$$

will be the case if

$$dU_\mu(n) \xrightarrow{a\to 0} dA_\mu(x = an) \tag{9}$$

$$S_L[U] \xrightarrow{a\to 0} S[A] \tag{10}$$

$$\Psi_L(U) \xrightarrow{a\to 0} \Psi(A) \tag{11}$$

where we label the lattice sites by the integer (or set of integeres) n. As we shall shortly see, it is not difficult to satisfy these requirements, with the variables $U_\mu(n)$ being SU(N) matrices assigned to the links that join neighbouring lattice sites. (That is to say, $U_\mu(n)$ is assigned to the link that leaves the site n in the positive μ direction.)

The lattice FPI in eqn(8) is no easier to calculate analytically than the original continuum version. However because the number of integrations is now finite, we can attempt a numerical evaluation. The number of integrations is large and so the natural method to use is the (Markovian) Monte Carlo. The Monte Carlo generates 'points' in the integration space. Each such 'point' is an explicit lattice gauge field i.e. an SU(N) matrix on every link of the lattice. These fields are generated with the measure

$$\mathcal{D}U = \prod_{n,\mu} dU_\mu(n)e^{-S_L[U]} \tag{12}$$

so if we generate n_c such 'points', i.e. $\{U_\mu(n); \mu = 0,..,3; n = 1,...,L^4\}^I$; $I = 1,...,n_c$, then the expectation value of Ψ will be just the average over these fields:

$$\langle\Psi_L(U)\rangle = \frac{1}{n_c}\sum_{I=1}^{n_c}\Psi_L(U^I) \pm O(\frac{1}{\sqrt{n_c}}). \tag{13}$$

I have made explicit here the statistical error which decreases as the square root of the number of field configurations - as one would expect for such a probabilistic estimate.

The above is a very rapid overview of how and why one performs lattice calculations. There are various details - in particular how the Monte Carlo works - that I will have no time to go in to. Fortunately there are now a number of good textbooks [2, 3, 1, 4] for lattice field theory which can be used to make up for the inevitable shortcomings of these lectures. In the next subsection I will give a detailed realisation of a gauge theory on the lattice and I will discuss how one relates it to the continuum theory.

2.2 From lattice to continuum

Consider the SU(N) gauge theory in 3+1 dimensions. The continuum degrees of freedom are gauge potentials, $A_\mu(x)$, that take values in the SU(N) Lie Algebra. That is to say they can be represented by $N \times N$ (anti)Hermitian matrices. If we apply a gauge transformation $G(x) \in SU(N)$ and transform

$$A_\mu(x) \to G(x)A_\mu(x)G^\dagger(x) + \frac{1}{g}G(x)\partial_\mu G^\dagger(x) \tag{14}$$

then gauge invariance (plus renormalisibility and the other usual symmetries) demands that the action have the usual Maxwell form as in eqn(7). The derivative term in eqn(14) tells us that the gauge potential $A_\mu(x)$ relates the gauge transformation at the point x to that at the 'neighbouring' point $x + \hat{\mu}dx$. (Here $\hat{\mu}$ is the unit vector in the μ direction.) Thus we expect that the lattice gauge field, $U_\mu(n)$, will need to relate the gauge transformations at neighbouring lattice sites $x = an$ and $x = an + a\hat{\mu}$, where a is the lattice spacing. That is to say, we should think of $U_\mu(n)$ as being assigned to the lattice link that leaves the site n in the positive μ direction. In the continuum theory one can construct such a variable as the path ordered exponential of the line integral,

$$U(x,y) = P\exp i\int_x^y A_\mu(x)dx_\mu, \tag{15}$$

which transforms as

$$U(x,y) \to G(x)U(x,y)G^\dagger(y) \tag{16}$$

under a gauge transformation G. Since A_μ belongs to the SU(N) Lie algebra, $U(x,y)$ will take values in the SU(N) group. Motivated by these observations we choose our lattice fields $U_\mu(n)$ to be SU(N) matrices and we define the gauge transformation by

$$U_\mu(n) \to G(n)U_\mu(n)G^\dagger(n+\hat{\mu}) \tag{17}$$

where $G(n)$ is a site-dependent SU(N) matrix, i.e. a local SU(N) gauge transformation. We now want a lattice action invariant under these gauge transformations. It is easy to see, by direct substitution, that if we have any closed path c that starts and ends at the site n, then if we multiply the group elements along c (with the convention that we use U_l when going in the positive direction of the link l and $U_l^\dagger$ when we go in the negative direction) we get an SU(N) matrix $U[c]$ which transforms as $U[c] \to G(n)U[c]G^\dagger(n)$. So if we consider $Tr\{U[c]\}$ we can use the cyclic property of the trace to bring the $G^\dagger(n)$ next to the $G(n)$, and since $GG^\dagger = 1$ for unitary matrices, we see that $Tr\{U[c]\}$ is gauge invariant and a suitable candidate for part of a lattice action.

The simplest non-trivial closed path is around the elementary square of the lattice. This is called a plaquette p. Denote by U_p the ordered product of the four U_l around p. Then a suitable lattice action is

$$S_L[U] = \frac{2N}{g^2}\sum_p\{1 - \frac{1}{N}ReTrU_p\}. \tag{18}$$

Summing over all plaquettes p ensures the action is invariant under rotations and translations. Taking the real part of the trace ensures it is $C = +$. The factor of $1/g^2$ is there because that is the factor we need to recover in the continuum limit, as in eqn(7). To see that the coefficient of $1/g^2$ should be $2N$ requires a detailed calculation. We can add any constant to the action and it will cancel from all calculations. We have chosen to add 1 into the action since that ensures that $S_L[U] \geq 0$, which is convenient and so has become conventional. It is also conventional to define $\beta \equiv 2N/g^2$. This convention risks confusion with the running coupling β-function and with the $\beta \equiv 1/kT$ of finite temperature field theory, but it is too late to change it and one can usually tell which β is being referred to by the context.

We have in eqn(18) the simplest and most local lattice action which is gauge invariant. If we take the continuum limit, $a \to 0$, the symmetries alone assure us that, barring some accidental cancellation, the leading term we pick up will be the desired continuum Maxwell term, $\propto \int TrF^2$. There will be terms of higher dimension as well, but these non-renormalisible interactions will be multiplied, for dimensional reasons, by appropriate powers of a which will ensure that their contribution on physical length scales will vanish at $a = 0$.

There remain some crucial questions. First, how do we vary a? The lattice spacing does not appear as an explicit tunable parameter in the lattice FPI, eqns(8,18). The only parameter that we can vary is the $1/g^2$ in the action. Now, g^2 is the bare coupling and like any bare coupling it can be regarded as providing a particular definition of a running coupling on the ultraviolet length scale at which the fields of the bare theory are defined. Here that scale is just a. Thus we may write $\beta = 1/g_L^2(a)$ where the L denotes that we are in some particular lattice scheme for defining the running coupling. Choosing a value for $g_L(a)$ thus corresponds to choosing a value of a (dimensional transmutation). Thus we vary a by varying $\beta \equiv 2N/g^2$.

A second crucial question is: how do we approach the continuum limit? In the case of SU(N) gauge theories it is easy to answer this question. Since these theories are asymptotically free, we know that $g_L^2(a) \to 0$ as $a \to 0$. So we can tune $a \to 0$ by tuning $\beta \equiv 2N/g^2 \to \infty$.

It is instructive to ask how one can find a continuum limit in the general case where the theory need not be asymptotically free. Suppose the theory has a mass gap. Call this lightest mass in the theory m_G. For any value of the parameters (here just β) we can calculate this mass from the Euclidean correlation functions on the lattice:

$$\langle \Phi^\dagger(t = n_0 a)\Phi(0)\rangle_E = \sum_{l=1} |c_l|^2 e^{-E_l n_0 a} \stackrel{n_0 \to \infty}{\longrightarrow} |c_1|^2 e^{-a m_G n_0} \tag{19}$$

(where we use a vacuum subtracted operator). This can be expressed as a lattice FPI and the latter can be evaluated numerically as outlined above. From such a calculation (and below I will provide an illustration) we extract m_G in lattice units i.e. am_G. Now, we want $a \to 0$. That is to say, we want $am_G \to 0$. So we simply calculate am_G for a

fine mesh of values of β and from the calculated values find any β_c such that

$$am_G \stackrel{\beta\to\beta_c}{\longrightarrow} 0. \tag{20}$$

In this way we can search and find candidate continuum limits. I say 'candidate' because $am_G \to 0$ can arise not only because $a \to 0$ but also if $m_G \to 0$. For example if the parameters are being tuned to a value where the theory acquires an exact symmetry that is spontaneously broken, leading to massless Goldstone bosons. So in practice one needs to calculate several particle masses: if $a \to 0$ then all masses am_i vanish, while in other cases only some will vanish. So while care must be taken, it is clear how one needs to proceed.

For SU(N) gauge theories the fact that the theory is asymptotically free, so that we know the continuum limit to be at $\beta = \infty$, makes the argument that much simpler. First we can easily see how the lattice action in eqn(18) becomes the continuum action in eqn(7). It is clear from eqns(8, 18) that as $\beta \to \infty$ the important values of the plaquette matrix are $\frac{1}{N}ReTrU_p = 1+O(\frac{1}{\beta})$ which means that the link matrices $U_\mu(n) \to 1$ (up to a gauge transformation). So we can write

$$U_\mu(n) = \exp\{iaA_\mu(n)\} \stackrel{\beta\to\beta_c}{\simeq} 1 + aA_\mu(n). \tag{21}$$

Here we have written the group element as an exponential of an element of the Lie algebra and the factor of a in the exponent is there to ensure that A_μ has dimensions of mass, just like the usual continuum gauge potential. If we now substitute the expression in eqn(21) for every occurrence of U_l in the plaquette action, we can show, after some algebra, that we recover the usual Maxwell action for $A_\mu(n)$ with corrections that are $O(a^2)$. This kind of operator expansion occurs on scales $\sim a$ and the fact that the theory is asympotically free then allows us to control that limit perturbatively. For small enough lattice spacings we can use the usual one-loop relation

$$g_L^2(a) = \frac{(4\pi)^2}{\frac{11N}{3}\log\{\frac{1}{a^2\Lambda_L^2}\}}. \tag{22}$$

This can be inverted to provide a as a function of $\beta \equiv 2N/g_L^2(a)$ so that we can write

$$am_G(a) = \frac{m_G(a)}{\Lambda_L} e^{-\frac{12\pi^2}{11N^2}\beta} \tag{23}$$

where the first factor on the right differs from the corresponding continuum ratio $\frac{m_G}{\Lambda_L}$ by a correction that is $O(a^2)$. Thus for small enough value of a we would expect the value of $am_G(a)$ that we extract from our numerical calculation of approporiate correlation functions to deacrease with β as shown in eqn(23). In practice it is hard to do calculations for such small values of a and we calculate mass ratios instead, as described in the next subsection.

The continuum limit $am_G \to 0$ can be trivially restated as $\xi_G/a \to \infty$, where $\xi_G \equiv 1/m_G$ is the (longest) correlation length (since the mass gap is the lightest mass). Viewing the lattice EFPI in eqn(8) as just the statistical mechanical partition function of a system on a 4 dimensional lattice, the continuum limit is thus the place where the correlation length diverges in units of the lattice spacing. That is to say, it corresponds to a 2'nd order phase transition. Conversely if we have some D-dimensional lattice model with a second order phase transition at some values of the couplings, then that is a candidate continuum limit of some field theory in $D-1$ space dimensions. What that field theory might be is not necessarily obvious. It can be inferred both from the symmetries and by investigating the behaviour of various mass ratios.

2.3 Example : the SU(3) mass spectrum

Consider the SU(3) gauge theory. To calculate its mass spectrum we calculate correlation functions of some appropriate operators ϕ. Consider the following operator:

$$\phi(n_t) = \sum_{n_s} \sum_{\mathcal{R}} ReTr\, U_c(n_t, n_s) \tag{24}$$

where c is a closed space-like curve, beginning and ending with the site (n_t, n_s) (n_t and n_s are the temporal and spatial co-ordinates repectively) and $U_c(n_t, n_s)$ is the product of link matrices around the boundary of the curve c. We take the trace to make the operator a colour singlet, because we believe SU(3) is colour confining and asymptotic states are colour singlets. We take the real part to project on $C = +$ states: we can take the imaginary part to obtain $C = -$ states. The sum over spatial sites means that we project onto states of zero momentum. The sum over $\mathcal{R}$ denotes a sum over all spatial rotations of the curve c, so that we project onto states that have maximum rotational symmetry i.e. "spin zero". So, if we call the lightest 0^{++} particle state $|G\rangle$, then we expect:

$$\langle \phi^\dagger(n_t)\phi(0)\rangle = |\langle\Omega|\phi^\dagger|G\rangle|^2 e^{-am_G n_t} + \sum_{E_i > m_G} |\langle\Omega|\phi^\dagger|i\rangle|^2 e^{-aE_i n_t} \tag{25}$$

(where, as usual, we use vacuum subtracted operators so that the vacuum does not contribute a term to the RHS). For large t the first term on the RHS dominates, the correlation function has a simple exponential fall-off, and we can extract am_G from the exponent. If we want other spins then we insert an appropriate phase factor that depends on the rotation $\mathcal{R}$. (There are ambiguities in extracting continuum spins using only the lattice symmetries, but I will not dwell on these here.)

An obvious candidate for U_c in eqn(25) is U_p, the simple plaquette operator that appears in the lattice action. It is essentially 'pointlike' and so one would expect it to have a non-zero projection onto all states, the lightest included. However for this very reason calculations with it turn out to be disappointing. Because it projects more-or-less equally on all states, its normalised projection onto any given state, the lightest

included, is very small. And it becomes smaller as $a \downarrow$ since the number of excited states increases as the cut-off is removed. Thus we have to go to very large values of n_t before the correlator in eqn(25) is dominated by the lightest state. In a numerical calculation this is fatal: the statistical error in eqn(13) is weakly dependent on n_t and so as we increase n_t it will quickly dominate over the exponentially decreasing value of the correlator, making an extraction of the lightest mass impossible.

The message from this failure is that we need operators that have a good overlap onto the state whose mass we want to calculate – in this case the lightest mass. While we don't know in advance what will work best, we do expect that a good operator will have a size of order the physical length scale (just like the particle itself) and that it should be smooth, since we are looking for the ground state. So our curve c should be like a big ball of string! This sounds forbidding to construct, but there are fast and simple iterative methods [9, 10] for achieving such operators. (See e.g. [11] for a more recent description of these techniques.) I will now show a simple example of an SU(3) calculation that uses these methods.

I choose three values of a corresponding to $\beta = 5.7, 5.9, 6.05$. The lattices are $8^4, 12^4$ and 16^4 respectively. That such lattices are large enough will only be apparent after the calculation of the masses. (Of course, I know from experience that they will be.) From asymptotic freedom we know that $a \downarrow$ as $\beta \uparrow$, and so we would expect $am_G(a) \downarrow$ as well. In Fig. 1 I show correlation functions, as evaluated using Monte Carlo. We see approximate exponentials that flatten as $\beta \uparrow$ indicating that $am_G(a) \downarrow$, exactly as expected. So far so good, but how do we extract continuum physics from all this? If we calculate two masses on the lattice, $am_1(a)$ and $am_2(a)$, then we can remove the overall lattice scale by taking the ratio of the calculated masses. Now for the plaquette action we know theoretically that the leading correction is $O(a^2)$. That is to say,

$$\frac{am_1(a)}{am_2(a)} = \frac{m_1(a)}{m_2(a)} \stackrel{a\to 0}{=} \frac{m_1(0)}{m_2(0)} + c(am)^2 \tag{26}$$

where am is some mass, which we may choose to be $am_1(a)$ or $am_2(a)$. (Note that the difference between using $am(a)$ rather than $am(0)$ in the correction term only makes a difference at $O(a^4)$, and so does not matter.) If we calculate such a mass ratio for several values of β and find that it can be well fitted with eqn(26) for all β greater than some value, then we can obtain the continuum mass ratio, $m_1(0)/m_2(0)$, from the fit.

When applying eqn(26) I will often use $m = m_2 = \sqrt{\sigma}$, where σ is the string tension. By the latter I mean the coefficient of the linear part of the confining potential which dominates at large separations between charges that are in the fundamental representation of colour (e.g. quarks):

$$V(r) \stackrel{r\to\infty}{=} \sigma r. \tag{27}$$

If one uses a potential that is the sum of Coulomb and linear terms in the calculation of the charmonium (or bottom) spectrum one finds that to reproduce the experimental

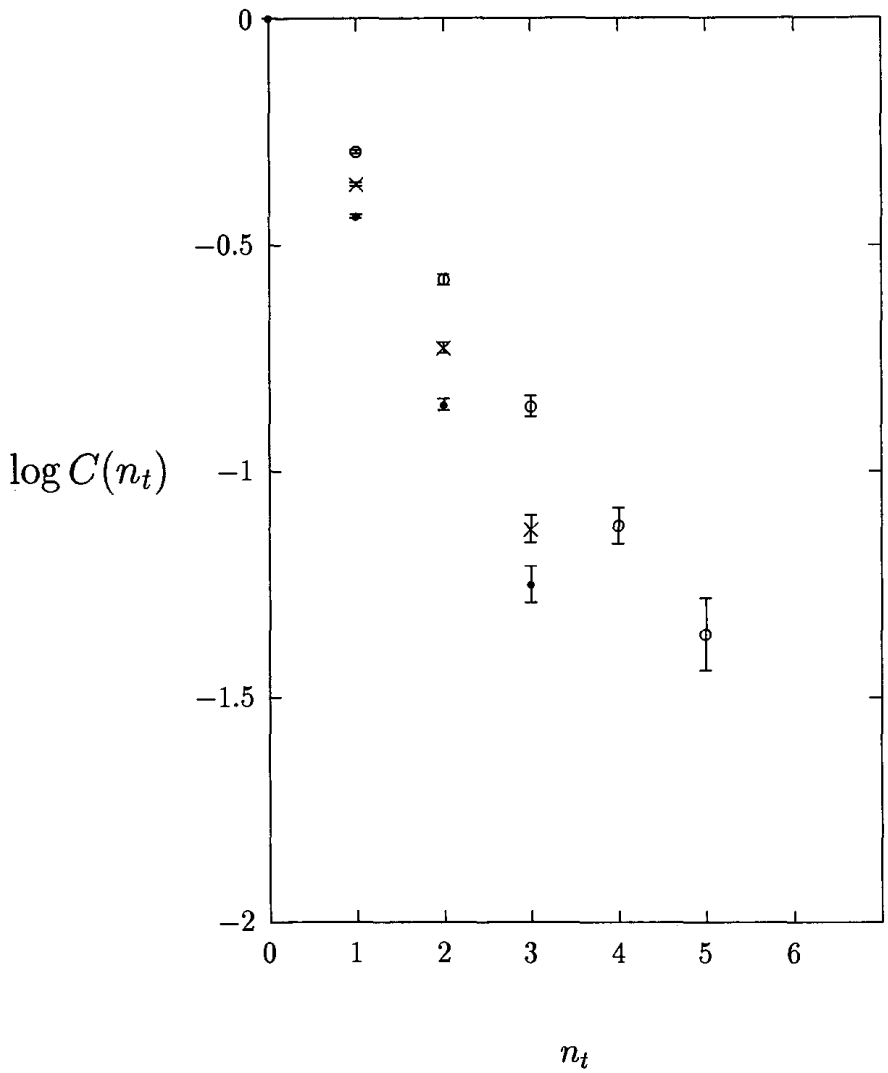


Figure 1: Correlation functions of our 'best' vacuum-subtracted $J^{PC} = 0^{++}$ glueball operators at $\beta = 5.7$, $\bullet$, $\beta = 5.9$, $\times$, and $\beta = 6.05$, $\circ$, versus the Euclidean time separation $t = an_t$.

spectrum one needs $\sigma \sim (440\mathrm{MeV})^2$. This is also close to the value that comes out if one models the states along a Regge trajectory with a simple rotating string joining the q and $\bar{q}$; a simple calculation (see e.g. [12]) gives the Regge slope to be $\alpha' = 1/2\pi\sigma$, and using a slope of $\alpha' \sim 0.88\mathrm{GeV}^{-2}$ gives a σ in this ball-park. Thus the value for σ that I shall use here is

$$\sigma \simeq (440 \pm 30\mathrm{MeV})^2. \tag{28}$$

The large error reflects both the fact that the above arguments are all rather heuristic, and, more importantly, the fact that we are in the pure gauge theory rather in full QCD, and so there is an inherent ambiguity in trying to introduce physical MeV units into the calculation. In a strict sense physical units can only be applied in the physical theory (i.e. SU(3) gauge fields with u, d, s quarks with the correct physical masses). As one moves away from the physical theory the use of 'MeV' units becomes more ambiguous and, eventually, completely misleading. The $\pm 30\mathrm{MeV}$ error in eqn(28) is intended to be some reasonable estimate [13] of this ambiguity in the pure SU(3) gauge theory.

So we calculate the mass of the lightest 0^{++} glueball (as described above) and the string tension (the details of which are described later on) at the three values of β specified above. So at each β we calculate the ratio

$$\frac{am_{0^{++}}(a)}{a\sqrt{\sigma(a)}} \equiv \frac{m_{0^{++}}(a)}{\sqrt{\sigma(a)}} \tag{29}$$

In Fig. 2 I plot this ratio against the value of $a^2\sigma(a)$ obtained at the same value of β. Since the leading correction to the continuum limit will be $\propto a^2\sigma(a)$, see eqn(26), the continuum extrapolation is a straight line, as shown in Fig. 2. The continuum limit $a = 0$ corresponds to the $a^2\sigma(a) = 0$ intercept, and so from our fit we obtain

$$m_{0^{++}} = 3.67(11)\sqrt{\sigma} \simeq 1.61(12)\mathrm{GeV} \tag{30}$$

where we use eqn(28) to express σ in physical units. Note that the error here is dominated by the error that arises from expressing σ in GeV units, so better lattice calculations will not help with that. It turns out, by performing similar calculations with other J^{PC} quantum numbers, that the lightest glueball is this one. The next lightest is the 2^+ tensor whose mass we find to be

$$m_{2^{++}} = 5.03(21)\sqrt{\sigma} \simeq 2.21(18)\mathrm{GeV}. \tag{31}$$

As I remarked earlier, I performed this calculation just before coming here. You will be interested to know that it took about 1 week on a local α workstation. This two year old work station is slower than the typical Pentium IV PC you can buy for 1000 Euros today. So the lesson is that you can obtain interesting continuum physics with the kind of computing resources that will be available in almost any Physics department. You will not, of course, be able to compete with large lattice groups using specialised

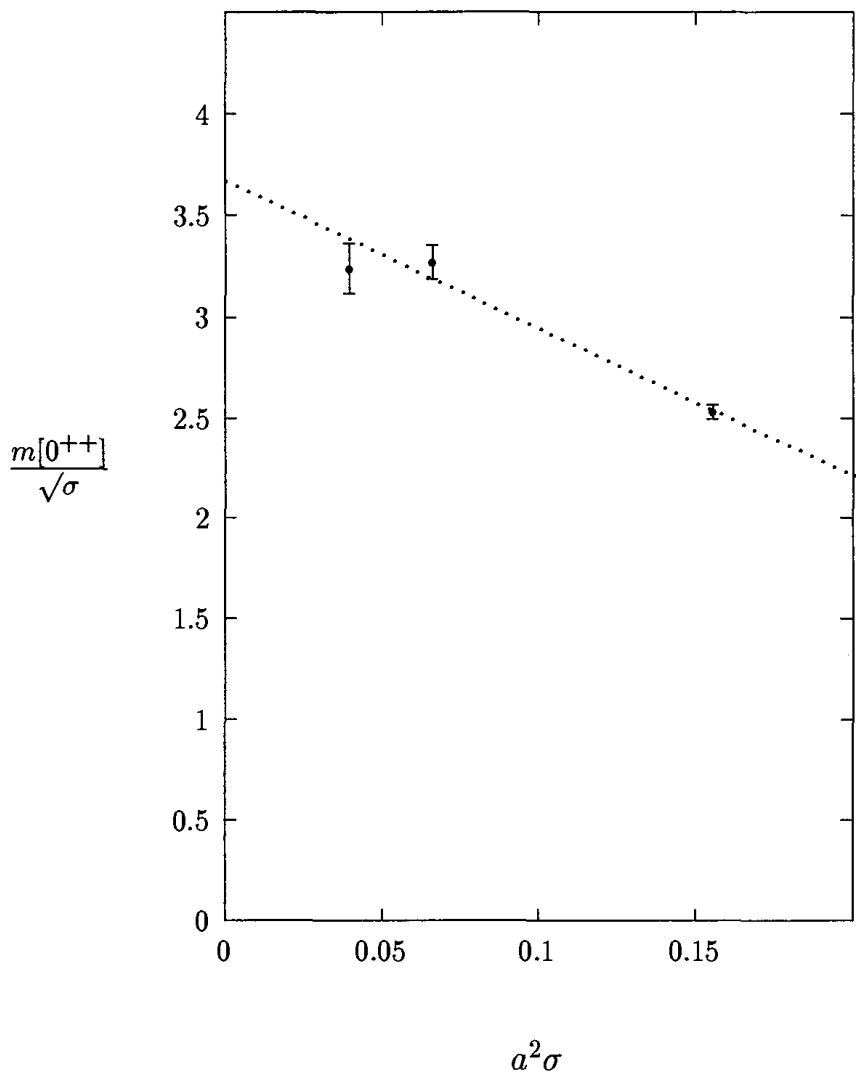


Figure 2: The lightest SU(3) $J^{PC} = 0^{++}$ glueball mass, in units of the string tension, as obtained from the correlators in Fig. 1. Plotted against the string tension in lattice units. Shown also is an extrapolation to the continuum limit.

Teraflop computers but such computers are being used for only a few problems (they are hard to program) and these problems, such as full lattice QCD with light quarks, are not ones that you could attempt on a PC anyway. But if you are interested in gauge field theories, perhaps with scalar fields, at finite as well as zero temperature (and even with light fermions as long as you keep only one fermion loop) then these problems will probably be accessible to you.

2.4 Glueballs: an aside about the 'real world'

We have just seen that the lightest two glueballs in the SU(3) gauge theory are the scalar 0^{++}, with a mass of about 1.6 GeV, and the tensor 2^{++}, with a mass of about 2.2 GeV. Before discussing the phenomenological implications of this let me say a bit more about the spectrum.

Firstly, you may be unimpressed with just the three values of β that I used to extrapolate to the continuum limit in Fig. 2. That is fair enough. So in Fig. 3 I show you a selection of values for this ratio from a number of different calculations [13]. (These are all with the same plaquette action as used above. There are other calculations using other lattice actions with similar results.) As you can see the linear fit is now much more convincing – and the fit is entirely consistent with the one in Fig. 2.

As for the rest of the glueball spectrum, the lightest 0^{-+} pseudoscalar is slightly heavier than the tensor, and the lightest 1^{+-} and the first excited 0^{++} lie in the $2.5 - 3.0$GeV range. Many glueballs with other J^{PC} quantum numbers lie in the $3 - 4$GeV mass range. (For what is probably the most extensive recent study of the SU(3) glueball spectrum see [14].) So we see that most glueballs are relatively heavy compared to the quarkonia ($q\bar{q}$) states that appear to dominate the experimentally determined spectrum of states and which appear in the spectrum of full QCD.

What do we expect to happen to these glueball states if we go to QCD by adding light quarks to this SU(3) field theory? The simplest possibility is that the glueball states will be relatively unaffected unless there is a quarkonium state with similar quantum numbers that has a similar mass. In that case the states will mix. In addition some of the lighter glueball states, that are stable in the pure gauge theory, will be able to decay into quarkonia. Although we expect (by the 'square root' of the OZI rule – or by the fact that SU(3) is 'close to' SU(∞)) that both such mixings and decays will be somewhat suppressed, it is clear that identifying glueballs will be tricky in such a situation. The thing we can look for is 'too many' states. The glueball has vacuum quantum numbers and so will mix with $q\bar{q}$ states with these quantum numbers. In the typical $q\bar{q}$ nonet there are two such states. Thus an experimental signal for the presence of a glueball is if we find 3 rather than 2 such states. Unfortunately most of the glueballs have high masses where we can expect many excited quarkonia (not to mention hybrids) but do not know precisely how many. So it is only for the lightest

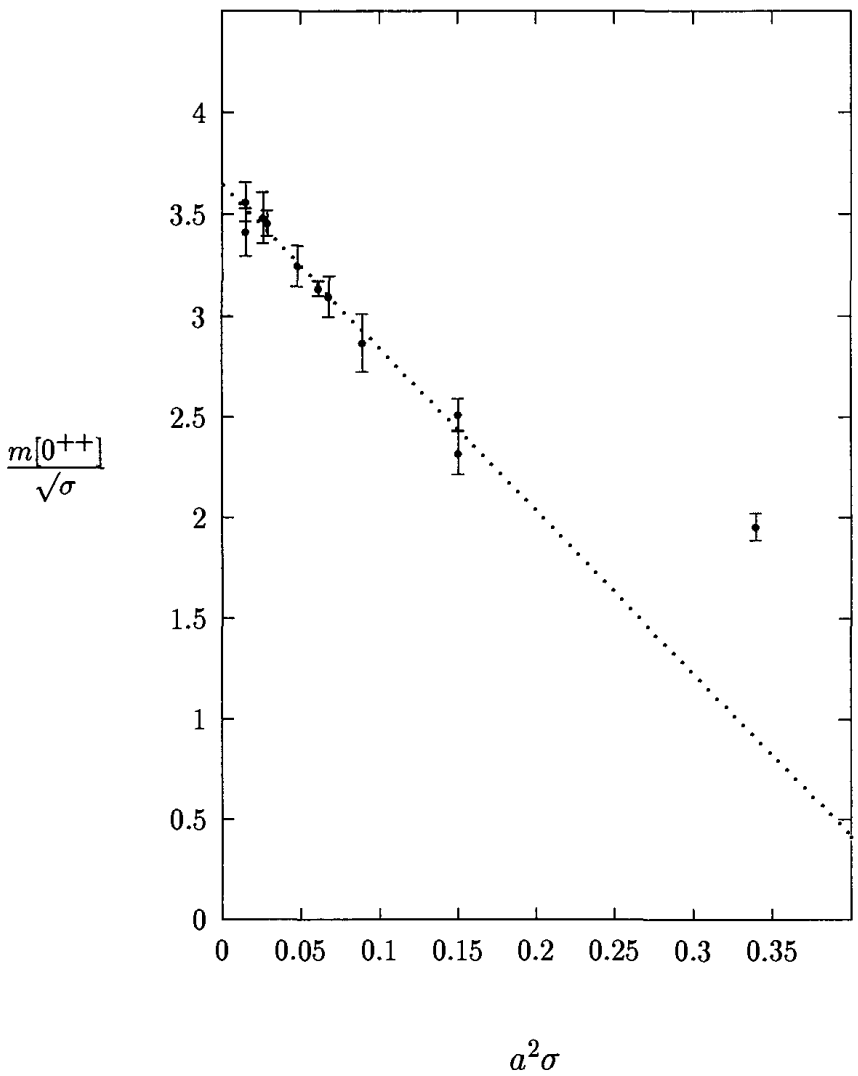


Figure 3: The lightest $J^{PC} = 0^{++}$ glueball mass, in units of the string tension, from various SU(3) calculations [13]. Plotted against the string tension in lattice units. Shown also is an extrapolation to the continuum limit.

glueballs that one has much hope at present, and indeed it is here that there has been very exciting recent progress.

In particular consider the lightest glueball of all, the 0^{++} at around 1.6 GeV. This is close to the lightest $q\bar{q}$ 0^{++} nonet. Motivated in significant part by the lattice calculations, which have consistently claimed such a mass for the glueball since the late 80's, a great deal of experimental and phenomenological effort has been going into elucidating the flavour singlet spectrum in this mass range. In the last few years it has become clear that there are in fact three such mesons in the experimental spectrum: the $f_o(1370)$, the $f_o(1500)$, and the $f_o(1710)$, and one interprets these states as arising from the mixing of $u\bar{u}+d\bar{d}$, $s\bar{s}$ and glueball states. I do not have time to dwell further on this fascinating development – see for example [15] for recent reviews.

3 The physics of SU(N) gauge theories

To study SU(N) gauge theories numerically one performs calculations for increasing values of N until the calculated values cease to change significantly. Once a smooth large-N limit has been established in this fashion one can extrapolate to $N=\infty$ using the expected leading $O(1/N^2)$ correction. Typically one finds that it is enough to perform calculations over the range $2 \leq N \leq 5$ or 6.

There is an alternative approach that involves the reduction of the leading large-N limit to a field theory on a single site (and correspondingly larger matrices). This twisted Eguchi-Kawai [16] approach is attracting interest again [17] after long neglect. (See [18] for reviews of earlier work.)

In this Section I shall describe some of the things we have learned about SU(N) gauge theories using lattice calculations. I will begin by asking if SU(N) gauge theories are linearly confining and if the confining flux tube behaves like a string. Then I look at the lightest masses and ask if they have a smooth $N \to \infty$ limit. This allows us to address what is physically the most interesting question: is SU(3) close to SU(∞)? I then move on to investigating the properties of strings between charges that are in representations other than the fundamental. I conclude by summarising what we learn from these calculations as well as from some other relevant calculations that I do not have time to cover in any detail here.

3.1 The confining string and its tension

Consider the minimum energy of two static charges that are in the fundamental representation and that are a distance r apart. This energy is, up to a (possibly infinite) constant, what we shall mean by the potential energy, $V(r)$. We can calculate this energy using Euclidean correlation functions as in eqn(6). For this we need a suitable operator. Let $\phi(\vec{r},t)$ be the charged matter field that provides our charges. Then a

suitable operator is

$$\Phi(r,t) = \phi^\dagger(0,t)U(0,r;t)\phi(r,t). \tag{32}$$

Here the U is the path ordered exponential of the gauge potential defined in eqn(15), but with $U(x = \{0,t\}, y = \{r,t\})$ abbreviated to $U(0,r;t)$, and we have simplified the notation, showing only the spatial coordinate r that will be varied. We can easily see by applying eqn(16) that $\Phi(r,t)$ will be gauge invariant. Now we also want r to be a good quantum number; this will only happen if the charges ϕ are static, i.e have infinite mass. If we calculate such a correlation function then we get precisely the Wilson loop, from whose expectation value we can extract $V(r)$. But if we are interested only in the string tension, there is a simpler way to proceed, which takes advantage of the fact that we are on a spatial torus of size l. Imagine separating the charges until $r = l$ and the charges are at the same space-time point. We can 'annihilate' them and we are left with

$$\Phi(l,t) = TrU(l;t) \tag{33}$$

which is the path ordered exponetial of the gauge potential around a curve that closes on itself around the spatial torus. If the theory is confining such an operator will couple to states that contain a flux tube that encircles the spatial torus. From the Euclidean correlator of such an operator we will obtain the lightest mass, $m_f(l)$ of a flux tube of length l:

$$\langle TrU^\dagger(l,t)TrU(l,0)\rangle \stackrel{t\to\infty}{=} |c|^2 e^{-m_f(l)t}. \tag{34}$$

On a lattice of spatial size L in lattice units, the calculation is exactly the same with

$$U(l,t) \to U_P(L,n_t) = \prod_{n_k=1}^{n_k=L} U_k(n_k,n_t) \tag{35}$$

where the subscript P refers to the fact that such an operator is conventionally called a Polyakov loop. The correlator in eqn(34) then gives us the flux loop mass $am_f(L)$ in lattice units.

If the theory is linearly confining, the flux loop cannot break and its mass will be proportional to its length once the length is long enough that sub-leading corrections can be dropped:

$$am_f(L) \stackrel{L\to\infty}{=} a^2\sigma L. \tag{36}$$

By performing calculations for different L we can check the hypothesis of linear confinement. In Fig. 4 I show the result of such a calculation in SU(2) [19]. We can clearly see that $am_f(L)$ increase roughly with L; that is to say SU(2) is linearly confining. Of course a numerical calculation can only check this over a finite range of distance L. It is clearly interesting to know how large our largest distance is in physical units since that affects how significant is the result. We obtain this as follows. Using eqn(36) we extract $a^2\sigma$ from our loop masses. From this we extract a in units of $1/\sqrt{\sigma}$ which is a natural physical length scale in the theory. We find $a \simeq 0.157/\sqrt{\sigma}$. To turn this into

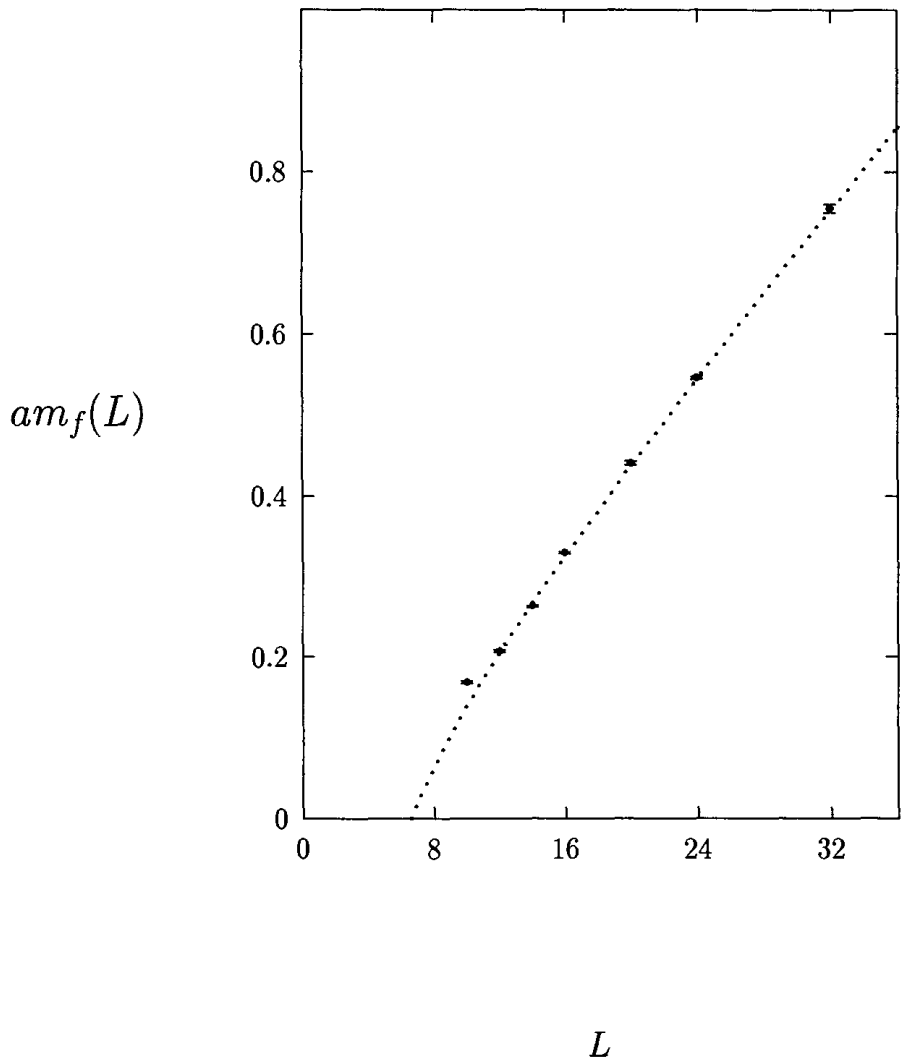


Figure 4: The mass of the lightest (periodic) flux loop of length $l = aL$. In SU(2) at $\beta = 2.55$.

more familiar units I use $\sqrt{\sigma} \sim 440\mathrm{MeV}$ to give $a \sim 0.07\mathrm{fm}$. This is small enough that one would expect lattice corrections to be small: we are 'close' to the continuum limit. In these units the largest value of $l = aL$ in Fig. 4 is $32a \sim 2.2\mathrm{fm}$. This is quite large compared to the dynamical length scale of the theory, and so we have in Fig. 4 a very significant test of linear confinement.

It is also apparent from Fig. 4 that we see some deviations from linearity. These deviations are in fact of deep theoretical interest. If the confining flux tube is described by an effective string theory at large distances, then the leading correction is known [20]:

$$m_f(L) \stackrel{L\to\infty}{=} aL\sigma - \frac{\pi}{6}\frac{D-2}{aL}c_s. \tag{37}$$

Here D is the dimensionality of space-time, and c_s is the central charge of the effective string theory. The correction is essentially the string Casimir energy coming from the sum of zero point energies when we quantise the fluctuations of the string. There is an infinite piece $\propto L$ that is absorbed into σ and the above $\propto 1/L^2$ correction which is dominated by contributions from the long wavelength fluctuations. This correction is universal (in the technical sense) with c_s labelling the universality class. The simplest possibility is $c_s = 1$ corresponding to a simple bosonic string. However the string effective theory could be different, perhaps with fermionic 'matter'. For example, the Neveu-Swartz string has $c_s = 3/2$. We can take neighbouring lattice sizes L_i, L_{i+1} from our calculation and fit the two polyakov loop masses to the formula in eqn(37) (with D=4) so as to extract an effective value of the central charge $c_s^{eff}(L_i, L_{i+1})$. If c_s^{eff} becomes independent of L_i as $L_i \uparrow$, then this tells us that the functional form of the leading correction in eqn(37) is indeed correct, and

$$c_s^{eff}(L_i, L_{i+1}) \stackrel{L_i\to\infty}{=} c_s. \tag{38}$$

In Fig. 5 I plot the result of the calculation and we see some real evidence that the long distance physics of confining flux tubes is given by a simple bosonic string theory, with central charge $c_s = 1$.

What we would now like to do is to repeat these calculations for SU(3), SU(4), SU(5), ... so as to see if we continue to have linear confinement at large=N. There have, of course, been many calculations in SU(3), some of comparable quality to the above SU(2) one, and they come to the same conclusion: linear confinement, with an effective string theory that is probably of the simple bosonic kind. For the higher N that is our real interest here, the calculations are less accurate but still useful. As an example I show in Fig. 6 a calculation [19] of the lightest flux loop mass versus length in the case of SU(4). We see the approximate linear rise with L that is a signal of linear confinement. However this is for a rather coarse lattice spacing, $a \simeq 0.14\mathrm{fm}$ (introducing 'fermi' units in the same way as for SU(2)), and although the largest loop length, $14a \sim 2\mathrm{fm}$, is quite long, the accuracy is too poor to provide a significant test of the universality class. All one can say is that the deviation from linearity is consistent

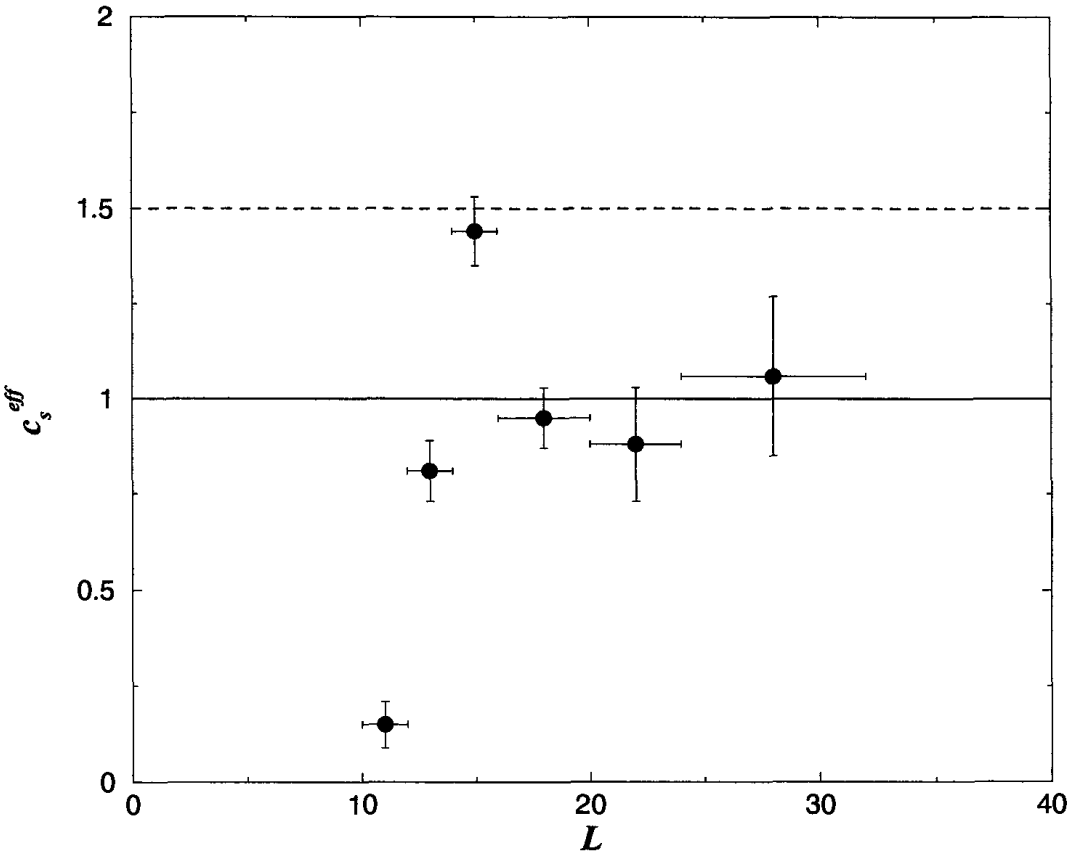


Figure 5: The effective value of the central charge from the flux loop masses plotted in Fig. 4. The values for the simple bosonic (—) and Neveu Swartz (– – –) string theories are indicated.

with what one expects from $c_s = 1$. Similar comments apply to calculations in SU(6) [21] where the approximate linearity is quite clear but the value of the correction is less well determined.

Nonetheless, while the precise nature of the effective string theory that describes the confining flux tube as $N \to \infty$ needs more work, it is clear that confinement is linear for $N \leq 6$ at least, and this provides strong support for the idea that the SU(∞) theory is linearly confining.

3.2 The mass spectrum

The simplest thing we can do, now that we have confirmed linear confinement, is to repeat the calculation of the light glueball masses, that we performed in SU(3), for other SU(N) gauge theories. This is precisely what was done in [22] where the lightest and first excited 0^{++} masses and the lightest 2^{++} mass were calculated for $N = 2, 3, 4, 5$ and the appropriate continuum limits were taken (just as we did in our SU(3) example). In Fig. 7 I display these glueball masses in units of the string tension. The masses are plotted against $1/N^2$ because the usual diagrammatic arguments tell us that at large N the leading correction should be:

$$\frac{M}{\sqrt{\sigma}}(N) = \frac{M}{\sqrt{\sigma}}(\infty) + \frac{c}{N^2}. \tag{39}$$

Such an extrapolation will be a straight line in our Figure and we show best fits of this kind to all the three mass ratios. There are some things we observe.

• The mass ratios show a remarkably weak N-dependence and we are able to fit with the leading correction all the way down to SU(2). This is of course not very significant in the case of the 0^{++*} because the errors are so large, but for the 2^{++} and particularly for the 0^{++} the errors are small enough for this observation to be significant.

• The rapid convergence as $N \to \infty$ is also telling us that $\sigma \not\to 0$ as $N \to \infty$. This plugs a gap in our earlier reasoning: even if we have linear confinement it could be that the string tension vanishes smoothly as $N \to \infty$. Apparently it does not.

Thus we conclude that not only SU(3) but even SU(2) is close to SU(∞). (At least if we look at the lightest states in the mass spectrum.) This increases our expectation that for QCD, where the leading corrections are $O(1/N)$ rather than $O(1/N^2)$, $QCD_{N=3}$ will indeed be 'close' to $QCD_{N=\infty}$.

3.3 k-strings

Having shown that SU(3) is close to SU(∞) we are now going to look more closely at one of the most interesting differences.

In particular we ask what kind of different stable confining strings there are in the theory. One might think that there are very many, corresponding to the many

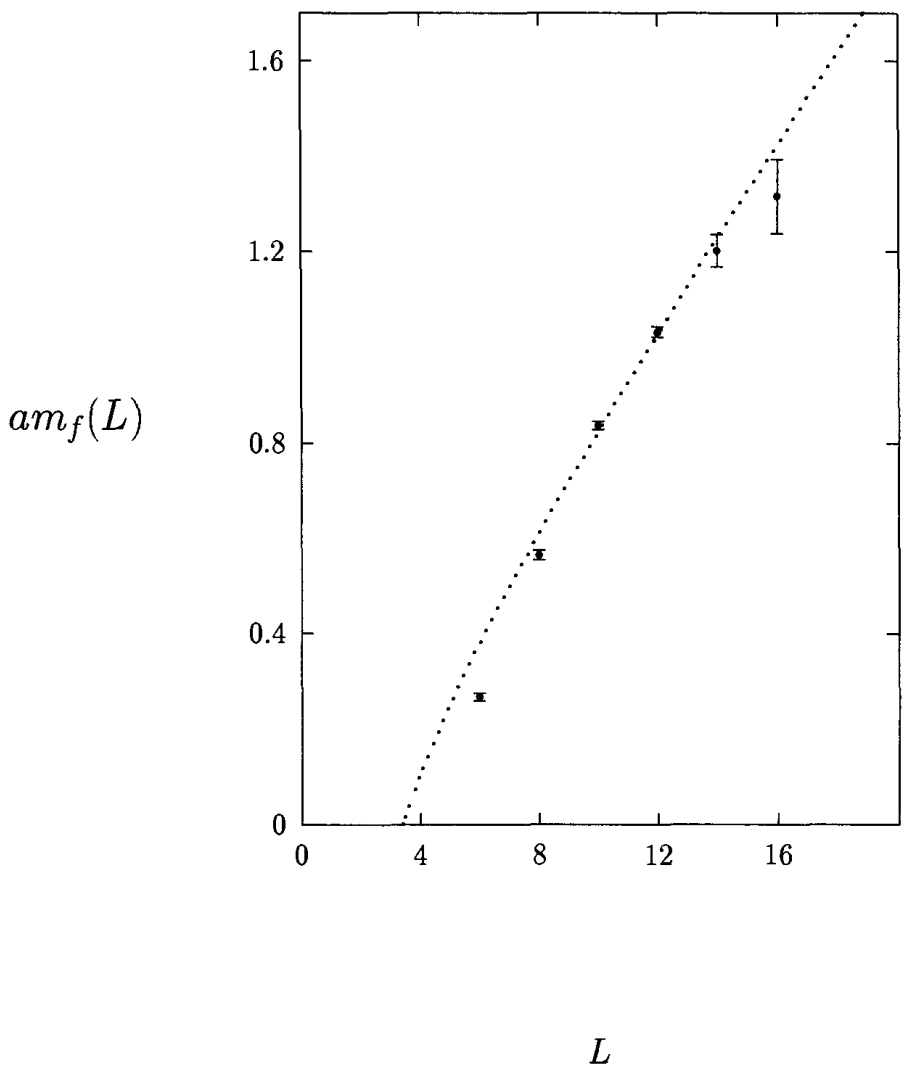


Figure 6: As in Fig. 4, but for SU(4) at $\beta = 10.7$.

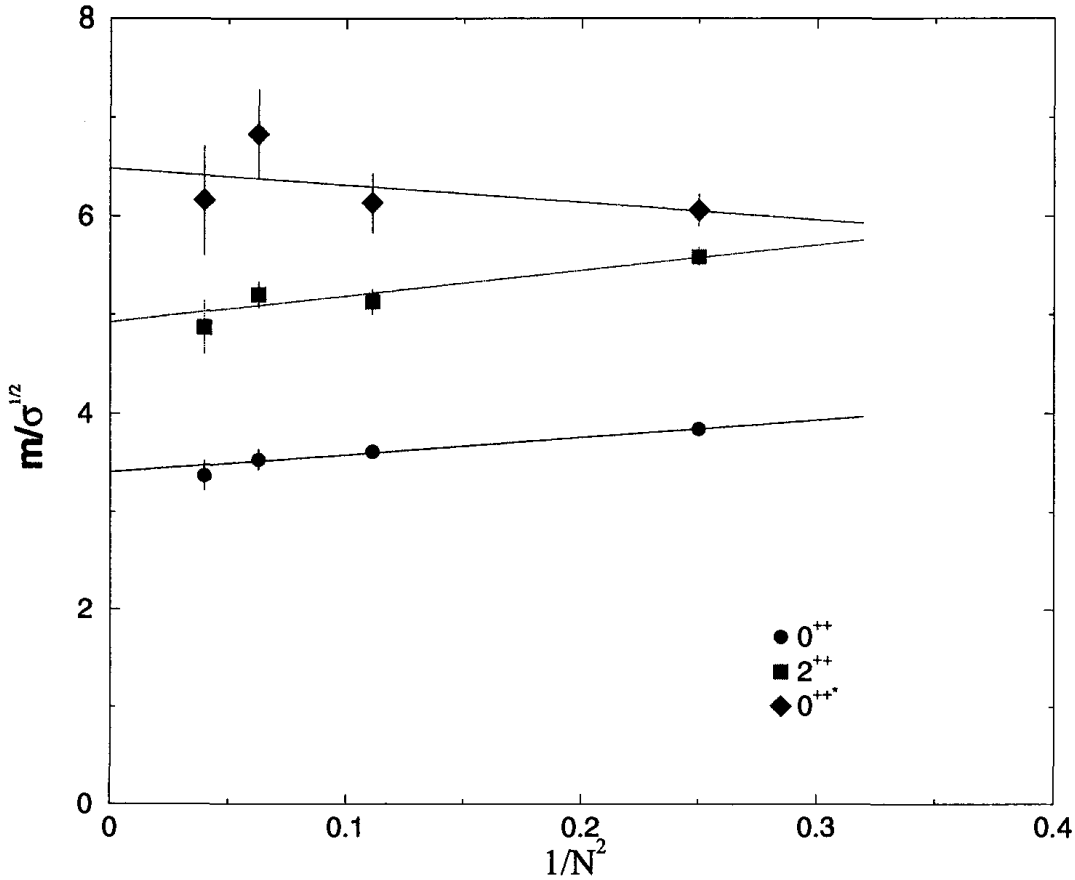


Figure 7: The lightest scalar, first excited scalar and lightest tensor glueball masses in units of the string tension, for various SU(N) gauge theories. Extrapolations to $N = \infty$ are shown.

different representations that the separated charged sources might belong to. However one must remember that the sources can be screened by gluons from the vacuum, and that this will certainly happen if it is energetically favourable. And it will certainly be energetically favourable, at large enough separations, if the two string tensions (with and without screening) are different, because the energy difference grows linearly with separation and becomes arbitrarily large at large separations.

Therefore there will only be one stable string corresponding to the set of sources obtained from a given representation by multiplying by an arbitrary number of adjoint representations (since gluons are in the adjoint representation). Now recall that under a gauge transformation z_i that belongs to the centre Z_N of the SU(N) group, a source consisting of a clump of k particles in the fundamental representation (think of e.g. k quarks) will transform as z_i^k and because $z_i A_\mu z_i^\dagger = A_\mu$ (by definition the centre elements commute with all elements of the group and hence of the Lie algebra) we have the same transformation property however many gluons we add. Thus for each k we can have, in principle, a different stable string. We call it a k-string. Of course, because $z^N = 1$, strings with k and $N-k$ are in the same class so we only have different strings corresponding to $k = 1, ..., [N/2]$ where $[N/2]$ corresponds to the integer part of $N/2$. So for SU(2) and SU(3) there is only one stable string, while for SU(4) and SU(5) we have two ($k = 1$ and $k = 2$) and for SU(6) three, etc.

These strings and their corresponding string tensions, σ_k, are interesting because they serve to distinguish between various models and theoretical pictures. For example:

• If we find

$$\sigma_k = k\sigma_1 \tag{40}$$

this means that the k-string is really just k independent fundamental strings i.e. the fundamental strings are not bound. Actually there are general expectations that $\sigma_k \to k\sigma_1$ as $N \to \infty$ at fixed k. What is interesting is the k and N dependence at finite k and N or the dependence for k/N fixed as $N \to \infty$.

• If $\sigma_k < k\sigma_1$ then the k fundamental strings are bound; the precise value should tell us something about the dynamics of confinement.

• There are old ideas that go under the name of 'Casimir scaling' [23] which predict that the string between charges in a representation $\mathcal{R}$ will have a tension that is proportional the quadratic Casimir, $C_\mathcal{R}$, of the representation

$$C_\mathcal{R} \equiv Tr_\mathcal{R} T^a T^a \tag{41}$$

with the T^a being the generators of the group. The stable k string will belong to the totally antisymmetric representation since that has the smallest Casimir in the set that transform as z_i^k. This predicts that

$$\frac{\sigma_k}{\sigma} = \frac{k(N-k)}{N-1}. \tag{42}$$

• A number of calculations in QCD-like theories that are approached from brane(M)-theory (see [24] and references therein), which are generically referred to as MQCD, find that that the string tension of k-strings satisfies

$$\frac{\sigma_k}{\sigma} = \frac{\sin\frac{k\pi}{N}}{\sin\frac{\pi}{N}}. \tag{43}$$

This led to the conjecture [24] that this might be a universal result and that QCD (and SU(N) gauge theories) fall into this universality class. (Although conter-examples to such universality have since been found.)

• In the bag model [25] the flux between distant sources is confined to a cylindrical bag of cross-section A and is homogeneous, $E_a A = gT_a$. Since the vacuum energy difference between the inside and outside of the bag is given by the bag constant B, the energy per unit length is [25]

$$\frac{E}{l} = 2\pi\alpha_s \frac{C_{\mathcal{R}}}{A} + AB \tag{44}$$

where $C_{\mathcal{R}}$ is the quadratic Casimir of the source and α_s is the strong coupling constant. One now fixes the area A by minimising the energy. This predicts the string tension to be

$$\sigma_{\mathcal{R}} \propto \{C_{\mathcal{R}}\}^{\frac{1}{2}} \propto \left\{\frac{k(N-k)}{N-1}\right\}^{\frac{1}{2}} \tag{45}$$

– which differs markedly from Casimir scaling result in eqn(42).

There have been two sets of relevant lattice calculations in the last couple of years. (The pioneering calculation of this kind [26] is too inaccurate to be useful here.) The calculations in [19] were for the $k = 2$ and $k = 1$ strings in SU(4) and SU(5), while those in [21] studied SU(4) and also SU(6) where there is a $k = 3$ string. The tensions of the k-strings are calculated from correlations of operators that involve all different ways of making a product of k fundamental Polyakov loops – the latter as described in Section 3.1. Otherwise the calculation is exactly as for the fundamental ($k = 1$) string tension. At each β we obtain $a^2\sigma_k$ for all possible k. In Fig. 8 I show the ratio $a^2\sigma_{k=2}/a^2\sigma_{k=1} = \sigma_{k=2}/\sigma$ plotted against $a^2\sigma$ for both SU(4) and SU(5) gauge theories [19] together with continuum extrpolations, of the form in eqn(26), in each case. One obtains for the continuum limit:

$$\lim_{a\to 0}\frac{\sigma_{k=2}}{\sigma} = \begin{cases} 1.357 \pm 0.029 & \text{SU(4)} \\ 1.583 \pm 0.074 & \text{SU(5)}. \end{cases} \tag{46}$$

It is obvious from Fig. 8 that $\sigma_{k=2}$ is substantially smaller than $2 \times \sigma_{k=1}$: k-strings are indeed strongly bound. The values, moreover, are much larger than the bag model prediction in eqn(45): so the bag model is also excluded. On the other hand the values in eqn(46) are close to both MQCD (1.41.. and 1.61.. for SU(4) and SU(5) respectively) and to Casimir scaling (1.33.. and 1.50 respectively). The accuracy of these calculations

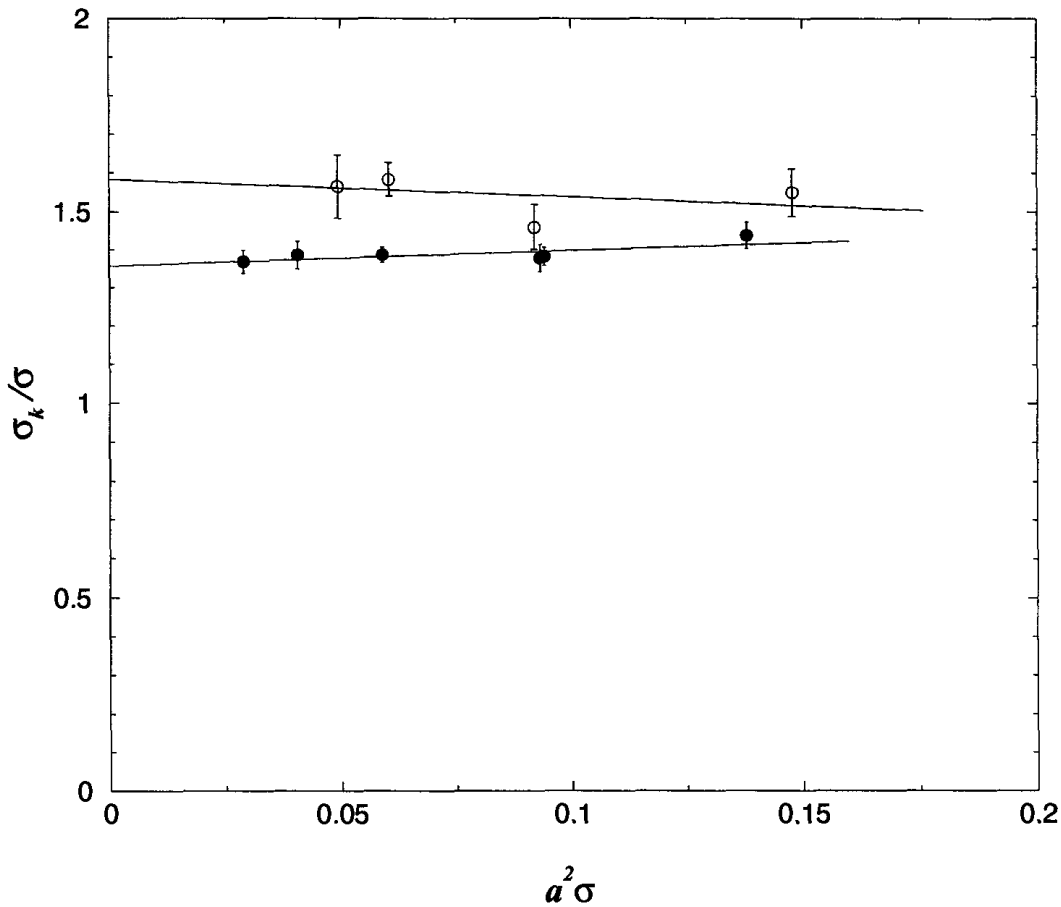


Figure 8: Ratio of the $k = 2$ to fundamental string tensions in SU(4) (•) and SU(5) (○) lattice gauge theories, plotted against $a^2\sigma$. Continuum extrapolations are as shown.

is not quite good enough to distinguish between these two predictions, largely because they are in fact rather close together.

The more recent calculations in [21] are much closer to the MQCD prediction than to Casimir scaling. However in all these calculations one needs to remember that as N increases the calculations obviously get harder, and as k (and N) increases the masses of the loops get much heavier and get *very* much harder to calculate – leading to what is best termed an increasing systematic error. So my personal view is that it is too early to say whether MQCD or Casimir scaling or something in between is favoured. There is still considerably more to do here, even if we have already learned a great deal.

3.4 Some other results

I have not had the time to discuss many other interesting calculations that are in the process of teaching us about the large-N limit. Here I will refer to some of this work so that you can go and look at it in your own time.

First let me mention recent attempts to obtain accurate information on the order and value of the finite temperature deconfinement transition in SU(N) gauge theories [27]. Secondly, the values of the tensions of k-strings should tell us something about the dynamics of confinement. In models where vortices produce confinement we can learn about these vortices [19, 28]. There has also been work on the Abelian projection approach to confinement [29]. Thirdly, topological fluctuations are of particular interest: the loss of identifiable smaller instantons at large N has been confirmed in lattice calculations [22]. and the large N limit of the topological susceptibility has been obtained [22, 30]. In addition speculations about interlacing vacua as θ is varied from zero at large N, have begun to be confirmed in very accurate lattice simulations [30]. At the same time the relationship betweeen topology and chiral symmetry breaking, and how that changes as $N \to \infty$, has been investigated using recent lattice formulations (Ginsparg-Wilson/Overlap) that respect chiral symmetry and the Index Theorem [31]. As for conventional expectations, the lattice calculations in [22] confirm that the large-N limit is reached by keeping fixed the 't Hooft coupling, $\lambda \equiv Ng^2$. The large-N limit has also been explored in 2+1 dimensional gauge theories, with many of the same conclusions [11, 32, 19]. There the calculations are accurate enough that one can observe that the first non-leading correction is indeed $O(1/N^2)$ as expected from the usual diagrammatic counting arguments. All these calculations use very standard lattice methods, just increasing the SU(3) to larger SU(N) groups. The alternative approach, of space-time reduction, is now also being explored [17].

4 Some final remarks

I hope that in these two lectures I have managed to convince you that numerical lattice field theory provides a powerful technique for elucidating the physics of otherwise intractable field theories. All the calculations I have described in these lectures have been done (or could be done) on modest workstations/PCs of the kind that you will find in any Physics Department. If some of you, when you go back home, find some interesting questions to answer using these techniques, then these lectures will have been well worthwhile.

References

[1] G. Münster and I. Montvay, *Quantum Fields on a Lattice* (CUP, 1994).

[2] M. Creutz, *Quarks, Gluons and Lattices* (CUP, 1983).

[3] H. Rothe, *Lattice Gauge Theories 2nd Ed.* (World Scientific, 1997).

[4] J. Smit, *Introduction to Quantum Fields on a Lattice* (CUP, 2002).

[5] G. 't Hooft, Nucl. Phys. B72 (1974) 461.

[6] S. Coleman, 1979 Erice Lectures.
E. Witten, Nucl. Phys. B160 (1979) 57.
G. 't Hooft, Physics Reports 142 (1986) 357.
A. Manohar, 1997 Les Houches Lectures, hep-ph/9802419.
R. F. Lebed (Ed.), *Large N_c QCD* (World Scientific, 2002).

[7] S. Weinberg, *Quantum Theory of Fields Vol II* (CUP, 1996).

[8] G. Parisi *Statistical Field Theory* (Addison-Wesley, 1988)

[9] M. Albanese et al (APE Collaboration), Phys. Lett. B192 (1987) 163.

[10] M. Teper, Phys. Lett. B183 (1987) 345.

[11] M. Teper, Phys. Rev. D59 (1999) 014512 (hep-lat/9804008).

[12] D. Perkins, *Introduction to High Energy Physics 4th Ed* (CUP 2000).

[13] M. Teper, hep-th/9812187.

[14] C. Morningstar and M. Peardon, Phys. Rev. D56 (1997) 4043 (hep-lat/9704011); Phys. Rev. D60 (1999) 034509 (hep-lat/9901004).

[15] F.E. Close, Int. J. Mod. Phys. A17 (2002) 3239 (hep-ph/0110081).
F.E. Close and A. Kirk, Eur. Phys. J. C21 (2001) 531 (hep-ph/0103173).

[16] T. Eguchi and H. Kawai, Phys. Rev. Lett. 48 (1982) 1063.
G. Bhanot, U. Heller and H. Neuberger, Phys. Lett 113B (1982) 47.
A. Gonzales-Arroyo and M. Okawa, Phys. Rev. D27 (1983) 2397.

[17] R. Narayanan and H.Neuberger, hep-lat/0303023.
J. Kiskis, R. Narayanan and H.Neuberger, Phys. Rev. D66 (2002) 025019 (hep-lat/0203005); hep-lat/0208021, hep-lat/0208060.

[18] O.Haan in *Lattice gauge Theory* (Edited by B. Bunk, K. Mutter, K. Schilling, Plenum Press 1986).
S. R. Das, Rev. Mod. Phys. 59 (1987) 235.

[19] B. Lucini and M. Teper, Phys.Lett. B501 (2001) 128 (hep-lat/0012025);
Phys. Rev. D64 (2001) 105019 (hep-lat/0107007).

[20] M. Lüscher, K. Symanzik and P. Weisz, Nucl. Phys. B173 (1980) 365.
Ph. de Forcrand, G. Schierholz, H.Schneider and M. Teper, Phys. Lett. B160 (1985) 137.

[21] L. Del Debbio, H. Panagopoulos, P. Rossi and E. Vicari, JHEP 0201 (2002) 009 (hep-th/0111090); Phys. Rev. D65 (2002) 021501 (hep-th/0106185).

[22] B. Lucini and M. Teper, JHEP 0106 (2001) 050 (hep-lat/0103027).

[23] J. Ambjorn, P. Olesen and C. Peterson, Nucl. Phys. B240 (1984) 189, 533; B244 (1984) 262; Phys. Lett. B142 (1984) 410.
P. Olesen, Phys. Lett. 160B (1985) 408.
M. Flensburg, A. Irbäck and C. Peterson, Z. Phys. C36 (1987) 629.
M. Caselle, R. Fiore, F. Gliozzi and R. Alzetta, Phys. Lett. 200B (1988) 525; 224B (1989) 153.
A. Loewy and J. Sonnenschein, hep-th/0103163.
V. I. Shevchenko and Yu. A. Simonov, Phys. Rev. Lett. 85 (2000) 1811 (hep-ph/0001299).
Y. Koma, E. -M. Ilgenfritz, H. Toki and T. Suzuki, hep-ph/0103162.
S. Deldar, Phys. Rev. D62 (2000) 034509 (hep-lat/9911008).
G. Bali, Phys. Rev. D62 (2000) 114503 (hep-lat/0006022).

[24] A. Hanany, M. Strassler and A. Zaffaroni, Nucl. Phys. B513 (1998) 87 (hep-th/9707244).
M. Strassler, Nucl. Phys. Proc. Suppl. 73 (1999) 120 (hep-lat/9810059).
M. Strassler, Prog. Theor. Phys. Suppl. 131 (1998) 439 (hep-lat/9803009).

[25] K. Johnson and C. B. Thorn, Phys. Rev. D13 (1976) 1934.
T. H. Hansson, Phys. Lett. B166 (1986) 343.

[26] M. Wingate and S. Ohta, hep-lat/0006016; Nucl. Phys. Proc. Suppl. 83 (2000) 381 (hep-lat/9909125).

[27] B. Lucini, M. Teper and U. Wenger, Phys. Lett. B545 (2002) 197 (hep-lat/0206029)

[28] J. Greensite and S. Olejnik, JHEP 0209 (2002) 039 (hep-lat/0209088)

[29] L. Del Debbio, A. Di Giacomo, B. Lucini, and G. Paffuti, hep-lat/0203023

[30] L. Del Debbio, H. Panagopoulos, and E. Vicari, JHEP 0208 (2002) 044 (hep-th/0204125)

[31] N. Cundy, M. Teper and U. Wenger, Phys. Rev. D66 (2002) 094505 (hep-lat/0203030)

[32] B. Lucini and M. Teper, Phys. Rev. D66 (2002) 097502 (hep-lat/0206027)

CHAIRMAN: M. TEPER

Scientific Secretaries: D. Chakrabarti, C. Dawson

DISCUSSION I

- *Shuryak:*

A comment about 0^{++} and 2^{++} glueball masses, and their dependence on lattice spacing. The dependence of 0^{++} is probably stronger than for 2^{++} simply because of its smaller size. Larger objects, like the 2^{++}, simply do not notice the lattice while the 0^{++} does.

- *Teper:*

It is difficult to say because it is not a universal effect. I think it is not the right way to look. What you want to do is to look at the effects of putting e.g., the glueball in a finite volume in the continuum limit. Perhaps what we see is that the glueball shrinks more rapidly as the lattice spacing goes larger, which is simply a lattice artifact effect. While these order a-squared effects may be written in terms of continuum operators, they will be lattice action dependent, and so are lattice artifacts.

- *Maillard:*

Are we living on a lattice? And if not, why can we describe physics by using a lattice?

- *Teper:*

You know, for example, that if you write down a continuum lagrangian it is not defined because anything you calculate is infinite, so you have to introduce a cut-off and then you remove the cut-off in a careful way, keeping the physics unchanged. In a perturbative calculation, you can use dimensional regularisation, but for a non-perturbative calculation you use the lattice. Because the theory is renormalisable as you remove the cut-off, only a few parameters have to be changed such as the coupling, masses and the wavefunction renormalisation. So it is really just a regularisation, and while I show you calculations at a fixed lattice spacing, the purpose is always to take the continuum limit to get the final result.

- *Stamen:*

What is the motivation to go for N tends to infinity? Is it to get rid of the quarks because quarks are hard to put on the lattice? Or is there some other reason behind that?

- *Teper:*

The N tend to infinity limit is in many ways a lot simpler than N=3, for example there are no mixings or decays. The hope was that the N tends to infinity limit would allow many problems to be approached analytically and this is in fact true. For example the Ads/CFT calculations and calculations in super-string theory actually end up with results in the N equals infinity limit.

In general it is more likely that you will be able to get an analytic result for the N equals infinity limit than for N=3. You can also say many exact statements about the N equals infinity limit, like the fact that there is no decay, without solving it fully, providing an excellent starting point for further study.

- *Papinto:*

Are the simulations quenched or unquenched?

- *Teper:*

All the simulations I will show use pure gauge theory, with no quarks. In the N tends to infinity limit, this makes no difference as the quark loops start at order 1/N. So anything I learn from the pure gauge theory will be the same as I would have learned by looking at glueballs in QCD in the N equals infinity limit, as long as I keep the quark mass finite. For the N=3 case the effects are much less clear.

- *Maas:*

Lattice calculations can also access gauge-variant quantities, although normally measuring only gauge-invariant quantities. Is it possible to access ghost dynamics in lattice simulations?

- *Teper:*

On the lattice, we are integrating over the gauge group, rather than integrating over the lie algebra and this is finite, so no gauge fixing is needed. For some quantities, however, you need to be in a fixed gauge to calculate, for example if you want to attempt to calculate a gluon mass, and some people have done this. There have also been studies of Gribov copies on the lattice. I am not sure if people have calculated anything related to the Fadeev-Popov term.

- *Ramtohul:*

Is it possible to calculate the order 1/N terms on the lattice and if so how computationally intensive would it be?

- *Teper:*

You can certainly fit to the 1/N dependence of lattice data by calculating at several values of N, but I am not sure if you can go further and actually formulate the problem to calculate the coefficient of 1/N directly.

- *Amelino-Camelia:*

You discussed for us lattice techniques for the study of SU(N) gauge theories, which are perturbatively renormalisable. Is perturbative renormalisability a key tool for your type of analysis? Could one use lattice techniques to analyse theories which are not perturbatively renormalisable? Any hope for quantum gravity?

- *Teper:*

People have explored discretised versions of gravity on the lattice, both in two and higher dimensions. However, in practice, having a perturbative result is very useful in guiding the analysis of lattice data. For example, as I am taking the continuum limit, it is useful to know that the leading correction goes as the square of the lattice spacing, allowing me to perform a simple extrapolation to a lattice spacing of zero. If you do not have the knowledge of the functional form, it becomes much harder to extract accurate results. It is not impossible to study theories that are not perturbatively renormalisable, it is just more difficult.

- *Korthals-Altes:*

Non-renormalisable gauge theories have been investigated on the lattice: SU(2) in 5 dimensions. This theory is not renormalizable for the lattice coupling beta going to infinity. Now the strong coupling (beta small) regime and the Coulomb regime are separated by a first order transition at some finite beta. So if we try to formulate a continuum limit, we have the difficulty that correlation lengths stay finite, so the lattice length does not become negligibly small.

- *Teper:*

This reminds me of the many people who tried looking for a continuum limit of strong coupling QED. There is phase transition like that in the pure gauge theory (compact U(1)) between strong and weak coupling. Then people discussed what happened when you add fermions. Part of the problem was that there was a lack of control which led the different groups to disagree. This is another example of what happens when you do not have control over the short distance behaviour.

- *Kenway:*

In principle, you can study non-renormalisable theories on the lattice, provided you can find a second order phase transition of which you can take the continuum limit. You also need to understand the functional behaviour of physical quantities as you approach this point if you are able to extrapolate them to the continuum limit. This is much easier when you have perturbative control of the continuum limit.

- *Teper:*

Calculation of properties such as the mass ratios as you approach can be quite hard if you do not know the functional form. For example, there is second order point that people believe is a lattice artifact that causes the dip in the mass of the O^{++}

glueball and if you look in an extended coupling space, you find a critical point. People do not think this is any kind of continuum limit, but I do not think that anybody has studied it that hard. In fact it is very difficult to determine the properties of the theory near that point.

- *Mass:*

Can lattice calculations test if multiplicative renormalisability, which is only proven in perturbation theory, holds also in the non-perturbative regime in SU(N)-gauge theory?

- *Teper:*

In principle lattice calculations could have shown that the perturbative prediction of a continuum limit at beta = infinity was incorrect, showing that there was a problem with perturbative renormalisability. However what people have found is that perturbative results seem to hold. For example, Luscher et al have studied the running of the coupling constant non-perturbatively by comparing several lattice spacings following the coupling for quite large length scales down to a scale of around several GeV. They then compared this with high order perturbative calculations and found it to match very well. People have also expressed masses on the lattice in terms of lambda parameters like lambda $M\overline{S}$ by matching up with perturbative calculations. So in general, perturbative renormalisability seems to work well, even in the context of a non-perturbative lattice calculation.

- *Stamen:*

There are currently several experiments running which test the proton structure. These machines will be switched off within the next ten years. Which experiments do we have to do today such that lattice results can be compared to that in the future?

- *Teper:*

People are certainly currently calculating moments of structure functions on the lattice. These studies could be tried for other particles than the proton, as on the lattice it is easy to change the particle you are studying. In general I am not sure if the lattice community has anything specific that they would like experimentalists to study so we can question it against the lattice.

- *Kenway:*

There are two problems with current lattice calculations. Firstly, because we are working in Euclidean space, we can only calculate certain matrix elements, which is what restricts people to the moments of the structure functions rather than the structure functions themselves. The second point is that in the future it is critical to include light dynamical quarks in the calculations. As such, we have quite a way to go before we really know what we want from experiments.

- *Korthals-Altes:*

The lattice by definition finds quantities, Green's functions in the Euclidean region and it finds it numerically. In order to fully answer the question on such a proton structure, you have to analytically continue this into Minkowski space, which is a very difficult task from numerical data.

- *Teper:*

Well, the experimentalist can calculate the moments of structure function, which can then be compared to lattice calculations. Of course, if you want to calculate something like S-matrix elements, then it does become harder.

- *Levy:*

You predict and calculate glueball masses and I have not seen yet a glueball in an experiment.

- *Teper:*

This is a tricky question. The tricky part is because the lattice calculation predicts the mass of the lightest glueball to be around 1.6 GeV, a region where you will find a lot of mesons, quark-antiquark states, with vacuum quantum numbers which can mix with the scalar glueball. The simplest way to see if you have a glueball is in the presence of extra states. For example, if you have a scalar glueball, it can mix with $s\bar{s}$, $u\bar{u}$, $d\bar{d}$ and so you will find three states where you would otherwise expect two from the quark model.

Recently, the consensus seems to be growing among experimentalists and phenomenologists, partly motivated by the lattice results which have been firmly saying for the past 12 years that the scalar glueball is at 1.6 GeV, that there is indeed a scalar glueball mixing with two nearby q-$\bar{q}$ states. I think it is only in the last couple of years that people have really worked out that there are three spin 0 states here, although some are very broad. It is now thought that this comes from a mixture of the glueball and the two mesonic states. Different people have different ideas about how much of a glueball is in each different state and they have been analysing decays to try and work this out, but the data on that is not so good and for some decays there is not a lot of information. They also look at which of these are produced in so-called "gluon enhanced" channels like double pomeron collisions. So it is gelling that there is a scalar glueball together with two quark-antiquark states. Without pushing it too far, as there still is an element of speculation, I would say that the lightest glueball is here.

The tensor glueball is another question. It is harder because there are too many tensors in the region of 2.2 GeV where we expect the lightest tensor glueball to be, so it can mix with many different states, possibly even with four-quark states.

- *Markov:*

In hadronic phenomenology there is some speculation about exotic states. Could lattice QCD tell us about this?

- *Teper:*

Yes, people have been calculating hybrid states and looking at four quarks states. In fact, part of the whole complication with the 0^{++} is with the four quark states. One has to be honest and say that there are other scalar states present than those mentioned in my answer to the previous question. The s-$\bar{s}$ and the u-$\bar{u}$ plus d-$\bar{d}$ are what the quark model tells us, but there are some funny states that come in lower down. For example there is an $f_0(980)$, a $\sigma(600)$ and an $A_0(980)$ and people believe these are really some kind of four-quark states. Some time ago Jaffe pointed out that there is enhancement in the S-wave of the two quark, two anti-quark system which in some cases could be interpreted as a $K-\bar{K}$ molecule and in others just an attractive channel.

Also in the usual potential model, you have the quark and anti-quark moving in a potential, which has no quantum numbers. This does not have to be the case. For example, you could have the quark and anti-quark connected by some sort of flux tube, which might be the case if the quark and anti-quark are a fairly large distance apart, and the flux tube could be rotating with some net angular momentum which would change the quantum numbers of the state. This kind of state is called a Hybrid, and people have been calculating their masses with some success (especially when using charm and bottom quarks). There is an ongoing interaction between experimentalists and theorists in order to look for these states.

As far as states with exotic quantum numbers are concerned, unfortunately these seem to have very large masses, making them difficult to study on the lattice and experimentally.

- *Polosa:*

In your large-N approach should 4-quark states be forbidden?

- *Teper:*

Quark loops are suppressed in large N, but you do have the usual quark-antiquark hadron states which propagate and have a mass. It is just their virtual creation that is suppressed. You can make any number of hadrons using any number of quarks and anti-quarks and they will propagate. It is just that quark loops will not be part of that propagation as they are suppressed by 1/N. While there may be some situations for which quark loops are especially important, I know of no reason to think that this will be true for four-quark states. Their mass should be roughly the same as in SU(3) and has no reason to be large. Ed Shuryak has told me about analytic studies of $f_0(980)$ states in instanton models, which showed that things did not change very much in the large N limit.

- *'t Hooft:*

I would like to know what can be said about boundary effects due to the finite size of the lattice box. It seems to me that those effects can be estimated and modelled both using the lattice itself and effective hadron models. This way one may be able to save computation time and increase accuracies by going to smaller lattices.

- *Teper:*

I am sure that this is correct. However, people have spent more time studying the ultraviolet region devising actions that allow them to work at larger lattice spacings.

- *Korthals-Altes:*

In the early 80's, Eguchi and Kawai studied twisted boundary conditions.

- *'t Hooft:*

Yes, in the quenched approximation in the large N limit the boundaries do not talk to each other very much.

- *Teper:*

Using the Eguchi-Kawai technique may be a good idea, but it does have the difficulty that it only treats the large N limit, and does not allow you to see the approach to it as the corrections are not those that you would have in the normal SU(N) theory.

- *Korthal-Altes:*

Also it is difficult to take the continuum limit.

- *Teper:*

You can compare the Eguchi-Kawai results to the continuum limit calculated by the usual methods for SU(3) and SU(4). In fact, in the early 80's people did use the Eguchi-Kawai to calculate glueball masses by compactifying some directions, but not all of them. This did not work too well, but this may have been because the methods for calculating glueball masses were not very well developed at that time.

- *Shuryak:*

There is also some recent work by Luscher that exploits boundary effects by using small lattices, such that all hadrons fit except the pion, together with analytic predictions for the corrections to the infinite volume results to extract physical quantities such as scattering lengths. So there have been some very nice developments in that direction, although nobody has used it yet.

CHAIRMAN: M. TEPER

Scientific Secretaries: J. Kotanski, A. D. Polosa

DISCUSSION II

- *Catà:*

Since you have been dealing with lattice $SU(N)$ for different values of N $(2,3,4,5,6,...)$, what is your opinion about the fact that these theories look so similar?

- *Teper:*

The expected corrections for $SU(N)$ gauge theories are order $1/N^2$, e.g. $1/9$ for $N=3$, and that is quite a small number. So it is not astonishing to find that $SU(3)$ is quite close to large N. If you look at QCD with N colours the corrections are of order $1/N$ so it might well be that QCD is not as close to QCD $(N \Rightarrow \propto)$ as is the pure gauge theory $SU(3)$.

However the fact that even $SU(2)$ is quite close to $SU(infinity)$ in the pure gauge theory makes me hope that maybe QCD with $N=3$ will be close to its $N \Rightarrow \propto$ limit. Anyway one has to do the calculation and see eventually what comes out.

If you want to find the properties of QCD at $N \Rightarrow \propto$, then what you can do is essentially a quenched calculation at different N and at various fixed quark masses and then take the limit to $N \Rightarrow \propto$ at each of these fixed quark masses. That is much easier to do than a full dynamical calculation as a function of N. The quenched calculation could be done today using available techniques and resources and one could find out what is for instance the mesonic spectrum of QCD with $N \Rightarrow \propto$ and compare it with the spectrum in the real world which is presumably the spectrum of QCD with $N=3$. In a year or two, people will have good QCD spectra at $N=3$ anyway. When we started our calculation we had no strong bias to think that $N=3$ is very close to $N \Rightarrow \propto$. But it seems to be pretty close.

- *Kenway:*

Can you learn anything about $O(1/N)$ corrections to QCD by using the quenched approximation or do you have to do dynamical quark simulations?

- *Teper:*

No, not directly, because there are no virtual quark loops in the quenched approximation. However, as I just remarked, suppose that you want to know whether the large N limit is close to real QCD – what you can do is to approach the large N limit in a "wrong" way via quenched QCD, and then compare the results using the kind of plot I have shown you. I think it would be worthwhile to do.

- *Stamen:*

What are the current limitations for your mass calculations? Why are the errors so large?

- *Teper:*

Part of the answer is that our calculation is not a large-scale calculation by modern standards. With larger machines than those we used here, we could have produced much more accurate results. Also, as I discussed earlier, you need to improve your large smooth operators to perform more accurate computations. You could improve the calculation, if you had more computer time, by going to smaller lattice spacings. Basically one reason the errors get large when you go to large N is that in your Monte Carlo calculations you start multiplying larger complex matrices which are even more time consuming. In other cases the error becomes larger for reasons which are more difficult to deal with because they have a physical origin (for example you lose small instantons at large N).

Also we only calculate three masses, the scalar, tensor and the first excited scalar masses. If you wanted to compare with people doing analytical or model calculations you would need more than three numbers (one of which has very large errors on it). So what I have shown you should be regarded as a preliminary calculation just to explore what can be done.

- *Maas:*

You showed the graph with the glueball against N, in which the points for $SU(5)$ seem to lie systematically below the extrapolation. What happens with $SU(6)$, $SU(7)$ and say $SU(100)$?

- *Teper:*

Lattice people have to think a bit like experimentalists and have to think of their results like data when they look at them: this means that errors have to be taken seriously into account. Taking the indicated errors into account you cannot say that those points lie systematically below the extrapolation line. There is no evidence for those points to tend away from the straight line. The errors are simply large because the calculation takes a long time and we didn't do it so well. The only way to improve these calculations is to perform them with high statistics, do them over a larger range of lattice spacings to make the extrapolation to the continuum limit more reliable and so get more convincing results. There is however no effect here outside the errors.

- *Kenway*:

How do you ensure that the volumes used in the simulations of $SU(N)$ at different N are the same and therefore do not introduce any systematic effect?

- *Teper:*

We made sure that if l is the lattice size, then $a.l\sqrt{\sigma}$ is greater or equivalent to three and checked by a finite volume study that for these volumes the finite size effects are small.

- *Shuryak:*

Please explain how you look for instantons, and what do you find, especially at the large N you just described.

- *Teper:*

The way we try to see instantons is to take the usual topological charge density which is proportional to $F\tilde{F}x$ and we make a lattice operator out of that. The only problem is that lattice fields are not smooth and you get contributions from various operators which are not the ones you want.

So one has to locally smooth the fields artificially (say by minimizing locally the action of the fields) and one then gets a topological charge distribution which has bumps in it, like instantons. You can certainly identify narrow instantons this way. (If it is narrow, it is a small object and therefore there is a huge peak; there is absolutely no problem in seeing small instantons.) Large instantons are more difficult; and you never know if you are seeing a very large instanton or a very low bump left behind by your smoothing procedure which has nothing to do with an instanton. As you go to larger N in $SU(N)$ theories you expect to lose all your small instantons.

Instantons are theoretically best defined if they are isolated. But if they are really isolated, which will happen if they are small, then $g^2\rho$ is small and if this happens there is a huge exponential suppression and one just loses them.

This is the old argument of Witten according to which instantons are probably not important for physics in $SU(3)$ because they cannot be important at large N and if you got the same physics at large N as in $SU(3)$... Certainly, really isolated instantons will not survive at large N.

- *Shuryak:*

To the issue of instanton sizes: what matters is not a definition of a fm for different N, but simply how the instanton size relates to the distance between them: $R=n^{-1/4}$. It is liquid-like, but still each instanton is not terribly deformed by others.

- *Teper:*

That's right. What you would really like to do is to know how many instantons per configuration there are, which you could get by summing up over the topological distributions. As I was mentioning before in the case of $SU(3)$, I have used "Fermi" physical units because I wanted to say where the glueball masses were. You can convince yourself by various ways that the square root of the string tension can be sensibly set equal to (440 ± 30) MeV in that case. (Most of the error is due to the fact that if I use the ρ meson to set the scale I get one value and if I use e.g. the nucleon I

get another. So this error incorporates the fact that the pure *SU(3)* gauge theory is a different theory from QCD.)

That is the only place where I took the MeV unit seriously: it was when I calculated in *SU(3)* my continuum glueball mass and then I wanted to say what this mass corresponded to in the real world and what $q\bar{q}$ state it might be mixed with. In that case it is important to have some estimate in real physical units. At other times in these lectures, I have used Fermi units only to give some intuition for the scales involved. You can forget these "Fermi" and just express everything in units of the calculated string tension, if you prefer.

- *Korthals Altes:*

If I look at your plot for the instanton distribution as a function of size, I do not see any sign of the distribution bending over at large N. On the other hand at $N \Rightarrow \infty$, it is supported to be peaked at some finite value ρ_0.

- *Teper:*

Well, you cannot really tell it from this calculation. The problem is that if you get a large instanton, then the bump you are looking at is extremely low. In that case we do not know if it is a real instanton or just a bump left over from the smoothing process. So I do not really trust my large instantons. This is not very useful to you, and I have to apologise for that. At some point I have to deal with that problem more carefully.

Symmetries and Quasi-Particles in Hot QCD

Chris P. Korthals Altes
Centre Physique Théorique au CNRS,
Case 907, Campus de Luminy, F13288, Marseille, France

Abstract

These lectures start with a brief overview of salient features of the critical region of hot QCD. The main emphasis is on the accurate description of static plasma observables by the well-known hierarchy of reduced actions combined with 3D simulations above $T \sim 2T_c$. A striking pattern emerges, put in perspective by completing the quasi-particle picture.

1 Introduction

This school gathers experimentalists and theorists in about equal numbers. And that is not more than natural: what else is physics than a playground for experiment and theory? In fact, this is the Galilean concept of physics and the raison d'être of this school!

So I will try to oblige and explain to you the little I came to know over the last years about the plasma state of QCD. With the data from RHIC presently coming out, this is an ideal subject for such a mixed audience. Also, it ties in with the theoretical lectures given by Profs Strassler on flux, Shuryak on QCD at the transition point and Teper on lattice results.

One of the first duties of a theorist is being able to tell the experimentalist what to compare his data with. And the main part of this lecture will be dealing with exactly that: well above the critical temperature, say two to three times its estimated value of 170 MeV, a theorist can predict with good accuracy the behaviour of equilibrium properties of the plasma. The RHIC experimentalist may be disappointed: what he sees is a far cry from a plasma in equilibrium: only recently monojets, increasing proton to pion ratios as function of centrality of the heavy ion collision have been measured and are according to some [1] a sign of a plasma formed at about twice the critical temperature. It may be only at ALICE that we will attain temperatures like a GeV. At any rate it is beyond my competence to discuss the experimental signatures for the production of a plasma in equilibrium during what may be coined as a fleeting moment. In fact the creation of a plasma state at RHIC or ALICE is sometimes compared to a "little bang". But the comparison is somewhat biased: whereas it is clear that the expansion of the universe was involving time scales much larger than the typical time scales for the QCD plasma to come to equilibrium, the time scale of the heavy ion colliders is down by roughly the ratio of the Planck mass and the pion mass. So the reconstruction of the "little bang" from the data remains a daunting task.

So as a consolation of some sort: hot QCD in equilibrium may be useful for the cosmologist...

Despite recent interesting developments in high density QCD I will limit myself to zero density.

In section (2) I will discuss a very simplified version of the QCD transition. This to set the stage for the formal developments based on QCD. Then in section (3) QCD and its global symmetries and order parameters are discussed. Global symmetries are paramount in shaping the phase diagram of QCD. One of these global symmetries is explained in more detail in section (4).

The change in the range of the forces [3] from the hadron phase to the plasma phase is the subject of section (5). The confining force in the hadron phase gets screened in the plasma phase. In QED it is the Coulomb force that gets screened.

In QED one can introduce Dirac monopoles as external sources. One expects no screening at all (after all Galilei observed sun spots, witnesses of the long range of static magnetic fields in the sun's plasma). In QCD monopole sources are screened for any temperature!

In the next section the strategy for doing perturbative calculations is set up. The strategy is the use of a hierarchy of effective actions.

They are the electrostatic action, obtained from QCD by integrating out all hard modes of order T.

The magnetostatic action is obtained from the electrostatic action by integrating out all fields with an electric screening mass. This leaves us with a universal action (only depending through its coupling parameters on the original QCD action parameters). It is universal in its form, containing the magnetic field strength and covariant derivatives thereof. Six years ago Shaposhnikov gave a set of lectures at this school on this subject [10], and you are vividly advised to read those. In the remaining part of the lecture we will concentrate on developments in the last four years. The main topic is the perturbation series as produced by the reduced actions and how well it converges when confronted with non-perturbative lattice simulations. The results seem to indicate that the quasi-particle picture is still a good guide, provided we are willing to accept not only the transverse non-static gluons but also the static transverse (magnetic) gluons as quasi-particles.

In section (7) follow a few examples: Wilson loop, pressure, Debye screening mass, magnetic screening mass and 't Hooft loop.

In section (8) we come back to the gluonic quasi-particles and show how they determine through their flux the behaviour of the spatial 't Hooft loop. In particular, simple scaling laws follow from counting rules which are backed by perturbative calculation. For spatial Wilson loops, analytic calculations are not available. We hypothesize soft transverse gluons follow analogous counting rules, and arrive at a scaling law for the Wilson loop which is verified by lattice simulations.

2 Qualitative picture of the transition

To know better what we are talking about, have a look at fig.(1).

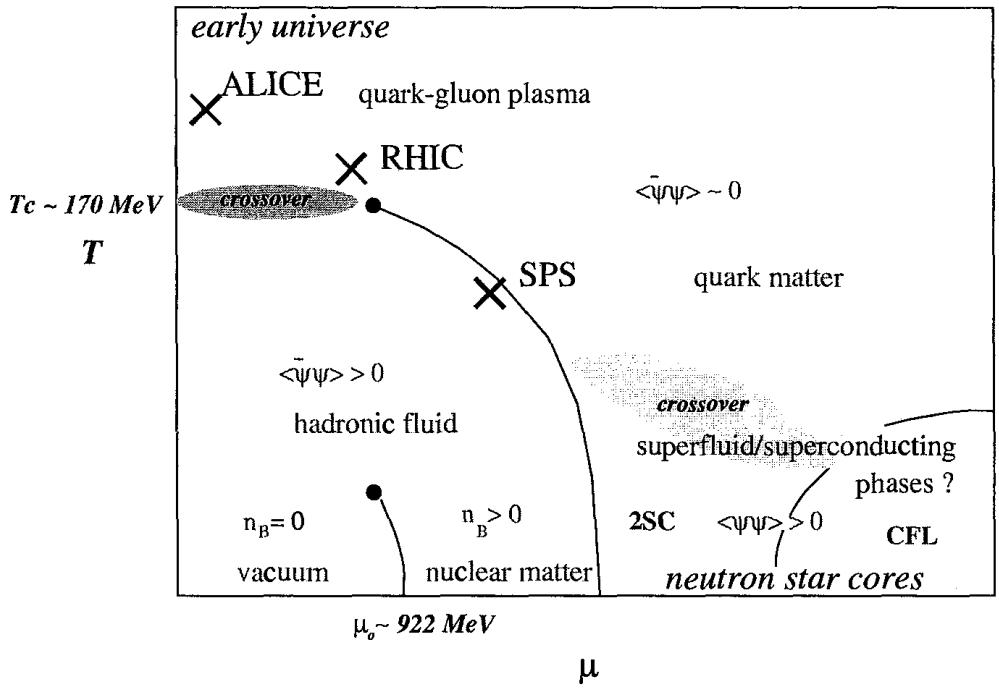


Figure 1: Proposed phase diagram for QCD. 2SC and CFL refer to the diquark condensates defined in ref.[2]. From ref.[5]

It is a schematic phase diagram of QCD as function of temperature and nucleon density, or more precisely the nucleon chemical potential. At the origin we have a groundstate where the quarks and the antiquarks combine into Cooper pairs. This condensate of Cooper pairs is sensitive to temperature and chemical potential changes. You see a familiar transition at zero temperature and chemical potential about 900 MeV: the formation of nuclear matter. For still higher chemical potential we get a a phenomenon called Pauli blocking. At high nucleon density the Pauli principle frustrates the formation of quark-antiquark pairs because the high density of nuclear matter renders all low lying quark levels occupied. So the chiral quark condensate will diminish with growing density. On the other hand at very high densities the gauge coupling becomes so small that perturbation theory is valid. It tells us the pairing of quark-quark pairs is preferred. Then the Cooper instability changes the groundstate into a state of matter where we have instead of a condensate of paired electrons with electric superconductivity a condensate of paired quarks with colour superconductivity [2]. This phenomenon may take place in neutron star cores.

Let us now increase the temperature T. On the vertical axis at zero μ probably a cross-over behaviour results for realistic quark masses. Cross-over behaviour means a gradual change of thermodynamic quantities, like pressure and internal energy. Numerical simulations for realistic quark masses are not yet decisive on this point. Some lattice simulations of QCD [6] [7] indicate a critical point for non-zero nucleon density. At any rate, it is not excluded that the part of the phase diagram dubbed "hadronic fluid" is smoothly connected to the region where ALICE and RHIC will probe the phase diagram.

The continous lines show a first order transition. First order means that quantities like the energy density jump at the transition. At the endpoint, thermodynamics tells us the transition must be second or higher order. In such points first derivatives of the free energy are still continuous (or even higher derivatives). A transition is often caused by a change in the way a global symmetry is realized[1]. We will see more about that in the next sections.

So the diagram shows a rich variety in physics: collider experiments take place near the T-axis, cosmology on the T-axis, and astrophysics at low-T high density. That collider physics and cosmology have small density in common is a fortunate coincidence: one may have a direct bearing on the other and there is a rich litterature on this subject. From now on we concentrate on zero chemical potential.

2.1 A simple model of the transition

The simplest way to see there must be a transition is to take the bag model of hadrons. Increase the temperature up to energies E on the order of the pion mass$\sim$ $140MeV$. The Boltzmann probabilty $\exp -\frac{E}{T}$ for thermal excitation tells us a gas of relativistic pions has formed, with a Stefan-Boltzmann pressure:

$$p_L = 3\text{x}\frac{\pi^2}{90}T^4 \tag{2.1}$$

There is an isospin degeneracy factor 3 in this pressure at low T.

Similarly, coming in from temperatures T much larger than the pion mass, we can expect on the basis of asymptotic freedom a gas of free quarks and gluons. Taking the degrees of freedom into account (for a given number N_f of flavours) one finds:

$$p_H = p_{q\bar{q}} + p_{glue} = 2\text{x}2\text{x}3\text{x}\frac{7}{8}N_f\frac{\pi^2}{90}T^4 + 2\text{x}8\text{x}\frac{\pi^2}{90}T^4 \tag{2.2}$$

Near the critical temperature the bag pressure B of the hadrons is released and adds up to the pressure of the pionic gas. In other words, the individual hadron bags become one large bag. This is typically what percolation is about. Percolation of the pions in the gas means they are starting to overlap with

[1]The Curie point in ferromagnetism is a transition where rotational symmetry is restored. Below the Curie point the ground state of the system is not rotationally invariant: there is a permanent magnetization.

the result that quarks and gluons do not know anymore to what hadron they belong.

So, when comparing the two pressures at T_c one finds:

$$p_L(T_c) + B = p_H(T_c) \tag{2.3}$$

With a bag pressure $B \sim (200\text{MeV})^4$ and $N_f = 2$ one arrives at $T_c \sim 140MeV$.

Here we suppose the bag constant not to vary with temperature. This means that the internal energy density is related to the pressure by $\epsilon_L(T_c) = 3p_L(T_c)$ whereas $\epsilon_H(T_c) = 3p_H(T_c)$. So eq.(2.3) tells us that the latent heat $\Delta\epsilon_c = \epsilon_H(T_c) - \epsilon_L(T_c) = 3B$, a very strong first order transition indeed! Comparing the jump to the mean value ϵ_c one finds $\Delta\epsilon_c/\epsilon_c = O(1)$.

Such spectacular jumps would leave their marks in distributions and correlations of the hadronic decay products.

But we mentioned already a caveat: we supposed the bag constant not to vary with T, and this was what made the transition first order.

So the real question is: what does QCD say about the transition?

3 Global symmetries, order parameters and the phase transition in QCD

The QCD action has as input parameters the experimental values of $\Lambda_{\overline{\text{MS}}}$,the number of colours N, the number of flavours N_f and the masses m_i of the quarks. Together with the Lagrangian:

$$\mathcal{L}_{QCD} = \frac{1}{2}TrF_{\mu\nu}F_{\mu\nu} + \sum_{i=1}^{N_f} \bar{\psi}_i(\gamma_\mu D_\mu + m_i)\psi_i \tag{3.1}$$

these input parameters describe all of hadron physics[2].

This Lagrangian is a strongly coupled system. Its particle spectrum consists of glueballs, and quark bound states. To test this Lagrangian, numerical simulations with a lattice version of QCD are done. This lattice version of the gauge field action is in terms of SU(N) matrices $U(l)$ living on the links l of the lattice. The links have length a, the cut-off in our theory. The field strength matrix $F_{\mu\nu}$ is replaced by the product of the link matrices on every plaquette $U(P) = \Pi_{l\in P}U(l)$. This product is the exponentiated flux through the plaquette:

$$U(P) = \Pi_{l\in P}U(l) = \exp ia^2F_{\mu\nu} + ... \tag{3.2}$$

where we suppose the two sides of P are in the $\mu\nu$ direction and where the dots mean higher derivatives. And the action density is replaced by:

$$TrF_{\mu\nu}^2 \to 1 - \frac{1}{N}ReTrU(P). \tag{3.3}$$

[2] D_μ is the covariant derivative $\partial_\mu + igA_\mu$ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_{nu}A_\mu + ig[A_\mu, A_\nu]$.

The lattice coupling β is related to the bare coupling g by $\beta = \frac{2N}{g^2}$. For more details see the lecture notes by Prof. Teper in this volume.

Global symmetries in QCD depend on how the fermion masses are implemented. Two extremes determine qualitatively what we know for zero nucleon density. All quark masses zero or all infinitely heavy. In the first case left handed quarks ψ_L and right handed quarks ψ_R transform under the symmetry group $SU(N_f)_L \mathrm{x} SU(N_f)_R$. The Lagrangian stays invariant, but the symmetry is realized in the spontaneously broken mode: the left handed quark and its righthanded partner do couple in the real world through a term $\bar{\psi}_L \psi_R$, trace over flavour indices understood. And such a term is only invariant under a left handed symmetry operation in combination with the contragredient right handed partner. This is the diagonal subgroup. The massless Goldstone bosons transform as an adjoint multiplet under this group. Nature provides a non-zero vacuum expectation value of the left right coupling

$$< \bar{\psi}_L \psi_R + \text{h.c.} > \sim (250\text{MeV})^3.$$

It transforms non-trivially under the group, so is a measure of the breaking of the symmetry. It is an order parameter. The Goldstone theorem assures then the existence of an adjoint multiplet of massless pseudoscalars.

We have left out the two $U(1)$ factors. One factor leaves the order parameter invariant and is connected to baryon number conservation. The other factor $U(1)_A$ transforms the condensate. But due to quantum corrections - the axial anomaly - it is not a symmetry of the system and the corresponding Goldstone boson gets a mass due to the instanton mechanism [27].

3.1 Universality

The order parameter Φ of many statistical systems is zero above the critical temperature T_c. Below T_c it is non-zero, so its behaviour is non-analytic. Should it jump at T_c, the transition is called first order. If it is continuous but its first derivative jumps, it's called second order and so on. The order of the transition is the same for a whole class of statistical systems and this is called the universality class of the transition:

- The order of the transition is determined by the symmetry and the dimensionality of the system as described by the order parameter.

So to know the order of the transition of QCD we just take the most general 3D action consistent with the symmetries of QCD one can write down for the order parameter $\Phi_{ij} = \bar{\psi}_{L,i} \psi_{R,j}$. It is the following:

$$\mathcal{L}_{eff\chi} = Tr(\vec{\partial}\Phi^\dagger \vec{\partial}\Phi) + m^2 Tr\Phi^\dagger\Phi + g(Tr\Phi^\dagger\Phi)^2 + hTr(\Phi^\dagger\Phi)^2 + c\det\Phi + \text{h.c} \quad (3.4)$$

The critical behaviour of this action is the same as that of QCD according to universality. For $N_f = 2$ the global symmetry is that of $O(4) \sim SU(2)\mathrm{x}SU(2)$ and is known to be 2nd order. For $N_f = 3$ the determinantal term drives it first order [26].

3.2 Z(N) symmetry

Till now the masses of the quarks were zero. Let us go to the other extreme: infinitely massive quarks. They leave us with only gluons as dynamical agents.

Then there is a global symmetry: $Z(N)$ symmetry [8] with $N = 3$ in case of QCD. $Z(N)$ stands for the subgroup of $SU(N)$ that commutes with all elements of $SU(N)$. It consists of N matrices $z_k 1_N$, 1_N the NxN unit matrix, and $z_k = \exp ik\frac{2\pi}{N}$, k=0,1,...N-1. For notational convenience we'll drop the unit matrix henceforth.

Where does this symmetry come from? In contrast to chiral symmetry, Z(N) symmetry is not a symmetry acting on quantum states. It is a symmetry of the free energy of the system expressed as a path integral.

To get this point we have to understand a basic fact about the description of static phenomena at finite temperature T. Any observable O has a thermal expectation value given by the Gibbs sum:

$$< O >_T = \mathrm{Tr}\, O \exp(-H/T)/\mathrm{Tr} \exp -H/T \tag{3.5}$$

The trace is over physical states only. Physical states are by definition gauge invariant states, that is, invariant under gauge transformations that are regular in configuration space. Of course only gauge invariant observables are admitted. The factor $1/T$ in the Boltzmann factor is like an imaginary time span in a quantum mechanical amplitude. The transcription to a path integral is then straightforward [9]. The trace means the path integral will be periodic in this imaginary ("Euclidean") time for bosons [3].

An immediate consequence is the transcription of Feynman rules. For finite temperature the Feynman rules in Euclidean space undergo one single and simple change: instead of integration over energies, energies are now discrete because of the (anti)- periodicity. For bosons they equal $\omega_n = 2\pi T n$, for fermions $\omega_n = 2\pi(n + \frac{1}{2})T$. In both cases n is integer. This change goes into the propagators, vertices and energy momentum conservation at the vertices.

Long ago 't Hooft [8] realized that the periodicity in time does not necessarily mean you need gauge transformations to be periodic in time. A gauge transformation can be periodic modulo a center group element $\exp ik\frac{2\pi}{N}$ of the gauge group $SU(N)$. The gluon field being an adjoint does not feel any centergroup element. So action and measure of the path integral are insensitive to such a gauge transformation [4].

The order parameter for such a discontinuous gauge transformation is the Wilson line running in the periodic time direction:

$$P(A_0)(\vec{x}) = \frac{1}{N}\mathrm{Tr}\mathcal{P} \exp ig \int_0^{\frac{1}{T}} d\tau A_0(\vec{x}, \tau) \tag{3.6}$$

[3] And antiperiodic for fermions. In this formalism boundary conditions tell an important distinction between bosons and fermions: they tell us the distinction in statistics!

[4] But quark fields are sensitive to the center group: antiperiodic boundary conditions are changed, and hence the statistics.

where the path ordering is defined by dividing the interval $[0, 1/T]$ into a large number N_τ of bits of length $\Delta\tau = \frac{1}{N_\tau T}$:

$$\mathcal{P}(A_0) = \lim_{N_\tau \to \infty} U(\tau = 0, \Delta\tau) U(\tau = \Delta\tau, 2\Delta\tau) \ldots\ldots U(\tau = \frac{1}{T} - \Delta\tau, \frac{1}{T}) \quad (3.7)$$

in an obvious notation. We have dropped the $\vec{x}$ dependence to avoid clutter in the formulae. Every factor is a string bit $U(\tau, \tau + \Delta\tau) = \exp ig\Delta\tau A_0(\tau)$. Every string bit in this product transforms under a gauge transform as $\Omega(\tau)U(\tau, \tau + \Delta\tau)\Omega^\dagger(\tau + \Delta\tau) + O(\frac{1}{N_\tau^2})$. So periodic gauge transforms transform the Wilson line like $\mathcal{P}(A_0^\Omega) = \Omega(\tau = 0)\mathcal{P}(A_0)\Omega^\dagger(\tau = 0)$. And so the trace is invariant. But a discontinuity will multiply the Wilson line with the center group phase z_k, if

$$\Omega_k(\tau = \frac{1}{T}) = z_k^* \Omega_k(\tau = 0) \quad (3.8)$$

Note that a discontinuity other than the center group would be fixed at the time $\tau = \frac{1}{T}$ inside the trace of the Wilson line. Only the center group is a global group, i.e. it does not matter where in time the discontinuity was defined to be.

Although the Z(N) transformation leaves the path integral, hence the free energy invariant, the question whether it commutes with the Hamiltonian of the system makes no sense. There is no such thing as a conserved charge.

This in contrast to canonical $Z(N)$ symmetries that do commute with the Hamiltonian. They can be broken at low temperature but not at high temperature as intuition has it. In section (4) we illustrate this point in QCD.

3.3 Wilson lines, Z(N) symmetry and the deconfining phase transition

The thermal average of the Wilson line is related to the free energy excess ΔF_ψ of a state with a very heavy test quark ψ_i, averaged over all gauge transforms of the state and averaged over the N colour indices i:

$$\exp -\Delta F_\psi / T = \int DA \frac{1}{N} \mathrm{Tr}\mathcal{P}(A_0) \exp -S(A) / \int DA \exp -S(A) \equiv \langle P(A_0) \rangle. \quad (3.9)$$

In appendix A we prove this relation. It is not only valid for a heavy point source in the fundamental representation. A source in any representation r of the group will just change the representation of the Wilson line into r. More precisely, the unitary matrices U in the ordered product eq.(3.7) are in the representation r.

The thermal average of the fundamental Wilson line has been simulated and it is zero at low temperatures, but at T_c it rises abruptly to acquire the value 1 at very high T. A little thought makes this clear, because of its connection to the heavy fundamental charge.

The energy excess equals the energy of the fluxtube pointing from the test charge. As the flux cannot return to the test charge, the length of the fluxtube is typically the spatial size of the box. The energy equals the string tension times the length so is proportional to the size of the box.

In the phase where the flux lines are screened, this energy is finite and will become zero if screening is total.

However we have swept a problem under the rug, that of the short distance effects on the self energy. They are still contained in the thermal average, eq.(3.9), and contribute in terms of the lattice cut-off a:

$$\Delta F_\psi \sim \frac{1}{Ta}. \tag{3.10}$$

For fixed temperature the lattice cut-off goes to zero exponentially fast in the lattice coupling. On the other hand the inverse temperature equals size $N_\tau a$ with N_τ the number of lattice points in the Euclidean time direction. So the free energy excess due to thermal effects is to be corrected for this short distance effect, and to do this is in practice quite intricate [6].

Let us illustrate how one determines the transition, say in SU(3). You can ask the question: what is the distribution of expectation values of the Wilson line $\overline{P}$, averaged over the space volume of our box. Mathematically one asks the probability $E(\tilde{P})$ of a given value $\tilde{P}$ of the line to occur:

$$E(\tilde{P}) \sim \int DA\delta(\tilde{P} - \overline{P(A_0)}\ ) \exp -S(A) \tag{3.11}$$

Under a gauge transform Ω_k as in eq.(3.8) the measure and the action are invariant. Only the line average $\overline{P(A_0)}$ picks up the factor z_k, so

$$E(\tilde{P}) = E(z_k^* \tilde{P}) \tag{3.12}$$

For $SU(3)$ the distribution E is shown in fig.(2) at the transition temperature T_d. The three peaks at the center group values are equally populated, and the figure clearly shows invariance under multiplication with $\exp i\frac{2\pi}{3}$. The central peak at $\tilde{P} = 0$ is a sign that the system likes to be in the hadron phase at the same time. This suggests coexistence of the hadron and deconfined phase at $T = T_d$. At higher T the central peak disappears rapidly and we are left with the three peaks at the center group values. This is confirmed by perturbative calculation of the distribution [24].

So the behaviour of the Wilson line indicates that at low temperature the Z(3) symmetry is restored and broken at high temperature, at first sight counter-intuitive. It seems that at high T the Wilson line spins want to align. This would be understandable if the surface tension between regions where the Wilson lines point in different Z(3) direction becomes very large at high T. The surface tension has dimension (mass)^2. In quarkless QCD there is no scale so the tension must be proportional to T^2. Hence alignment is energetically favourable at high T and the $Z(3)$ symmetry is spontaneously broken[5].

[5]Note that the QCD scale Λ was left out of the argument. Why were we allowed to do so? At

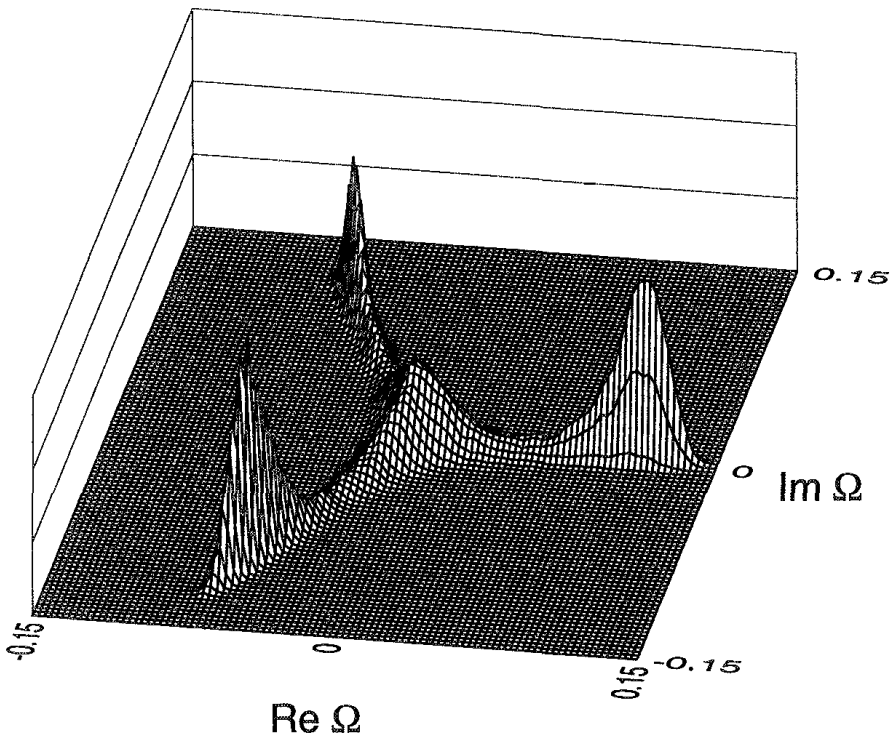


Figure 2: Thermal Wilson line histogram in the SU(3) gauge theory at the deconfining transition point obtained on a $24^2 \times 36 \times 4$ lattice, QCDPAX collaboration[4].

3.4 Z(N) universality

Let us now discuss universality in the context of the Wilson line. The Wilson line is an order parameter and serves therefore to define the universality class [6]. In analogy to the discussion of chiral symmetry, especially eq.(3.4), we now look for a 3D action which has $Z(N)$ symmetry. The Wilson line $P(A_0)$ is now written as a complex number p. In any lattice point $\vec{x}$ we have an independent "spin" $p(\vec{x})$, and it transforms under global $Z(N)$ as

$$p \to z^k p \text{ with } k = 0, 1, 2, .., N-1 \text{ and } z = \exp i\frac{2\pi}{N}. \tag{3.13}$$

So p takes only values in the center group.

Now it is an easy task to write down actions that are Z(N) invariant. For N=2 it is the famous Ising model in 3D, that models spontaneous magnetiza-

high T the Λ parameter is absorbed in the running coupling and is nowhere else present in high T observables (see section(6)).

[6]For a thorough discussion see the lecture notes of Pisarski [31].

tion:

$$S_{N=2} = \beta \sum_{n.n.} (1 - p(\vec{x})p(\vec{y})) \tag{3.14}$$

The sum is over all links that connect two neighbouring lattice points. One sums the Boltzmann factor $\exp -S_N(\{p(\vec{x})\})$ over all configurations $\{p(\vec{x})\}$ to get the free energy per lattice point (V is the number of lattice points):

$$\exp -\beta V f_{N=2}(\beta) = \sum_{\{p(\vec{x})\}} \exp -S_{N=2}(\{p(\vec{x})\}). \tag{3.15}$$

This gives the free energy and a transition is found at a $\beta_c \sim O(1)$. Below this point the order parameter $\langle p \rangle = 0$. This is understandable: at $\beta = 0$ the relative sign of neighbouring spins does not matter in the Boltzmann factor, so disorder will prevail and no magnetization $\langle p \rangle$ results.

Above this point it starts to grow to attain the value 1 or -1 at large β. At large β the spins at the end of any link align because that lowers the action. So the Boltzmann probability will be higher. Whether the resulting magnetization is positive or negative depends on the way we prepare the system. Even an infinitesimal applied magnetic field h (in the guise of a term $h \sum_{\vec{x}} p(\vec{x})$ in the action) will decide about it.

One can induce a region of up-magnetization next to a down-magnetization region by changing the interaction on the links that pierce the wall between the two regions. The change is from $1 - p(\vec{x})p(\vec{y})$ to $1 + p(\vec{x})p(\vec{y})$. The effect of that change or "twist" is for large β that the spins on such links will anti-align, because that optimizes the Boltzmann probability on such links. So this creates a domain wall around the twist with a surface tension $\rho(\beta) \geq 0$ as $\beta \geq \beta_c$. An equivalent way to create the same domain wall is to fix spins to be up at one end of the volume, and down at the other end. In exact analogy, in gauge theory one can fix the thermal Wilson line and compute a "domain wall" tension. Alternatively, one can define a twist in gauge theory most naturally in the lattice formulation.We will do so in later sections

The critical properties of the 3d Ising model have been well established, by numerical means, series expansions etc.. The transition is known to be second order. So magnetization and surface tension go smoothly to zero above the critical point. In particular $\rho(\beta) \sim |\beta - \beta_c|^{2\nu}$, with $2\nu = 1.26....$ And indeed the corresponding transition in $SU(2)$ is second order, it turns out by lattice simulation. In a later section we shall see more manifestations of this universality for $SU(2)$, namely for the surface tension. The surface tension of a wall separating two regions where the thermal Wilson line has opposite signature has been simulated [47] and is shown in fig. (11). Universality is well satisfied by the exponent.

For SU(3) the spin action is Z(3) invariant and reads:

$$S_{N=3} = \beta \sum_{n.n.} (1 - p(\vec{x})p^*(\vec{y}) + \text{c.c.}) \tag{3.16}$$

The first term is obviously Z(3) invariant. The reality of the action reflects the charge conjugation invariance of the SU(3) theory. Charge conjugation

guarantees that the average of P and of P^* is the same. And so does the reality of the spin action for the average of p and p^*.

Now the transition is first order for the spin system. And the SU(3) transition is indeed first order, though weakly so [34]. With weakly is meant that the ratio in the jump in energy over the energy is small, in contrast to what we found in the bag model before.

So universality seems to be well established for N=2 and 3.

Not only the order of the transition but quantities like exponents are identical according to universality. We will come back to those when discussing correlations.

3.5 Universality for large number of colours

Though QCD has three colours, GUT theories feature more, and it is therefore not academic to look at $N \geq 4$. For N=4 and larger we find a lack of predictivity. In fact Z(N) spin models usually comprise different universality classes. Z(2) and Z(3) are rather the exception!

The most general model with $Z(4)$ invariance has two parameters instead of one:

$$S_{N=4} = \beta \sum_{n.n.} (d_1(1 - p(\vec{x})p^*(\vec{y})) + \text{c.c.} + d_2(1 - (p(\vec{x})p^*(\vec{y}))^2)). \tag{3.17}$$

The normalization of β is clearly a convention so we have indeed two physical degrees of freedom.

The two-dimensional phase diagram of this model is well known. It has first and second order transitions as one varies the ratio d_1/d_2. Setting $d_1 = 0$ the resulting model is Ising-like because the remaining interaction term fluctuates between ± 1 and we saw before it was second order. Putting $d_1 = d_2 = d$ gives us an action with value $3\beta d$ if the spins are aligned, and $-\beta d$ if otherwise. This is a class of models - Potts models - which has a first order transition from $N = 3$ on. Depending on the ratio d_1/d_2 the order of the transition changes. In other words there are at least two universality classes in Z(4) spin models , and the question is to which one the SU(N) gauge theory belongs.

Simulations of SU(4) [35] gauge theory show a first order transition, and the same is true for $SU(6)$ [36].

In conclusion: where it is defined, universality works well. For $N \geq 4$ we have to invent extra criteria to pinpoint the universality class in the spin model.

3.6 The phase diagram of QCD

Based on what we learnt above, fig.(3) indicates schematically where one can expect first and higher order transitions. What varies in the diagram is the value of the $m_u = m_d$ mass and the mass of the strange quark m_s. The upper right corner contains the $Z(3)$ transition. As discussed before, it is known to be first order [24]. The deconfining transition T_d is rather high. The lower

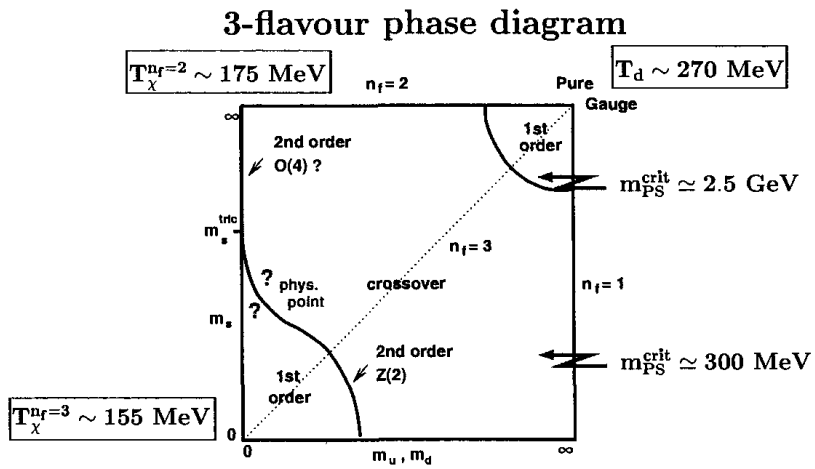


Figure 3: The QCD phase diagram of 3-flavour QCD with degenerate (u,d)-quark masses and a strange quark mass m_s [6].

left corner contains the case of the $N_f = 3$ chiral limit, which is first order as well, according to the renormalization group analysis of eq.(3.4). The transition temperature $T_\chi^{N_f=3}$ is lowest. The case $N_f = 2$, upper left corner, is second order [26]. Its transition temperature is not as low as that of three flavours. So the borderline between first order and crossover ends in a second order point at $m_u = m_d = 0, m_s = \infty$. Gavin et al. [25] find the critical behaviour of the lower left borderline is governed by an effective action with a $Z(2)$ symmetry. This $Z(2)$ symmetry is not present in the original QCD action.

Clearly the determination of the exact location of this line vis à vis the physical values of the quark masses is of paramount importance.

3.7 Chiral and Z(3) orderparameters in flavoured QCD

In fig.(4) the transition region for two flavour QCD is shown. We have only one symmetry, chiral symmetry. So we expect one transition at $T = T_\chi$, where the chiral order parameter drops to zero . There are two striking observations:

- Despite explicit breaking of $Z(3)$ invariance the Wilson line drops steeply below some $T = T_d$.
- The transition temperature is the same for both order parameters: $T_c = T_d = T_\chi$ as peaks of the susceptibilities show.

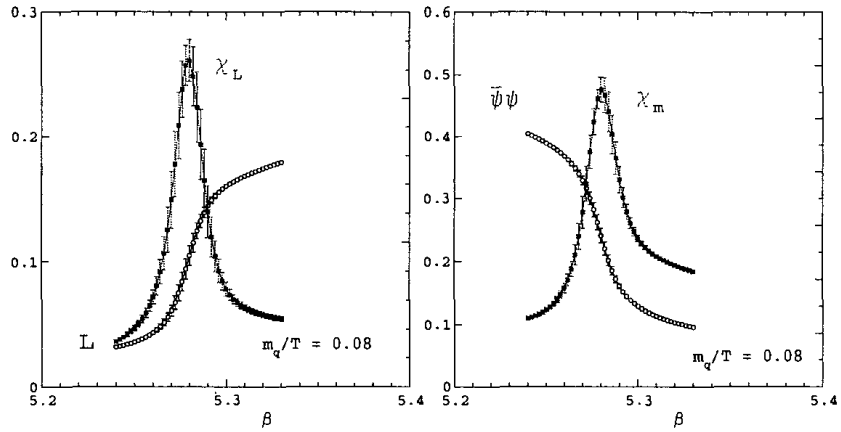


Figure 4: Deconfinement and chiral symmetry restoration in 2-flavour QCD: Shown is $\langle P \rangle$ (left), which is the order parameter for deconfinement in the pure gauge limit ($m_i \to \infty$), and $\langle \bar{\psi}\psi \rangle$ (right), which is the order parameter for chiral symmetry breaking in the chiral limit ($m_i \to 0$). Also shown are the corresponding susceptibilities as a function of the lattice coupling $6/g^2$ [6].

The first point is in seeming contradiction with the expectation that a heavy test quark forms easily a bound state with a dynamical light quark. However, you can argue that in the broken chiral phase the dynamical quarks acquire a mass heavy enough to recover approximate $Z(3)$ symmetry. If so, the Wilson line is a sensible order parameter in the broken chiral phase.

Confinement implies chiral symmetry breaking. After all, confinement implies a bound state of two massless quarks. But in a bound state the quarks must be able to flip their helicity. If so, then chiral symmetry cannot be restored before deconfinement sets in: $T_d \leq T_\chi$. In between, the Wilson line *could* be almost unity with the chiral symmetry still broken! However, Nature tells us that the quark condensate gets unstable above T_d, and that $T_d = T_\chi$. Why is not understood.

In sharp contrast, if quarks are in the adjoint representation [29] the system has two exact symmetries, $Z(3)$ and chiral symmetry. So two different transition temperatures are expected. Flux tubes cannot end on adjoint matter, so form glueballs. The region in between has no glueballs anymore, but still a fermion condensate and a hadron spectrum. The adjoint fermion condensate stays stable till $T_\chi \sim 8T_d$ [29].

4 Canonical Z(N) symmetries in SU(N) gauge theory

In this section we start with QCD in 3+1 dimensions. We then render one of the *spatial* directions periodic and study the effects due to varying its size. The results will be of later use, especially in section (7). Then we will switch on the temperature and see what happens.

First we fix some general notions.

4.1 Electric and magnetic fluxes

Below we give a quick review of electric and magnetic fluxes and their free energy. Let only fields with trivial N-allity couple to the gauge fields.

4.1.1 Electric fluxes

For example, take the y-direction periodic mod L_y and variable. The periodic τ direction is supposed to be very long $L_\tau = \frac{1}{T} \to \infty$, as well as x and z directions.

To explain the essentials we will first put $L_x = L_z = 0$ and consider the two-dimensional system.

Then a time independent gauge transformation Ω_k is allowed to be discontinuous mod z_k as in eq.(3.8) but now in the periodic y-direction:

$$\Omega_k(y + L_y) = z_k^* \Omega_k(y), k = 0, 1, .., N-1. \tag{4.1}$$

An example of such a transformation is:

$$\Omega_k = \exp -i \frac{y}{L_y} \frac{2\pi}{N} Y_k \tag{4.2}$$

with $Y_k = \mathrm{diag}(k, ..., k, k-N, ...k-N)$, the diagonal k-hypercharge, an NxN traceless matrix with $N-k$ entries k and k entries $N-k$. It has the property that

$$\exp i \frac{2\pi}{N} Y_k = z_k \tag{4.3}$$

is a center group element. The Y_k are a natural generalization of the familiar hypercharge Y_1. They span the Cartan subalgebra, which consists of all elements of the Lie algebra su(N) that can be diagonalized simultaneously. It is N-1 dimensional. There is a lattice L_c of elements in this subalgebra that give upon exponentiation a center group element. So the elements Y_k are special points on this lattice. They have an important property: let q be a number between 0 and 1. Then the elements qY_k trace a rectilinear path in the Cartan algebra on which only begin and end points correspond to centergroup elements (respectively 1 and z_k). Do this for all elements with $k = 1, ..., N-1$. Then we have an elementary cell of the lattice L_c. The reader can see by inspection

that this is correct for $N = 3$ and $N = 4$. This is the property that is important for the $Z(N)$ invariant Wilson line distribution $E(\widetilde{P})$ and will be used throughout the dynamical calculations in subsection (7.3).

The gauge transformation $\Omega_k = \exp i\omega_k$ is represented in Hilbert space by $\exp i \int d\vec{x} Tr \vec{D}\omega_k.\vec{E}$. As operator, Ω_k commutes with all local gauge invariant operators, in particular the Hamiltonian. It does not commute with the Wilson line in the y-direction:

$$\Omega_k P(A_y)\Omega_k^\dagger = z_k P(A_y) \tag{4.4}$$

It is important that on the physical Hilbert space these operators have a unique effect, *only* depending on the discontinuity. To understand this take Ω_k and Ω'_k with the same value for k. Form the product $\Omega_k^\dagger \Omega'_k$. In this product the singular behaviour drops out: both transformations belong to the same equivalence class through a regular gauge transformation. And a regular gauge transformation leaves a physical state invariant. The product of two elements from equivalence classes k and k' gives the equivalence class k+k' mod N. And finally Ω_k^N is regular.

As a consequence the eigenphases must be of the form $\exp ik\frac{2\pi}{N}e$. The number e is integer and conserved mod N.

And the physical Hilbert space divides into N orthogonal subspaces $\mathcal{H}_e$, e integer mod N, on which Ω_k is diagonal with eigenvalue $\exp ike\frac{2\pi}{N}$. The projector P_e on such a subspace is given by

$$P_e|phys>= \frac{1}{N}\sum_k \exp(-ike\frac{2\pi}{N})\Omega_k|phys\rangle. \tag{4.5}$$

And since the Wilson line $P(A_y)$ in the y-direction obeys $\Omega_k P(A_y)\Omega_k^\dagger = z_k P(A_y)$ a state with charge e can be written as a state with $e = 0$ and with the line $P(A_y)$ wrapping e mod N times around the y-direction. So e is the promised conserved charge, and counts the number of "strings" or Wilson lines wrapping around the y-direction (mod N). There is a free energy F_e for each of these electric flux sectors, defined by tracing only over the physical states of a given sector:

$$\exp -L_\tau F_e = Tr_e \exp -L_\tau H. \tag{4.6}$$

These free energies can be inferred from simulations on the lattice. First we need a formula relating the F_e to partition sums $Z_k^{(e)}$ [8] . To this end substitute eq.(4.5) into eq.(4.6) and rewrite the latter as:

$$\exp -L_\tau F_e = \frac{1}{N}\sum_{k=0}^{N-1} Z_k^{(e)} \exp -ike\frac{2\pi}{N}. \tag{4.7}$$

The partition functions on the r.h.s. of this equation are now Gibbs sums over physical states, with the operators Ω_k acting:

$$Z_k^{(e)} = Tr_{phys}\Omega_k \exp -H/T \tag{4.8}$$

To understand the physical meaning of the partition functions there is an alternative definition of the operator Ω_k. It is *only* valid on the physical subspace, where it reads:

$$\Omega_k = \exp i\frac{4\pi}{gN} Tr E_y(y_o) Y_k. \tag{4.9}$$

Only on the physical subspace the two are identical. In fact they differ by a regular gauge transformation, as you can infer from the exercise below.

- Show that the operator (4.9) has the same effect on the Wilson line $P(A_y)$ as Ω_k in eq.(4.4).
- Show that any regular gauge transformation Ω of Ω_k in (4.9), $\Omega\Omega_k\Omega^\dagger$, has the same effect on $P(A_y)$. This means that physical states stay physical states after applying Ω_k.

So the discontinuous gauge transformation is brought about by a single dipole of strength Y_k at the point $y = y_0$.

This generalizes easily from d=2 to d=4, adding the x and z dimensions. The single dipole at y_0 becomes a dipole sheet on the (x,z) surface at the point $(y_0, \tau = 0)$ as representing the operator Ω_{k_y} (see fig.(5)). We have added the suffix y on k to distinguish it from a similar operator in z or x direction. Once this is done we have to admit not only operators $\Omega_{\vec{k}}$ labeled by the vector $\vec{k} = (k_x, k_y, k_z)$, each component running from 0 to N. Obviously, we also have fluxes $\vec{e} = (e_x, e_y, e_z)$ and the connexion between the free energy $F^{(e)}_{\vec{e}}$ and the partition functions $Z^{(e)}_{\vec{k}}$ generalizes to:

$$\exp -L_\tau F_{\vec{e}} = \frac{1}{N^3}\sum_{\vec{k}} Z^{(e)}_{\vec{k}} \exp -i(\vec{k}.\vec{e}\frac{2\pi}{N}) \tag{4.10}$$

And so we have now a physical interpretation of the partition function eq. (4.8) as the thermal average of the electric dipole sheet. It monitors the electric flux activity in the system as we will see in sections (7) and (8).

Note that the partition functions with an electric twist are related through a Fourier transform to the free energies $F^{(e)}_{\vec{e}}$. They do not define by themselves a free energy as they are off-diagonal matrix elements. 1

4.1.2 Magnetic fluxes

Of course, one can define partition functions from physical states with a magnetic vortex line running along a space direction, say the z direction. In continuum language the operator creating such a vortex is:

$$V_k = \exp i\vec{D}(A) v_k(x,y)\vec{E} \tag{4.11}$$

with $v_k(x,y) = \arctan(\frac{y}{x})\frac{1}{N}Y_k$. Remember Y_k is the generalization of the colour hypercharge we introduced in eq. (4.3). When encircling the point $x = y = 0$ the gauge transformation $\exp iv_k(x,y)$ picks up a factor $\exp i\frac{2\pi}{N}Y_k = z_k$. This

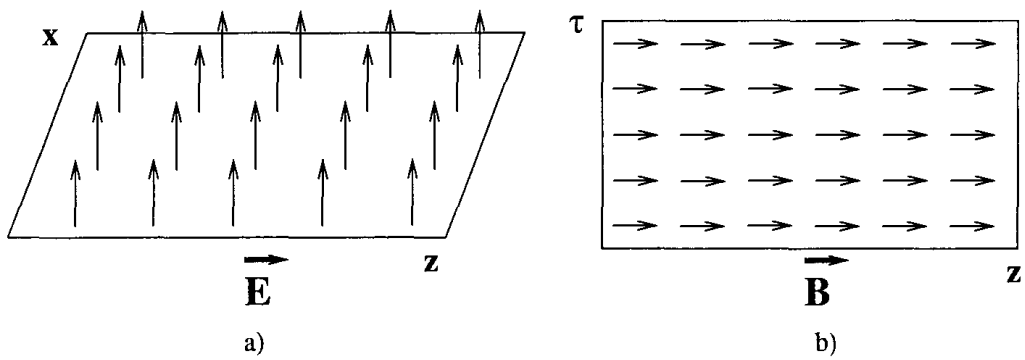


Figure 5: Electric twist a) and magnetic twist b) partition functions. The electric twist is a dipole sheet at fixed (τ, y). The magnetic twist is a Dirac flux in the z-direction, propagating in time, at fixed (x,y). In lattice language the plaquettes $P_{\tau,y}$ are twisted in a), $P_{x,y}$ in b).

gauge transformation remains unchanged along the z-direction. We say that V_k creates a vortex or "Z(N) Dirac string". That means, a Wilsonloop W in the fundamental representation that encircles the vortex will pick up the z_k factor:

$$V_k W V_k^\dagger = z_k W. \tag{4.12}$$

Any Wilsonloop with non-zero N-allity l will pick up a factor $(z_k)^l$. But Z(N) neutral loops will not sense the Z(N) Dirac string, hence the name.

This Dirac string is then repeated in every τ slice, tracing out its history. The notation for those magnetic partition functions is $Z_k^{(m)}$, with

$$Z_k^{(m)} = Tr_{phys} V_k^\dagger \exp -H/T V_k \equiv \int DA \exp -S_{\vec{k}}(A). \tag{4.13}$$

They define directly a magnetic free energy $Z_k^{(m)} = exp{-L_\tau F_k^{(m)}}$, being diagonal matrix elements of the Hamiltonian. Note the difference with the electric twist partition function $Z_k^{(e)}$ in eq.(4.8).

The situation is shown in fig.(5 b). The vortex creates a singular dipole field $\vec{B}$ along the z-direction in the Y_k colour direction.

One can define simultaneously define Dirac fluxes in x, y and z directions with strengths $\vec{k} = (k_x, k_y, k_z)$.

Still we need to understand how to simulate numerically the partition functions in the form of a path integral. For the magnetic partition function this is immediate. The lattice Hamiltonian consists of plaquettes. Those that encircle the vortex will, like the Wilson loop in eq. (4.12), pick up a z_k factor (or a z_k^* factor, depending on the orientation of the plaquette). So the magnetic twist partition function is simulated by replacing the usual lattice action by the lattice action with a twist. Replace along the string:

$$1 - \frac{1}{N} TrU(P_{x,y}) \to 1 - z_k(\frac{1}{N} TrU(P_{x,y})). \tag{4.14}$$

and repeat this for all time slices.

For the electric twist partition function one just exchanges x and τ. As the plaquettes $P_{y,\tau}$ deliver the electric field in the continuum limit, it is intuitively clear that this prescription will give the thermal average of the electric dipole sheet.

4.1.3 Behaviour of flux free energies in the confined phase

The behaviour of the electric and magnetic free energies is quite different in the confined phase, where all sizes are macroscopic.

When the size L_y of the periodic direction is macroscopic, the VEV of the Wilson line in the y-direction is zero. The system is confining with string tension σ. This means that $F_e - F_0 = \sigma_e L_y$ and states with $e = 0$ mod N are energetically favoured. The energy $E_e = \lim_{L_\tau \to \infty} F_e$ of a state with $e = 0$ mod N is the lowest, all others are higher by an amount $\sim \sigma L_z$, and the symmetry is "restored" because we have one unique ground state. Only the space $\mathcal{H}_{e=0}$ is of importance for confining physics. It contains the glueball states, the localized eigenstates of the Hamiltonian. The periodicity in e mod N comes about because N strings decay into glueballs.

The magnetic free energy is decaying exponentially fast [8]:

$$F_{m_z}^{(m)} - F_0 \sim \sigma L_z \exp -\sigma L_x L_y.$$

This means the magnetic flux is screened. We will come back to this type of screening in the next section.

4.1.4 A simple property of electric and magnetic twisted partition functions

On the other hand the partition functions have a simple property. Suppose only one size L becomes small (meaning of hadronic size or smaller), whereas the others stay macroscopic. Consider any partition function Z_k with one single twist of strength k like in fig.(5a) or b)). If L corresponds to one of the directions orthogonal to the planes shown (y or τ in a), x or y in b)) then Z_k obeys an area law with the area as shown in the figure, and a *universal* coefficient ρ_k:

$$Z_k \sim \exp\left(-\rho_k Area\right). \tag{4.15}$$

In the lattice formulation of the twist this universality is just a consequence of the Euclidian invariance under exchange of the τ and x axis of fig. (5) a) into b) and vice versa. The function ρ_k is computed perturbatively in subsection (7.3) in terms of the running coupling $g(1/L)$.

4.1.5 Behaviour of the partition functions in the hot phase

Physically one would expect that electric flux free energies will show screening behaviour. And this is verified easily by using the simple property of the partition function mentioned in the previous section at high temperature. All partition functions with a twist in the time direction will behave according to eq. (4.15). These are precisely the partition functions appearing in the Z(N) Fourier transform that leads to the electric flux free energies in eq. (4.10):

$$\exp -L_\tau F^{(e)}_{\vec{e}} = \frac{1}{N^3}\sum_{\vec{k}} Z^{(e)}_{\vec{k}} \exp -i(\vec{k}.\vec{e}\frac{2\pi}{N}) \tag{4.16}$$

With the partition functions on the right hand side decaying like $\exp -(\rho_k Area)$ one can easily infer that the free energy differences $F_{\vec{e}} - F_{\vec{0}}$ decay exponentially as well.

- Deduce that $F^{(e)}_{e_x,0,0} - F_{\vec{0}} \sim (1-\cos(e_x\frac{2\pi}{N}))\exp -\rho_1 L_y L_z$. Use eq.(4.16) and that ρ_1 is the smallest tension (as will be shown in section (7)).

So the electric fluxes behave radically different from the confining phase. They become exponentially fast degenerate, and the electric Z(N) symmetry is spontaneously broken.

The magnetic flux energies behave qualitatively the same as in the low T phase: they stay screened.

4.2 Breaking canonical Z(N) symmetry

What happens as L_y becomes smaller? In fact, if L_y becomes on the order of the hadron size R we have a transition just as we had before with the inverse temperature and L_y interchanged. So we expect the Wilson line $P(A_y)$ to acquire a VEV. But the physics will be that of 2+1 dimensional Yang-Mills plus the degrees of freedom of A_y that do not depend on y anymore, i.e. an adjoint Higgs. In the 2+1 dimensional space, A_y is a scalar with respect to rotations in the x-z plane. But there will still be a string tension σ.

So what *has* changed qualitatively in this phase?

- In the x-z plane domain walls, or rather domain "lines", appear: they separate two regions where the spatial Wilson line, curled up in the y-direction, has different $Z(N)$ values by a factor $\exp ik\frac{2\pi}{N}$.

  These walls have a tension $\rho_k(M)$ which is perturbatively calculable with the 4d running coupling $g(M)$ ($M = 1/L_y$ is any mass scale larger than the critical one M_c).

The tension $\rho_k(M)$ is computed from the normalized twisted magnetic partition function $\hat{Z}^{(m)}_k = Z^{(m)}_k / Z^{(m)}_0$. At small enough L_y it behaves as $\hat{Z}^{(m)}_k = \exp -\rho_k L_x L_\tau$. This is the expression for a domainwall stretching in the x-direction, with energy density or tension ρ_k.

What is obvious without calculation is that the tension of the wall is

$$\rho_k = d_k(1/g(M))M^2 \tag{4.17}$$

from dimensional reasoning and the fact that the calculation is semi-classical. Note the presence of $\frac{1}{g}$, not $\frac{1}{g^2}$ as in semi-classical instanton calculations.

Its width will be $O((g(M)M)^{-1})$. The calculation will be done in detail in section (7).

So we find that the magnetic flux free energy in x and z direction is no longer screened! The domain lines are made of unscreened magnetic flux.

This summarizes the effect of the breaking of canonical $Z(N)$ symmetry.

4.3 Intrinsic Z(N) symmetry in 2+1 dimensional Yang-Mills

In the 3d Yang-Mills system there is also an *intrinsic* canonical $Z(N)$ symmetry as opposed to the "extrinsic" Z(N) symmetry discussed above. It was discovered long ago [8] and is due to the appearance of magnetic vortices in 2+1 Yang Mills theory. Hence the name "magnetic Z(N)" symmetry with as order parameter the vorticity.

We give here a quick recapitulation of how this symmetry is realized and its relation with confinement in 2+1 dimensions [8] [51]. The results are going to be useful in section (7).

A vortex is created by a gauge transformation $V_k(x,z) \equiv \Omega_k(x,z)$ with a discontinuity $\exp ik\frac{2\pi}{N}$ across a line starting from the point (x,z)). The discontinuity is not seen by the adjoint fields. $V_k(x,z)$ has a purely local effect. Only when we surround it by a Wilson loop in the fundamental representation it gives a phase factor to the loop:

$$V_k(x,z)W(C)V_k(x,z)^\dagger = \exp ik\frac{2\pi}{N}W(C) \tag{4.18}$$

if the point (x,z) is inside the loop C.

So $V_k(x,z)$ creates excitations that have a charge mod N with respect to a Wilson loop that surrounds the whole 2d system. This large Wilson loop is the generator of this intrinsic Z(N) symmetry. The large Wilsonloop commutes with the Hamiltonian. As a consequence the vorticity is conserved mod N.

If this symmetry is realized in the spontaneously broken mode, $< V_1 > \neq 0$, then we have N equivalent ground states. Each of these ground states corresponds to a different orientation of the VEV and they are all mutually orthogonal.

Pick a given ground state $|k>$ with $< k|V_1|k >= v\ z_k$, $0 < v \leq 1$.

Then the action of $W(C)$ transforms it into a state $W(C)|k>$ where inside the perimeter C the VEV of V_1 corresponds to that in the state $|k+1>$ because of eq.(4.18).

- Show that V_k commutes with $|W(C)|^2 \equiv W^\dagger(C)W(C)$. Deduce that the unitarized Wilson loop $\widetilde{W}(C) \equiv |W(C)|^{-1}W(C)$ has the same commutation relation with V_k as $W(C)$.
- Find $< k|\widetilde{W}(C)^\dagger V_1 \widetilde{W}(C)|k >= z_1 < k|V_1|k >$, if the vortex operator acts inside the contour C.

That is, the loop creates a domain "wall" [8] between the two groundstates.

What is the typical energy density and width of this wall? In 3d Yang-Mills theory there is a dimensionful coupling g_3 with its square having dimension of mass. So a dimensional argument leads to an energy density of g_3^4 and the width $1/g_3^2$.

The VEV of the loop, $< k|W(C)|k >$, consists of the overlap of the state turned into $|k+1>$ in the inside of the loop. So if we make the loop larger and larger the orthogonality of the ground states tell us the VEV goes to zero. That it decreases as fast as the exponent of the area follows from closer inspection of the overlap [51]. The overlap consists of a product of local overlaps $\gamma(\vec{x})$ between the vortices in a given point $\vec{x}$ inside the loop:

$$< k|W(C)|k >= \Pi_{\vec{x}\in S(C)}\gamma(\vec{x}) = exp{-\sigma S(C)} \tag{4.19}$$

This argument is correct to the extent that the vortices interact only locally, typically over a distance defined by the dimensionful coupling g_3^2. The tension $\sigma \sim g_3^4$ again on dimensional grounds.

Recall that our 2+1 dimensional system is embedded in a 3+1 dimensional world where the z-direction is periodic and of size M^{-1}. The 2+1 gauge coupling expressed in terms of the 4d coupling is:

$$g_3^2 = g^2(M)M. \tag{4.20}$$

So for large M the gauge coupling $g(M)$ is small.

This means that the energy of the domain walls due to breaking of intrinsic Z(N) is parametrically a factor $g(M)^5$ smaller than the tension due to the periodic z-direction in eq.(4.17)! And the width of the wall created by the Wilson loop is of the order of $(g(M)^2M)^{-1} >> M^{-1}$ i.e. much larger than the extra periodic dimension. From this one concludes that, once extrinsic Z(N) is broken, the system is essentially 2+1 dimensional. This hierarchy of scales is consequential for the next subsection.

4.4 The fate of broken Z(N) at high temperature

Let us now heat up this 2+1 dimensional system. For the case of two colours the procedure is illustrated in fig.(6). It shows a simulation of 3+1 dimensional SU(2) gauge theory with τ and y directions periodic and variable, of length $L_\tau = 1/T$ and $L_y = 1/M$ respectively [49].

The 3+1 dimensional theory lies in region D. Simulated are the thermal Wilson lines P_τ and the temperature $T_c(M)$ where it becomes non-zero. The

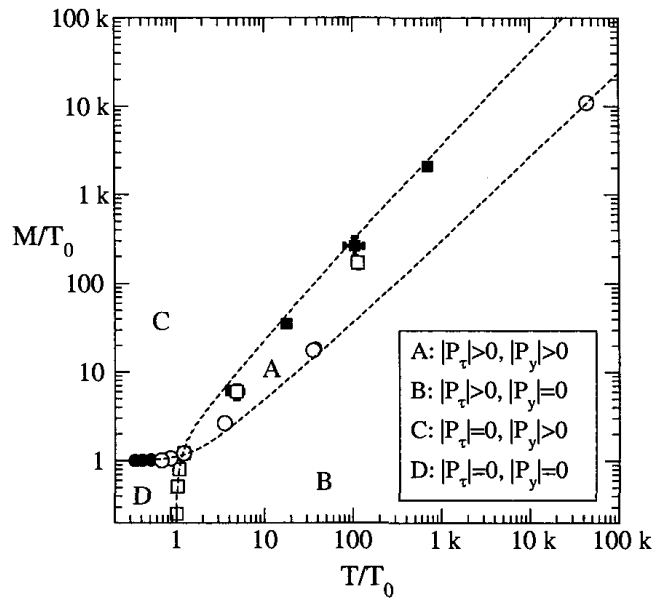


Figure 6: The 4d phase diagram in units of $T_0 \equiv T_c$, as shown by Montecarlo data, ref. [49]. The data show the points where the Wilson lines P_τ and P_y undergo the transition. Note the symmetry $\tau \leftrightarrow y$.

transition line $T_c(M)$ starts at the $M = 0$ axis at $T = T_0$, rises steeply and bends to the right. Also shown is the Wilson line P_y with its critical behaviour. The lattice data are shown as circles and the broken lines are for our purpose here just fits to the data. The two broken lines are mirrored through the diagonal. This should be obvious: the loci of the two types of transition cannot distinguish between τ and y: $P_\tau(T, M) = P_y(M, T)$.

Start from region C on the vertical axis somewhere above $M = T_0 = T_c$ in fig.(6). This portion of the vertical axis is the cold system in the broken "extrinsic" Z(N) phase. Here $P_y > 0$ and also the VEV of the 't Hooft vortex operator V_k. As we increase T along this line we first cross the broken line into region A where also the thermal Wilson line P_τ gets a VEV. But the VEV $< V_k >= 0$, as was argued in ref. [43] [44]: one can show with the arguments of subsection (7.3) that the correlation $< V_k(\vec{x})V_k(\vec{y})^\dagger >\sim \exp -m|\vec{x} - \vec{y}|$ for large separation, as soon as the thermal Wilson line becomes non-zero. So this is the region where the intrinsic Z(N) is restored.

But the extrinsic Z(N) is still broken because of the difference in energy scales. Note that the transition of the essentially 2+1 dimensional system is at a *higher* temperature than the 3+1 dimensional system. This is intuitively reasonable. After all, in 1+1 dimensions the transition is at infinite temperature.

Finally increasing T even more, the restoration of the extrinsic Z(N) takes place at the second crossing of the horizontal line into phase B. Then we are in

the high temperature phase B of 3+1 dimensional QCD.

The symmetry in the figure between phase B and C is deceptive from a physics point of view. In phase C there are one dimensional "domain" walls tracing out a two dimensional sheet in time τ. They do separate regions where the Wilson lines P_y have different center group values and an observer in the (x,z) world can observe those walls with their high energy density. Phase B is the hot QCD phase where there are regions with different center group values for P_τ, separated by two dimensional sheets as discussed at the end of subsection (3.3). These sheets cannot be interpreted as domain walls because they do not extend in time as solitons. In section (8) their role will be seen to be that of detectors of electric flux. The tension ρ_k is indeed symmetric with respect to the diagonal.

The reader may ask the justified question: how do the currently popular extra periodic space dimensions of size $1/TeV$ to our four dimensional world fit into this picture? The answer is simple and expected: for 5D SU(N) gauge theory, phase C now contains *two-dimensional domain walls*! The energy density of these walls is typically $(TeV)^4$. The form of the phase diagram is qualitatively the same. Hence the extrinsic Z(N) symmetry gets restored at high T and the walls put constraints on cosmological models [50] [49].

5 Forces and screening in the plasma

Till now we did not mention what happens to the forces in the QCD plasma.

In this section we will first discuss Debye screening on a perturbative basis. For QCD this will turn out to be insufficient. We will need a non-perturbaitve definition. Then in the next section an operator formalism is presented, needed to systematize the latiice calculations.

Fianally magnetic screening is defined. This is a new and important aspect of thermal QCD.

5.1 Electric screening

In a QED plasma one would like to know what happens to the electric field due to a heavy charge Q and to the Coulomb force between two static charges of opposite signature at distance r in the z-direction. Let us look at scalar QED, as it shows in leading order in the coupling some features in common with QCD.

The Coulomb force is transmitted by the the A_0 potential. The propagator of A_0 is renormalized by the one loop scalar and seagull diagram and gives the self energy, as calculated by the Feynman rules discussed below eq.(3.5):

$$\Pi_{\mu\nu}(p) = e^2 T \sum_{l_0} \int \frac{d\vec{l}}{(2\pi)^3} \left(\frac{(2l_\mu - p_\mu)(2l_\nu - p_\nu)}{(l-p)^2 l^2} - 2\frac{\delta_{\mu\nu}}{p^2} \right) \tag{5.1}$$

The self energy is transverse, $p_\mu \Pi_{\mu\nu}(p) = 0$. It has two independent tensors, that we choose to be Π_{00} and $\Pi_{\mu\mu}$. For $T = 0$ they are proportional.

For the Coulomb force we are interested in the static limit $p_0 = 0$. We resum all the self energy bubbles to get the propagator and the static part of the $< A_0 A_0 >$ propagator becomes :

$$\frac{1}{\vec{p}^2} \to \frac{1}{\vec{p}^2 + \Pi_{00}(p_0 = 0, \vec{p})} \tag{5.2}$$

To find the pole to lowest order in the coupling, we let $\vec{p} \to 0$ and find $\Pi_{00}(0, \vec{0}) = e^2 T^2/3$. We use dimensional regularization so that the $l_0 = 0$ contribution to Π_{00} is $\sim \int d\vec{l} \frac{1}{\vec{l}^2} = 0$. Only hard modes proportional to T inside the loop contribute to the pole mass!

In configuration space this leads to Coulomb screening:

$$\frac{1}{r} \to \frac{1}{r} \exp -m_D r \text{ with } m_D = \frac{e^2 T^2}{3}. \tag{5.3}$$

In scalar QED the self energy is gauge independent to all orders in perturbation theory. And so you can start to evaluate the corrections to the Debye mass by computing the corrections to the pole location. Indeed, one can reformulate perturbation theory by adding the screening mass term to the action of scalar QED, and use the Feynman rules with the modified static propagator above:

$$S_{QED} = \{S_{QED} + m_D^2 \int_{\vec{x}} A_0^2\} - m_D^2 \int_{\vec{x}} A_0^2 \tag{5.4}$$

To avoid double counting you have to subtract the screening term as well, and use it as an insertion whenever you have a self energy subdiagram. This procedure leads to a well-defined perturbation series. However the powerlike infrared divergencies are now cut-off at $m_D \sim eT$ so we can expect terms in the series like $e^4/m_D \sim e^3$, i.e. odd powers in the coupling. But otherwise the series can be computed to arbitrary order by taking the Debye screening into account.

The non-abelian case to lowest order is qualitatively the same. The Debye mass changes only by replacing $e^2 \to g^2(N + N_f/2)$. The salient differences are:

- In the non-Abelian case the self energy is not gauge independent. For the pole location one can argue that it is gauge independent.
- More important, already in next to leading order an infinity of diagrams in the static sector contributes [19] [30].

This means that an observable is needed to define the screening mass *independently* of perturbation theory.

Indeed, there exists a natural definition in terms of the correlation of two static charges in the fundamental representation. From the results of the previous sections and appendix A we can write this as the correlator of two Wilson lines:

$$< P(r)P(0)^\dagger >\equiv \exp -\frac{F_E(r)}{T} \tag{5.5}$$

This can be simulated on the lattice by non-perturbative means. To this end one takes a lattice periodic in all directions, and the Wilson lines separated over a distance r in the z-direction.

In the confining phase, below T_c, this correlator falls off at long distances as $\exp -\frac{\sigma(T)}{T} r$, due to the string tension $\sigma(T) \equiv V_T(r)/r$. Above T_c the string tension gets screened by the Debye mass and the potential becomes:

$$F_E(r) = F_{E0} - \frac{c_E}{r} \exp -m_D r. \tag{5.6}$$

The parameters in this free energy are only depending on T.

5.2 An operator formalism as a bookkeeping advice

The path integral of the spatial correlator, eq. (5.5), can be read in an alternative way. Consider the fictitious Yang-Mills Hamiltonian in the (x, y, τ) space, with its physical Hilbert space. This Hilbert space will contain eigenstates of the Hamiltonian, which are different from the one in which we live. One space direction, τ, is finite and periodic with period $1/T$. That means that the rotation group in these three dimensions is reduced to $SO(2)$ times the discrete group admitted by the periodic finite τ direction. The path integral reads in terms of this Hamiltonian H and the its physical Hilbert space:

$$< P(r)P(0)^\dagger >= (Tr_{phys} \exp -H\hat{L}_z P \exp -HrP^\dagger \exp -H\hat{L}_z)/(Tr_{phys} \exp -HL_z). \tag{5.7}$$

The Wilson lines P are now expressed in terms of the canonical operators A_0, $\hat{L}_z = (L_z - r)/2$, and $L_z >> r$. In the limit of $L_z \to \infty$ the correlation becomes a some of exponential decays:

$$< P(r)P(0)^\dagger >= \sum_n |\langle 0|P|n\rangle|^2 \exp -m_n r. \tag{5.8}$$

The mass gap m_n is the value of the energy compared to the groundstate energy, at zero momentum $p_x = p_y = p_\tau = 0$.

We want an efficient bookkeeping system for the states excited by the Wilson line (and eventual other observables).

To this end we define in our fictitious Hilbert space, apart from charge conjugation, a parity transform P, under which *only* the y-direction changes sign (and hence only E_y and A_y). Remember the rotation group is only SO(2), in which simultaneous sign flipping of x and y is a rotation.

Another useful quantum number is R-parity: it changes τ into $\frac{1}{T} - \tau$, and A_0 into $-A_0$[7].

[7]The combination of time reversal T ($A_0 \to A_0^t$, $A_i \to -A_i^t$) and charge conjugation C in the usual Euclidean version of the theory gives an operation $R = TC$ whose sole effect is to give a minus

So we have as symmetry group $SO(2) \times Z_2(R) \times Z_2(P) \times Z_2(C)$, hence sates labeled by J_R^{PC}. The $SO(2)$ group is generated by $J = x\partial_y - y\partial_x$. So we have $PJP = -J$. Look at any state $|j\rangle$ with $j \neq 0$. Then the state $P|j\rangle$ has $J = -j$ so is orthogonal to $|j\rangle$. This means we can form the parity doublet, degenerate in energy,

$$|j;\pm\rangle = (1 \pm P)|j\rangle. \tag{5.9}$$

The argument breaks down for spin zero states.

Now our Wilson line operator excites spin zero states. Also, under C and R

$$P(A_0) \to P(A_0)^\dagger. \tag{5.10}$$

Under parity the Wilson line stays invariant.

Arnold and Yaffe [41] noted that the potential consists of *two* channels of exponential decays! One is governed by the correlation of ImP and the other by that of ReP. The first is odd under charge conjugation and R, but even under P. The second is even in C, R and P. So they do not mix. The Debye mass corresponds to the R=C odd channel, if we want to have it coincide with our definition in terms of self-energy of the C-odd A_0 potential in the previous section.

In principle one can take as definition of the Debye mass any operator with $R = -1$, no matter what C and P are. $R = -1$ garantuees that there is an odd number of operators A_0. The state with the lowest mass at a given temperature determines then the value of the Deye mass.

There is no difference between the two channels if $< PP >=< P^\dagger P^\dagger > = 0$. And in the phase where Z(N) is unbroken it is not hard to see that both are exponentially small with respect to the correlation $< P^\dagger P >$. The exponent is controlled by the string tension σ. In that case the lowest energy state with energy $E_0(T)$ has $\vec{e} = (e_x, e_y, e_\tau) = 0$. All states with $e_\tau \neq o$ are exponentially suppressed by factors $\exp -\frac{\sigma}{T} L_z$, and the reader can convince himself by going through the exercise below that both correlators give the same area law in the hadron phase. This is intuitively expected from two like charges being unscreened: in a periodic volume their expectation value should be zero.

- Show $< PP >= Tr_{phys} e^{-(H-E_0)\hat{L}_z} P e^{-(H-E_0)r} P e^{-(H-E_0)\hat{L}_z}$, up to exponentially small terms, if $Z(N)$ is unbroken.
- From the above, show that $< PP >$ is a superposition of amplitudes involving the flux states $|e>$ by using the projection operator eq.(4.5).
- Find the suppression factors in front of $< e_\tau + 2|Pe^{-(H-E_0)r}P|e_\tau >$ for all e and their absence in front of $< 0|P^\dagger e^{-(H-E_0)r} P|0 >$.

In the hot phase there is no reason the two channels are the same. Two like charges with screening can coexist in a periodic volume.

---

sign to A_0. Time reversal has no effect on the Wilson line, because it inverts the time ordering and at the same time transposes A_0 as a colour matrix. So the Wilson line will get transposed by T, but the trace does not not feel this. So R has the same effect on the Wilson line as C alone.

How does the Debye mass depend on the temperature? Below T_c it is zero. At very high temperature it is perturbatively calculable as we saw above. And what happens in between will be the subject of the next sections.

5.3 Screening of heavy magnetic charges

Not only the force law between heavy electric charges like the heavy quark, but also the force between heavy magnetic charges tells us about the medium. The original idea of 't Hooft and Mandelstam [11] was that of a dual superconductor, with the electric Cooper pairs replaced by some form of magnetic condensate. Especially the lattice community has been fascinated through the last 25 years with this idea because it defies perturbative access.

If only matter in adjoint representations is coupled to the gauge fields, 't Hooft [8] taught us how to find local operators that create monopoles with a flux $\exp i\frac{2\pi}{N}$ or multiples thereof.

In section (4.1.2) we defined a vortex along a periodic space direction. It had no end points.

The monopole anti-monopole pair at points $(0, r)$ is simply defined as by a vortex running from 0 to r on the positive z-axis. The vortex is given by a gauge transformation $V_k(r)$ which is discontinuous modulo a center group element $\exp ik\frac{2\pi}{N}$ when going around the vortex. So we can write the free energy due to the monopole pair like the free energy of a magnetic vortex without end points like in eq.(4.13) in the form of a Gibbs trace:

$$\exp -F_M(r)/T = Tr_{phys} V_k^\dagger(r) \exp -H/T V_k(r). \tag{5.11}$$

The vortex is like the Dirac string in QED. It is unobservable by scattering with particles in the adjoint representation, as long as it has center group strength.

The Gibbs trace can be worked into a path integral along the now familiar tricks from section (4.1.2), and on the lattice it takes the form [45]:

$$\exp -F_M(r)/T = \left(\int DA \exp -S_{(k)}(A)\right) / \int DA \exp -S(A). \tag{5.12}$$

The action $S_{(k)}$ is the usual action, except for those plaquettes pierced by the Dirac string. Those plaquettes are multiplied by a factor $\exp ik\frac{2\pi}{N}$, as in fig.(7).

- Show that any deformation of the string can be obtained by a change of integration (link) variable through a centergroup element.

The reader will recognize from section (4) the magnetic vortex but now with endpoints, where the monopole pair resides. Varying the endpoints permits one to find the potential for all separations.

Screening is expected in both confined and deconfined phases:

$$F_M(r) = F_{M0} - c_M \frac{\exp -m_M r}{r}. \tag{5.13}$$

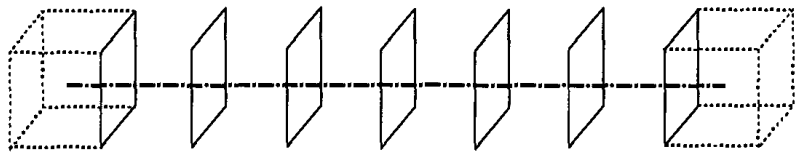


Figure 7: Monopole antimonopole pair induced by twisting the plaquettes pierced by the Dirac string.

All parameters are function of T. In the cold phase the screening is a consequence of the electric flux confinement. This is natural because the ground state contains a condensate of "magnetic Cooper pairs", according to the dual superconductor analogy. It is a screening mechanism whose details are not understood[8]. We dropped for notational reason the dependence on the strength k of the monopole. It is important to note that this strength comes in periodic modulo N! So the screening length is a periodic function of k and, because of charge conjugation, of $N-k$. It would be interesting to see what this dependence is.

In the hot phase there are indications from spatial Wilson loop simulations that there is additional thermal screening from magnetic quasi-particles, as discussed in section (8).

Analogous to the Wilson line correlator we consider the Hamiltonian H in the fictious system of (x, y, τ) variables. We search the operator V_k acting on the Hilbert space of physical states of this Hamiltonian, that reproduces the path integral eq.(5.12). So V_k should create a vortex in the (x,y) plane at every time slice τ and the Hamiltonian H should propagate every one of these vortices in the z-direction over a distance r. So V_k is the 't Hooft vortex operator discussed in section (4):

$$V_k = \exp i\vec{D}(A)v_k(x,y)\vec{E} \tag{5.14}$$

with $v_k(x,y) = \arctan(\frac{y}{x})\frac{1}{N}Y_k$.

Both under parity and charge conjugation the vortex V_k transforms into $V_k^\dagger$.

Its spin J equals 0, despite the appearance of the rotated singularity line in eq. (5.14). On physical states the location of the singularity does not matter, as we discussed before.

Hence the operator ImV_k excites $J=0$ states with $P=C=-1$. The magnetic screening mass should correspond, like the electric screening mass, to an odd charge conjugation state. The magnetic screening mass distinguishes itself from the electric screening mass by the opposite parity. We take here a slightly more restrictive definition of screening masses than Arnold and Yaffe in ref.[41]. Ours is inspired by the quantum numbers of the gauge invariant

[8]Nevertheless a quantitative understanding of the energy of a magnetic flux exists [8] as mentioned in subsection (4.1).

fluxes of the previous section. Theirs is inspired by the prominence of A_0 in the Debye mass, and hence for them it suffices to specify the R-parity. This is negative for A_0 and admits both 0_-^{-+} and 0_-^{+-} states. For us only the latter counts. Numerically the difference is not large, see section (7).

For $SU(2)$ gauge theory $ImV_1 = 0$ and we have to take ReV_1 exciting $J^{PC} = 0^{++}$ states [9]. The problem of how to distinguish electric and magnetic screening masses in $SU(2)$ is deferred to section (7).

- Convince yourself of the quantum number assignments (parity in two dimensions flips only one direction). Prove that $R = TC$ parity is $+1$.

Perturbation theory is not reliable and we need lattice simulations [48] [47].

6 A quantitative method: perturbation theory and dimensional reduction

Lattice simulations are our only tool today for tackling the critical region of QCD in a quantitative fashion, as far as the problematic fermions with small (including realistic) masses are avoided. But the region above a few times the critical temperature can be accessed by the method of dimensional reduction [16] without that problem. As we will see, fermions come in through the parameters of the reduced theory.

The method of dimensional reduction permits one to do perturbation theory, not only at very high temperatures but down to $T \sim 2T_c$. To obtain all coefficients of the perturbation series, one has to do dimensionally reduced lattice simulations, i.e. simulations in three dimensions. This is due to the three dimensional magnetic sector of the theory being a confining theory.

In fact the idea is similar to that in Kaluza-Klein theories: at high temperature the periodic dimension is very small with respect to the typical mass scale of 3D Yang-Mills, and the Fourier modes in the periodic direction are proportional to $2\pi Tn$ with n integer:

$$p_0 \to 2\pi nT, \text{ n integer for bosons.} \tag{6.1}$$

For the fermions the Fourier modes are anti-periodic so the n are half-integer. All non-zero modes are called hard modes.

So one works in three dimensional space with a 3D action. The field variables are the constant modes $A_\mu(\vec{x})$. The parameters in that action take the effects of the temperature into account. Some of the constant modes, the electric ones,

[9] Remember in SU(2) gauge theory all of the gauge-invariant sector has positive charge conjugation. This is due to the pseudo reality of SU(2): any element u is equivalent through $\sigma_2 u \sigma_2 = u^*$ to its complex conjugate. And the Pauli matrix σ_2 is precisely charge conjugation: $A_\mu^{cc} = \sigma_2 A_\mu \sigma_2$. As the gauge invariant sector involves integrating over charge conjugation, so is charge conjugation invariant. In particular, the gauge invariant electric and magnetic fluxes are C-invariant because they are defined mod 2.

get a mass due to Debye screening. The magnetic modes stay massless, at least in perturbation theory. So if one is not interested in distances on the order of the Debye screening but in much longer distances, integration of the electric screening modes is mandatory. We are then left with a 3D theory with only magnetic modes. They interact with a dimensionful coupling g_M^2 and describe a theory which is accurate on mass scales equal to or smaller than g_M^2.

6.1 Integrating out the hard modes

To be precise we want to integrate out all degrees of freedom in the original QCD action that relate to momenta and Fourier modes of order T. So we need to fix a cut-off Λ_E somewhere in between the scale T and gT. In low order we can do without. This is because we are interested in amplitudes with external legs with $n = 0$ and $p \sim gT$. To one loop order all modes with $n \neq 0$ in the loop introduce a scale of order T in the loop integration over $\vec{p}$. The mode with $n = 0$ has momenta of order gT injected from the external legs, so the momentum integration will involve only gT [10].

The form of our effective action S_E is dictated by all symmetries, global and local of the original QCD action and which are respected by the integration process. That implies all the symmetries we knew already, except that the electric term in the static action will have no $\partial_0 \vec{A}$ term. So A_0 appears as an adjoint Higgs term in our 3D gauge theory. The electrostatic QCD action density reads:

$$\begin{aligned} \mathcal{L}_E &= Tr(\vec{D}(A)A_0)^2 + m_E^2 Tr A_0^2 + \lambda_E (Tr(A_0^2))^2 + \\ &+ \bar{\lambda}_E (Tr(A_0)^4 - \frac{1}{2}(Tr A_0^2)^2) + Tr F_{ij}^2 + \delta\mathcal{L}_E. \end{aligned} \tag{6.2}$$

Because of R- conjugation invariance ($A_0 \to -A_0$) the electrostatic action must be even in A_0.

So far for the form of the action. The parameters in the action are all expressed in even powers of the QCD coupling. That is because only hard modes are present in the integrals. Odd powers will appear as soon as we admit modes of order gT.

The parameters in the A_0 sector are needed up to two loop accuracy [13] and we give the result for $N_c = 3$ and N_f:

$$m_E^2 = g(\mu)^2(1 + \frac{N_f}{6})T^2(1 + \Delta g^2(\mu)) \text{ and } \lambda_E = \frac{3g(\mu)^4}{8\pi^2}T(1 - \frac{N_f}{9})(1 + \delta g^2(\mu)). \tag{6.3}$$

For $\bar{\lambda}_E$ see reference [18]. The coefficients Δ and δ depend logarithmically on the scale μ and for their explicit form see refs. [13] [18].

[10] To two loop order there can be internal propagators with $n = 0$ and momenta on the order of T. To compute to this accuracy one either introduces the cut-off we just discussed [32] or one exploits $E(\tilde{P})$ as generating functional [13] [18] for the electrostatic action.

The gauge coupling g_E starts to run and in the $\overline{MS}$ scheme one finds [15]:

$$g_E^2 = g^2(\mu)T\{1 + \frac{g^2(\mu)}{(4\pi)^2}(\frac{22N}{3}\log\frac{\mu}{\mu_T} - \frac{4N_f}{3}\log\frac{4\mu}{\mu_T})\}. \tag{6.4}$$

The parameter $\mu_T = 4\pi T\exp(-\gamma - \frac{1}{22}) = 6.742..T$ follows from the one loop renormalization of the F_{ij}^2 term through the effects of scale T [15]. Eulers constant γ equals 0.577214... If one subtracts at this scale the renormalization effects appear only to two loop order and the coupling is then function of $T/\Lambda_{\overline{MS}}$:

$$\frac{g_E^2 N}{T} = \frac{24\pi^2}{11\log(\frac{6.742..T}{\Lambda_{\overline{MS}}})}. \tag{6.5}$$

Quenched QCD lattice simulations give us the critical temperature in terms of the QCD scale [39]:

$$\frac{T_c}{\Lambda_{\overline{MS}}} = 1.15 \pm 0.05, N = 3 \text{ and } \frac{T_c}{\Lambda_{\overline{MS}}} = 1.23 \pm 0.11, N = 2 \tag{6.6}$$

From the three dimensionful quantities in this Lagrangian we can form two dimensionless quantities:

$$x = \frac{\lambda}{g_E^2} \text{ and } y = \frac{m_E^2}{g_E^4} \tag{6.7}$$

We have made our promise true that the scale $\Lambda_{\overline{MS}}$ only goes in through the running of the coupling.

The dimensionless couplings x and y contain both $g^2(\mu)$. Eliminating the latter gives a very simple relationship between the former:

$$xy|_{4D} = \frac{2}{9\pi^2}(1 + \frac{9}{8}x + O(x^2)) \text{ for } N = 2 \tag{6.8}$$

and

$$xy|_{4D} = \frac{3}{8\pi^2}(1 + \frac{3}{2}x + O(x^2)) \text{ for } N = 3. \tag{6.9}$$

This is the trajectory in the x-y plane of the 4D physics, to order $O(x^2)$. We put $N_f = 0$. Remarkable is that it does not depend on the subtraction scale μ! The subtraction scale survives of course in the variable x but not in the relation between x and y. If this trend continues in higher orders, the series in x is probably well convergent.

Of course the physics of this effective action is specific to the quark-gluon plasma. First the coupling should be sufficiently small, and the presence of the mass term indicates that the electric flux is screened. And indeed m_E is to lowest order identical to the Debye mass m_D since both equal the one loop static self energy at zero momentum. The difference comes in the corrections. Whereas the corrections to m_E are $O(g^2)$ due to the absence of soft modes,

those to m_D are $O(g)$ due to presence of soft modes. The Debye mass m_D is a physical quantity. m_E is a parameter in the electrostatic action.

The mixed sector is known to one loop [14] up to six external legs. We lumped it into the term $\delta\mathcal{L}_E$. The reason for doing so is a question of accuracy. Already the superrenormalizable terms retained in eq.(6.2) do insure that the error we make in calculating some observable O with our electrostatic action is $O(g^4)$. This error to be numerically small constitutes one of several constraints on the value of the coupling. It warrants the calculation of the mass and four point coupling to two loop order above.

Let's see how this accuracy comes about. The argument is dimensional and based on the invariances of the reduced theory. These are the discrete spatial symmetries, 3D rotational and gauge invariance. Including two extra spatial potentials on any of the terms in $\mathcal{L}_E$ you get six independent terms [14] (from $F^3, (DF)^2, A_0^2F^2, A_0FDF$). A typical term reads:

$$\delta\mathcal{L}_E \sim \frac{g^2}{T^2}(DF)^2. \tag{6.10}$$

The square of the coupling appears because of the interaction of the stationary modes with the heavy modes. The scale T is there for dimensional reasons. The question is what this vertex is going to contribute. Irrespective of the observable in which it appears, we can say that the covariant derivative D concerns momenta in the effective theory of $O(gT)$. That gives an g^4 factor in front of the F^2 factor already present in the original Lagrangian $\mathcal{L}_E$, and provides the order of the relative error.

The estimate is generic. It can be higher order for specific observables. It motivates the two loop accuracy for m_E and λ_E above.

Another constraint is the following. The cut-off for our theory is $2\pi T$. The Debye mass ($\sim gT$), a typical scale of our electrostatic theory, should then be smaller than this cut-off. This means $g \leq 2\pi$, or $\alpha_s \leq \pi$.

In terms of temperature scales: if $\frac{T}{\Lambda_{\overline{MS}}} \sim 2$ then our coupling through eq.(6.5) with $N = 3$ equals $g \sim 1.7$, consistent with the cut-off limit.

6.2 Integrating out the electric screening scales

On the other hand there is the scale $g_E^2 = g^2T$. If this scale is much smaller than the Debye mass, we can integrate out the scale gT by integrating out the Higgs field A_0 from S_E. This necessitates the introduction of yet another cut-off Λ_M separating the scales gT from g^2T. That will lead to a new action S_M with only the magnetic fields present. This magnetostatic action density reads:

$$\mathcal{L}_M = TrF_{ij}^2 + \delta\mathcal{L}_M \tag{6.11}$$

with a magnetostatic gauge coupling g_M^2.

This coupling is related to the electrostatic coupling g_E through the renormalization of the magnetic gluon field strength:

$$F_{ij}^2 \to (1 + g_E^2 Z)F_{ij}^2$$

To one loop order the A_0 field is the only field contributing. To compute the Z factor we have to compute a diagram like in fig.(10d) with the wavy external legs the background magnetic potential with momentum of $O(g^2T)$. There is also a tadpole-like diagram contributing.

Simple power counting gives a linear infrared divergence for the transverse result Z. Since the infrared in S_E is cut off by m_E, we expect parametrically $Z \sim g_E^2/m_E \sim g$. So odd powers of g are to be expected. For g_M we get for three colours:

$$g_M^2 = g_3^2(1 - \frac{g_E^2}{16\pi m_E}) \tag{6.12}$$

for all reasonable couplings $g \leq 1$ a small effect.

Using the magnetostatic action at scales g^2T or smaller will induce an error $O(g^3)$ with respect to the results one would have got with the electrostatic action. Like in the previous subsection a generic estimate tells us for a typical term from the correction term in eq.(6.11):

$$\delta\mathcal{L}_M \sim \frac{g_E^2}{{m_E}^3}(DF)^2 \tag{6.13}$$

The coupling g_E describes the interaction between electric and magnetic modes, and m_E the scale of the integrated degree of freedom A_0. Now $D \sim g^2T$ and the relative error is $O(g^3)$.

This magnetic action has no dimensionless couplings like the electrostatic one. At scales g_M^2 it's obvious we cannot form a small dimensionless number with the coupling g_M^2. So the coupling in this theory is strong. A formal perturbation expansion of say the free energy gives from four loop order $(O(g_M^6))$ on powerlike infrared divergencies as naive power counting shows. Regulate with a mass m. Now, any free energy diagram with L loops has a power $(g_M^2)^{(L-1)}$ in front of the integral. So the integral must give a result $m^{(4-L)}$ to get the correct dimension for the the free energy. For $L = 4$ one expects a logarithm in the ratio cut-off over mass, calculated recently [20]. For $L \geq 4$ we have linear or higher divergencies. For a superrenormalizable theory all logs containing the cut-off are contained in $L = 4$.

When one regulates these divergencies with a mass of $O(g_M^2)$ higher loop diagrams are all of order g_M^6 modulo logarithms. This is Linde's argument [19].

What this means is that the coefficient of the sixth order free energy is not perturbatively calculable. We need non-perturbative input like the lattice.

If one needs higher order effects then the term $\delta\mathcal{L}_M$ has to be expanded in the magnetostatic partitionfunction Z_M:

$$Z_M = \int D\vec{A} \exp\left(- S_M(A) - \int d\vec{x}\delta\mathcal{L}_M\right) \tag{6.14}$$

$$= \int D\vec{A} \exp -S_M(A)(1 - \int d\vec{x}\delta\mathcal{L}_M + \ldots\ldots) \tag{6.15}$$

This gives an expansion for $-\frac{1}{V}\log Z_M = ag^6T^3(1 + bg + ...)$ where a, b and higher order coefficients have to be computed on the lattice or any other non-perturbative method.

7 Dimensional reduction at work

We will treat a few examples of relevant observables in order of mounting complication.

We will start with the spatial Wilson loop. Then the magnetic screening length m_M. Then its electric analogue, the Debye mass and finally the spatial 't Hooft loop and the pressure. We want to calculate the first terms in the series up to and including the term where the magnetostatic action enters for the first time. We just saw that this coefficient has to be computed from 3D lattice simulations. It turns out to dominate for any reasonable temperature, say from a few times T_c till $10^5 T_c$. Then one compares to 4D lattice data to determine the remnant of the series. This remnant turns out to be small, typically on the order of 30% up to $T \sim 2T_c$ at least for Wilson loop and Debye mass. For pressure, magnetic screening length and 't Hooft loop, this program is being pursued.

7.1 Spatial Wilson loop and magnetic screening mass: a window on the magnetic sector

The spatial Wilson loop is given in terms of a spatial loop L, and a representation r of $SU(N)$. The vector potential in this representation, $\vec{A}_r$, appears in the loop as:

$$W_r(L) = Tr\mathcal{P} \exp i \int_C g\vec{A}_r.d\vec{l}. \tag{7.1}$$

As for the Wilson line, eq.(3.6), the exponential is path ordered and hence invariant against regular gauge transformations . This spatial loop should measure the magnetic flux of any fixed gauge field configuration, as suggested by its abelian analogue and Stokes theorem. There is a useful version of Stokes theorem for the non-Abelian case [42] [43].

The thermal average of the spatial Wilson loop shows area behaviour with a surface tension $\sigma_r(T)$.

$$\langle W_r(L)\rangle_T = \exp -\sigma_r(T) A(L) + \ldots\ldots \tag{7.2}$$

The dots indicate perimeter terms. As far as the tension is concerned, it is very plausible [11] that it *only* depends on the number of quark minus the number of anti-quark representations constituting the representation r. This number is called the N-allity k of the loop. Also the tension is *periodic* in k modulo N. For the tension of the 't Hooft loop these properties are verified easily from its definition.

A useful corollary: a loop with N-allity k (a k-loop) and a $-k$-loop have by charge conjugation the same tension. So because of the periodicity also the $N-k$ loop has the same tension.

For $N = 2, 3$ only one tension results, because of charge conjuigation and periodicity.

[11] See the notes of Prof. Teper in this volume.

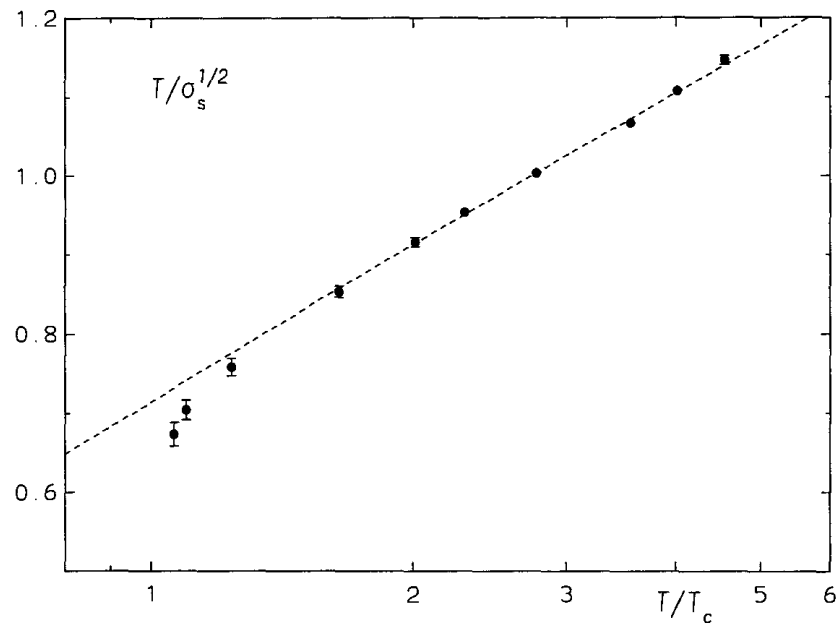


Figure 8: The temperature over the square root of the spatial string tension versus T/T_c for $SU(3)$. The dashed line shows a fit according to Eq. 6.5. From ref.[23].

We now establish that the loop is a perfect window on the magnetic sector. Its thermal average in path integral language is:

$$\langle W_k(L)\rangle = \int DA_0 D\vec{A} W_k(L) \exp -S(A) / \int DA_0 D\vec{A} \exp -S(A). \tag{7.3}$$

Integrate out all hard modes. They will not contribute to the tension σ_k of the loop, because the tension is due to correlations of the potential $\vec{A}$ over macroscopic distances. That will leave us with S replaced by S_E on the r.h.s. of the average. Since the spatial loop contains only spatial potentials $\vec{A}$ we can integrate over A_0 to obtain S_M from S_E. We arrive for the tension at the average:

$$\exp -\sigma_k(T) A(L) = \int D\vec{A} W_k(L) \exp -S_M(A) / \int D\vec{A}) \exp -S_M(A). \tag{7.4}$$

The hard and electrostatic free energies f_h and f_E drop out in the ratio.

The only dimensionful scale in the magnetostatic action is g_M^2. So the tension, having dimension $(\text{mass})^2$, can be written as:

$$\sigma_k(T) = c_k g_M^4 (1 + O(g^3)). \tag{7.5}$$

So the dominant contribution to the tension is entirely from the magnetostatic sector. In figure (8) you see a fit of the tension data to this parametric

expression for SU(3). The authors took for the magnetic coupling $g_M^2 = g_E^2$ so neglected renormalization effects of the scale gT, which are a few percent at $T = 2T_c$, see eq.(6.12). Notably the value of the tension at the critical temperature is within errors equal to the tension at zero temperature.

The conclusion is quite clear: down to temperatures a few times T_c, the loop behaviour is determined by leading order magnetic sector effects! These effects are embodied in the dimensionless number $c_{k=1}$. The rest of the T-dependence is through the hard-mode-running of the coupling, eq.(6.5). The number $c_{k=1}$ is within errors equal to the purely 3D simulation of the loop.

The spatial Wilson loop measures in a sense to be specified later the magnetic flux in the system. The tension is flat from $T = 0$ to $T = T_c$, according to the data. In all of the confined phase the magnetic activity does not change.

Above T_c it starts to grow like $g_M^4 T^2$. Apparently beyond the transition the activity goes up, and comes, as the data tell us, entirely from the magnetostatic sector.

This window on the magnetic sector spurs an obvious question [17]: what is the dependence of the coefficient on k and N? The lattice data by Teper and Lucini [28] are consistent within a percent to the simplest possible picture for the 3D magnetic sector: that of a gas of almost free quasi-particles, static transverse gluons. We'll come back to this in the last section.

7.1.1 The magnetic screening mass

The magnetic screening mass m_M was introduced in section(5) as the magnetic analogue of the Debye mass. It gets its leading order contribution from the magnetostatic sector, like the spatial Wilson loop.

Four dimensional data for SU(2) have been taken [48] [47]. But the numerics is much more involved than that for the Wilson loop. Qualitatively the data for the screening mass are compatible with the behaviour of the spatial Wilson loop tension. In the cold phase its value is about twice the lowest glueball mass. Beyond T_c it starts to rise, as you can see from the lower part in fig.(9):

$$m_M = r g_M^2 = r g^2(T) T \tag{7.6}$$

as one would expect from a mass in the 3d theory.

Because the mass is high, the signal to noise ratio becomes small and numerical extraction becomes tedious.

Data are being improved [48] for the region around $2T_c$, where we want to confront them with the 3d data.

Quantum numbers of the magnetic screening mass in SU(2) are given by $J^P = 0^+$. Precise 3D data for SU(2) are available from ref. [21] and [46]. The 4D data are certainly not compatible with the lowest 0^+ state. There are two recurrences, the highest of which (ref. [46], table 28) is compatible with the 4D data at $T = 2T_c$ in fig.(9), up to 30%. This certainly deserves clarification.

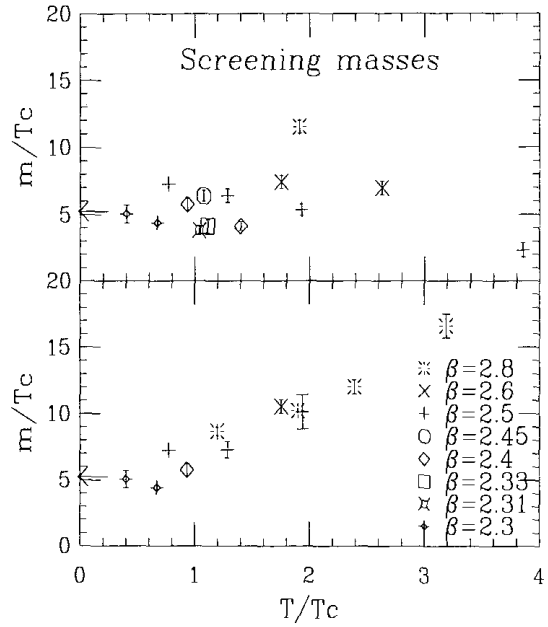


Figure 9: Screening mass as a function of temperature, both in units of T_c, as extracted from spatial (top) or temporal (bottom) 't Hooft loops. Below T_c both coincide. The arrow gives the mass of the scalar glueball at $T = 0$. From ref. [47].

7.2 The Debye mass

The Debye mass to next to leading order is given by:

$$m_D = (\frac{N}{3} + \frac{N_f}{6})^{\frac{1}{2}} gT + \frac{g^2 N}{4\pi} T(\log(\frac{\frac{N}{3} + \frac{N_f}{6})^{\frac{1}{2}}}{g} + 7.0) + O(g^3) \tag{7.7}$$

The first term is the result from the one loop self-energy discussed in section (5). Remember only hard modes contributed to one loop order.

The second term is only $O(g)$ smaller. The log term in the coefficient is due to Rebhan [30] and comes from scales between electric and magnetic ones. The large number comes from all scales equal or smaller than the magnetic one, g_M^2. It is non-perturbative and calculated by numerical simulation on the lattice [12] [21] [22] for N=2 and 3. How is explained below.

A natural choice in 3d is the reduced version of the Wilson line: $P = Tr \exp iA_0$. Its imaginary part is the negative C parity channel.

- Exercise: Use the electrostatic action $\mathcal{L}_E$ in the correlator.
  Show $< ImP(0) ImP(z) >_E$ has as dominant decay mode $\exp -3m_D z$

The mass can be extracted from this correlator. In terms of the electrostatic parameters eq.(6.7) it has a simple form:

$$m_D = m_E(1 + d\sqrt{\frac{1}{y}} + O(x^2)) \tag{7.8}$$

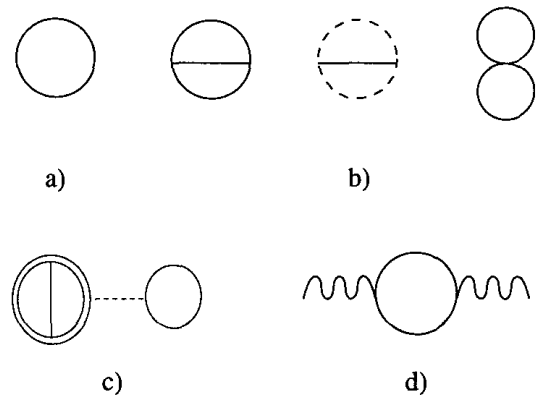


Figure 10: (a) is the one loop contribution, and stands for gluon and ghost loop.(b) is the two loop contribution.(c) is the renormalization of the thermal Wilson line(double circle) inserted (dotted line) into the one loop.(d) is the kinetic term with the loop being a gluon or ghost loop

We have to evaluate the correlator along the physics line eq.(6.9). To see that we can expect this form for the correction take the self energy correction to one of the three A_0 propagators contributing to $< ImP(0)ImP(z) >_E$. Let the momentum of this propagator be $\vec{p}$. Schematically, the self energy insertion reads:

$$g_E^2 \int \frac{d\vec{l}}{(2\pi)^3} \frac{(2l+p)^2}{(l^2+m_E^2)((l+p)^2} \tag{7.9}$$

We are interested in the contribution of this integral to the Debye mass in the exponent $\exp -m_D z$. So we have to evaluate the integral on-mass shell: $\vec{p}^2 + m_D^2 = 0$. That contribution is for dimensional reasons $\sim \frac{1}{m_E}$. That, together with the coupling constant factor g_E^2 gives indeed $\frac{g_E^2}{m_E} = \sqrt{\frac{1}{y}}$.

In eq.(7.7) the numerical value of the coefficient d for N=3 is given [12]. Its presence renders the next to leading term dominant till $T \sim 10^6 T_c$!

But what is even more so is that the remnant of the series in eq.(7.7) converges well up to $T = 2T_c$. This follows from comparing to the full 4D lattice data [37].

And so the pattern here is the same as for the spatial Wilson loop: once the magnetic sector has contributed to the series the rest is well convergent.

7.3 Spatial 't Hooft loop

In this subsection we finally make good our promise to compute the low orders in the tension of the spatial 't Hooft loop at high temperature. We met this

[12]Strictly speaking, this value comes from the 0_-^{-+} state instead of the 0_-^{+-} state in the ImP correlator. The difference in mass is small, see fig. 1 of ref.[22].

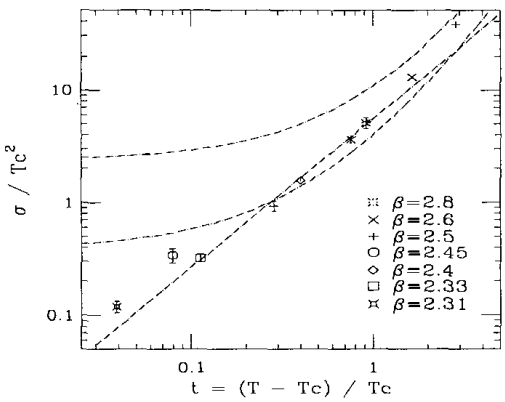


Figure 11: 't Hooft loop tension ρ_1 for SU(2), in units of T_c^2, as a function of the reduced temperature t. The straight line is a power law fit to $t < 1$. The fitted exponent is 1.32(6), to be compared with $2\nu \approx 1.26$ for the $3D$ Ising model. The curves show the perturbative result, to leading (upper) and next (lower) order, eq. (7.21) in the text. From ref.[47].

tension in section (4), where it came up as a ratio of twisted partition functions, eq.(4.15).

First we shall give a definition that makes clear its connection with the thermal Wilson line and with thermal Z(N) symmetry.

Consider a closed loop L in the (x,z) plane, and define the spatial 't Hooft loop $V_k(L)$ as a gauge transformation that has a discontinuity z_k on the minimal surface spanned by the loop.

This definition tells us that the loop is a closed magnetic flux loop with strength z_k, as we defined in section (4.1).

A simple realization of such a transformation is the solid angle $\omega_L(\vec{x})$ with which L is seen from a point $\vec{x}$. It jumps over 4π when crossing the surface. So the corresponding operator in Hilbert space is formed by taking Gauss's operator:

$$\tilde{V}_k(L) = \exp i \int d\vec{x} \frac{1}{g} Tr \vec{E}(\vec{x}).\vec{D}\omega_L(\vec{x}) \frac{Y_k}{2N}. \tag{7.10}$$

Remember from section (4) that $Y_k = \text{diag}\,(k, k, ..., k, k-N, ...k-N)$ with N-k entries k and k entries k-N, so that it generates the center group element:

$$\exp\,(i\frac{2\pi}{N}Y_k) = \exp ik\frac{2\pi}{N} = z_k. \tag{7.11}$$

What does a thermal Wilson line feel when it passes through the minimal area of the loop?

The answer is simple: since the Wilson line represents a heavy test quark it will pick up the Z(N) phase z_k.

So the operator acts like a twist. To recover the twisted partition functions $\Omega_{\vec{k}}$ of section (4) (with $\vec{k} = (0, k, 0)$ for the case at hand), just extend $S(L)$ over

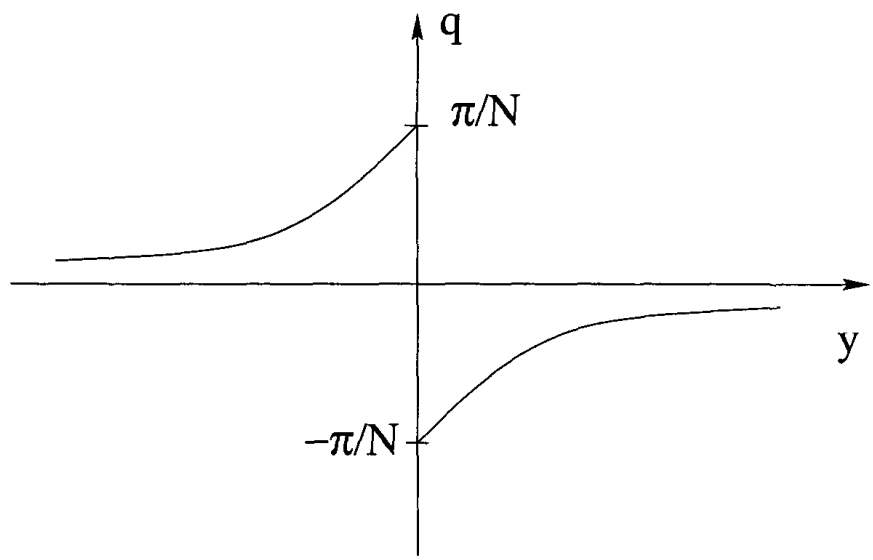


Figure 12: The Wilson line profile, that dominates the steepest descent calculation of the effective potential. The parameter q is defined in the text below eq.(7.13) for N=2.

the full x, z cross section [13]. As promised in that section we now compute the area law for the 't Hooft loop, when $T \geq T_c$:

$$< \tilde{V}_k(L) >= \exp -\rho_k(T) S(L) \tag{7.12}$$

and it is the tension $\rho_k(T)$ we are after.

7.3.1 The strategy for computing $\rho_k(T)$

Imagine the loop immersed in the plasma. Far away from the loop the value of the Wilson line is some fixed Z(N) value, as we learned in section (3.3). We will take the value $P(A_0) = 1$, but any other would have done equally well.

The geometry of the problem is such that all profiles orthogonal to the plane of the loop will be identical. Only near the border this is no longer true. We are only interested in the surface effects, not in the border effects. As we said before the Wilson line will jump at the surface of the loop. Such a typical profile is shown in the figure above.

The jump should be such, that the derivatives on both sides are equal. Once we have fixed these boundary conditions, all we have to do is to determine what profile minimizes the effective potential. This is what we are going to do in the next section.

[13]For the vortex correlation in the 2+1 dimensional theory in section (4) precisely the same holds true. Mutatis mutandis the reasoning of this section applies to that correlation as well.

7.3.2 Effective potential for constant profile

In this section the effective potential is computed. But, to keep the discussion as simple as possible, we will first consider the potential for constant profile. To be specific take SU(2) gauge theory. The key to computing the surface tension is the distribution function $E(\bar{P})$ of the thermal Wilson line that we introduced in section(3.3), eq.(3.11). We are interested in the logarithm of this distribution function:

$$E(\tilde{P}) = \int DA\delta(Tr\exp iC - Tr\overline{\mathcal{P}(A_0)})\exp -S(A) \equiv \exp -L^3(V(q)+f) \tag{7.13}$$

We have introduced the diagonal traceless matrix $C = \pi\mathrm{diag}(q,-q)$ to parametrize $\tilde{P}$. The free energy density f normalizes the distribution. L^3 is the spatial volume. The distribution function has two peaks, one at P=1, one at P=-1. So the effective action $V(q)$ has minima at q = integer. Even (odd) values correspond to $P = 1(-1)$.

The constraint in eq.(7.13) tells us that the potential A_0 fluctuates around the background matrix CT. So this matrix will be present in the Feynman rules as a background field. This is familiar. What is less familiar is that the background gets renormalized through the renormalization of the thermal Wilson line. This is an effect coming in from two loop order on, and is shown in fig. (10c). Only the sum of the two loop diagrams b) and c) is gauge independent.

The covariant derivative is $D_0(A_0 = TC)$. Such a derivative is diagonal in the Cartan basis for the fields $A = \frac{CT}{g} + Q^+\tau^- + Q^-\tau^+ + Q^3\tau^3$. The Pauli matrices $\vec{\tau}$ relate to $\sqrt{2}\tau^\pm$ as $\tau_1 \pm i\tau_2$.

So the Feynman rules consist of the Euclidean ones, except that p_0 is replaced by $p_0 \pm 2\pi Tq$ when it is the Matsubara frequency of $Q^\pm$. Hence it appears in propagators of $Q^\pm$ and in vertices when it pertains to an outgoing $Q^\pm$ line.

The one loop computation (see fig.(10a) is the calculation of the determinant of the quadratic part in the quantum fields Q of the action. In Feynman gauge you get:

$$V(q) + f = \frac{1}{L^3}\log\det(-D(C)^2). \tag{7.14}$$

The choice of gauge does not matter as is shown explicitely in Pisarski's lectures. $D(C)^2$ is the covariant Laplacian.

There is a well-known identity that relates $\log\det M = Tr\log M$ for any matrix M. Then, in the momentum basis you obtain in terms of the Fourier transformed Laplacian for $V(q)+f$:

$$T\sum_n\int\frac{d\vec{l}}{(2\pi)^3}(\log((2\pi T(n-q))^2+\vec{l}^2)+\log((2\pi T(n+q))^2+\vec{l}^2)+\log((2\pi Tn)^2+\vec{l}^2) \tag{7.15}$$

with the contributions from Q^+, Q^-, Q^3 integrations written out explicitely. If we sum over all values of n, clearly the result will be periodic in q mod 1 and

even in q. This is of course a consequence of the thermal Z(N) symmetry. In fact one gets:

$$V(q) = \frac{4}{3}\pi^2 T^4 q^2 (1 - |q|)^2 \tag{7.16}$$

As long as $qT = O(T)$ the integration over $\vec{l}$ is hard. For $q = O(g)$ or $q = 1 - O(g)$ we have for $n = 0$ or $n = -1$ soft momentum contributions.

7.3.3 Varying profile and gradient expansion

Till now we computed the potential as if the profile q were constant. But what we are really after is the profile of q as a function of y in between the values $P = 1$ and -1 at $y = \pm\infty$. That means we have to include the kinetic term $K = (\frac{T}{g^2}\partial_y C)^2$ to our potential V to get the tension:

$$U(q) = \int_{-\infty}^{\infty} dy(K + V) \tag{7.17}$$

So we tunnel through the potential mountain $V(q)$ from $P = 1$ to $P = -1$.

We swept a little problem under the rug. We forgot the contribution from the one loop potential due to the gradient in $q(y)$! This however is of higher order as we will see shortly.

To find the tension you have to minimize U with the boundary condition that the wall is between P=1 and P=-1 regions. This is done by the method of completing the square:

$$U(q) = \int dy((K^{1/2} - V^{1/2})^2 + 2(KV)^{1/2} \tag{7.18}$$

The integration of the second term over y can be replaced by an integration over q, using the chain rule for $\int_{-\infty}^{\infty} dy K^{1/2} \sim \int_{-\infty}^{\infty} dy \partial_y q = \int_0^1 dq$:

$$\int dy 2(KV)^{1/2} = \frac{2}{g}(8\pi^2 T^2)^{1/2} \int_0^1 dq V^{1/2} \tag{7.19}$$

which gives a number *independent* of the profile q!

Hence $U(q)$ in eq.(7.18) is minimized for that profile q that obeys the equations of motion

$$K^{1/2} - V^{1/2} = 0. \tag{7.20}$$

The tension of the loop is then given by eq.(7.19) and using the result (7.16) for the potential one gets:

$$\rho_1(T) = \frac{4\pi^2}{3\sqrt{6g^2}} T^2 (1 - c_2\, \alpha_s(T) + ...) \tag{7.21}$$

Two loop corrections have been computed [38] from the graphs in fig.(10b, c and d), with the result $c_2 = 2.0682....$

What is the typical width of the Wilson line profile? Just look at the equations of motion eq.(7.20). Write them out with eq.(7.16):

$$\partial_y q = m_D q(1 - |q|) \tag{7.22}$$

with $m_D^2 = \frac{2}{3}g^2T^2$, the lowest order Debye mass from $N = 2$. So it is the Debye mass that governs the width of the wall. This also answers the question about gradient terms. They are there, but are $O(g^2)$ compared to $V(q)$.

Only hard momenta do contribute to this order, as you can see by computing the contribution to ρ_1 from the wings of the profile near $P = 0, 1$.

The lattice simulated tension [47] in SU(2) is shown in fig.(11). For $T \sim 2T_c$ lattice data and perturbative prediction do agree reasonably well.

The cubic corrections are now known [17] and add a positive contribution. They are soft modes on the scale of the Debye mass. As we will see the same pattern shows in the pressure, calculated with the same graphs as in fig.(10), but with the background q set to zero. The three loop contribution is in progress.

Magnetic modes will contribute through the next to leading order of the Debye mass (see eq. (7.7)) and have not yet been computed.

For general N and strength k of the loop one finds to one and two loop order a remarkably simple scaling law in k [17]:

$$\rho_k(T) = \frac{k(N-k)}{(N-1)}\rho_1(T) \tag{7.23}$$

with $\rho_1(T) = \frac{4\pi^2}{3\sqrt{3}g^2N}(N-1)(1-1.0341..\alpha_s\, N+O(g^3))$, periodic in the strength k mod N.

In one loop order this is a simple consequence of additivity of the potentials for the various colour modes and of the tunneling path being along the Y_k direction. Let us look into that in more detail.

One tunnels from 1 to z_k by the path qY_k, from q=0 to q=1, so the diagonal background matrix is $C = \frac{2\pi q}{N}Y_k$. The operator whose determinant we have to compute is $-D(C)^2$ [14].

Let us adopt the Cartan basis for the fluctuating potentials Q in NxN matrix notation, like we did before for SU(2). This is a colour basis in which the effect of the profile is diagonalized in the Laplacian $-D^2(C)$.

Remember the profile C is diagonal by construction, with diagonal elements C_i, $i = 1, ..., N$. The profile appears only in the covariant derivative $D_0(C)$. There are diagonal fluctuations Q^d, that to one loop order do not contribute any C dependence:

$$D_0(C)Q^d = \partial_0 Q^d + iT[C, Q^d] = \partial Q^d. \tag{7.24}$$

Then there are off-diagonal fluctuations Q^{ij} $(1 \leq i \neq j \leq N)$ with Q^{ij} only non-zero in the (ij) entry. For those you have:

$$D_0(C)Q^{ij} = \partial_0 Q^{ij} + iT[C, Q^{ij}] = \partial_0 Q^{ij} + i(C_i - C_j)Q^{ij} \tag{7.25}$$

[14] We drop all indices exept colour indices.

and we have diagonalized the Laplacian. The background field C comes in through the diagonal elements of the *adjoint* representation of $C = \frac{q}{N} Y_k$. They equal 0, or $\pm q$ up to a factor 2π.

First we write the contribution to the C dependent part in the potential for a fixed combination (ij):

$$V_{ij}(C_i - C_j) = T \sum_n \int \frac{d\vec{l}}{(2\pi)^3} \log(T(2\pi n + C_i - C_j))^2 \tag{7.26}$$

The reader will recognize the form of the potential for SU(2), eq.(7.15).

By changing the sign of the summation over Matsubara modes, it follows that $V_{ij}(C_i - C_j)$ is even in $C_i - C_j$.

Remember that on the path $C = \frac{2\pi q}{N} Y_k$ the profile in its adjoint representation equals $|C_i - C_j| = 2\pi q$, or 0. So the C-dependent part of the full potential is obtained by multiplying V_{ij} with the number of off-diagonal modes (ij) with a non-zero eigenvalue $C_i - C_j$.

This number is determined from $Y_k = \mathrm{diag}(k, ..., k, k-N, ..k-N)$ by taking all differences of elements. There are k elements with value $k - N$, and $N - k$ elements with value k. Hence there are $2k(N-k)$ ways of picking a non-zero combination. We obtain for the full potential:

$$V(q) = 2k(N-k)(V_{ij}(2\pi q) - V_{ij}(0)) = k(N-k)\frac{4\pi^2}{3} T^4 q^2 (1 - |q|)^2. \tag{7.27}$$

The second equality follows from eq.(7.15) and eq.(7.16).

For the kinetic term one has the identity $Tr(\partial_y C)^2 = \frac{1}{2N} \sum_{i,j} \partial_y (C_i - C_j)^2$, valid because $TrC = 0$. So the same counting as above applies to the kinetic term:

$$K = \frac{T^2}{g} Tr(\partial_y C)^2 = k(N-k)\frac{4\pi^2 T^2}{g^2 N}(\partial_y q)^2. \tag{7.28}$$

Then eq.(7.23) follows from the minimization of $K + V$.

- The stability group of Y_k is $SU(N-k) \times SU(k) \times U(1)$. Determine the dimensionality of the coset space $SU(N)/SU(N-k) \times SU(k) \times U(1)$. Explain the masses appearing the diagonal and off-diagonal propagators $-D^{-2}(C)$.

Some comments on the 't Hooft loop tension, eq. (7.23):

- The result for the tension is $O(N)$ in the large N limit, in contrast to the tension of the Wilson loop. The latter is $O(1)$ as follows from the well-known index loop counting, valid to all orders in perturbation theory.
- The absence of a sacred cow is noteworthy in eq.(7.23). For k of order 1 the result has corrections $O(1/N)$! A priori we would have expected corrections of $O(1/N^2)$ in a theory with only gluons, like in the pressure. This anomalous correction is a simple consequence of the counting: not $N^2 - 1$ gluons, but only $2k(N-k)$ gluons contribute to the tension, whereas they all do contribute in equal amount to the pressure.

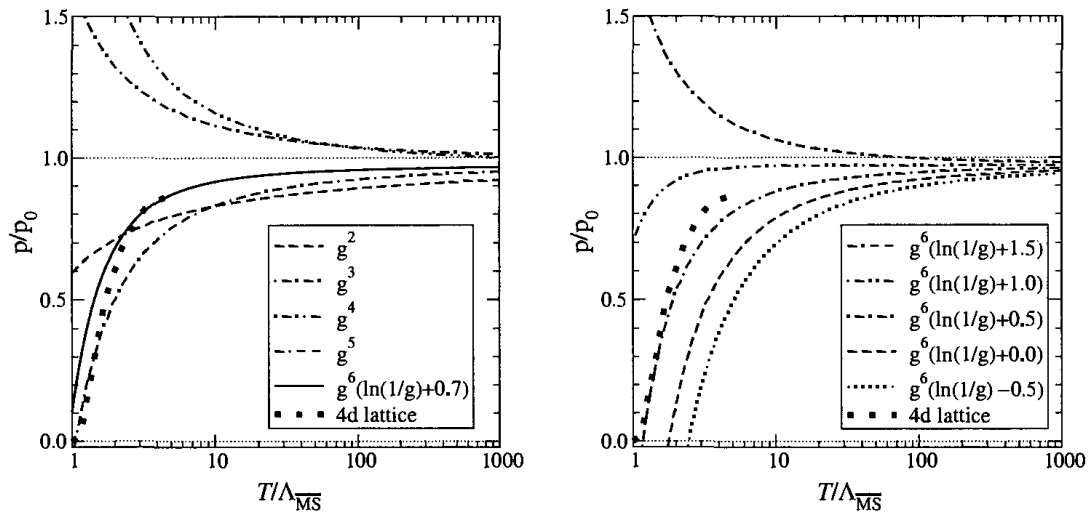


Figure 13: Left: perturbative results at various orders, including $\mathcal{O}(g^6)$ for an optimal constant. Right: the dependence of the $\mathcal{O}(g^6)$ result on the (not yet computed) constant, which contains both perturbative and non-perturbative contributions. The 4d lattice results are from [33]. From ref. [20].

Physically eq.(7.23) is understood as being due to the flux of the screened quasi-particles and will be discussed in section (8).

7.4 Pressure

In fig.(13) you see the pressure as measured by lattice simulation for three colours. It is compared to the analytically computed [20] terms in perturbation theory up and including $g^6 \log(\frac{1}{g})$. The series starts with the contribution from the gluons as free quasi-particles: the Stefan-Boltzmann gas with pressure $p_0 = \frac{8\pi^2}{45} T^4$ as computed from fig.(10a) . The interactions between the gluons give the two-loop contribution as in fig.(10b). The contribution of the Debye-screened gluons is the dominant one from $\mathcal{L}_E$ in eq.(6.2). It equals the logarithm of the determinant of the Higgs modes with mass m_E. Using dimensional regularization

$$-\frac{1}{2}(N^2-1)\int \frac{d\vec{l}}{(2\pi)^3} \log(\vec{k}^2 + m_E^2) = \frac{\Gamma(-\frac{3}{2})}{16\pi^{\frac{3}{2}}} m_E^3 \tag{7.29}$$

This being positive and large with respect to the two loop contribution (like in the case of the 't Hooft loop) spoils the convergence of the series except for academically high temperature. The two next orders g^4 and g^5 undo in part the effect of the cubic term. But it is clear that perturbation theory has little

predictive power in the region $2T_c$ to $100T_c$, as it varies there over more than 20%!

To put the calculation of the contributions of order higher than three in perspective and to see how the different scales come in, we recall once more the hierarchy of scales, cut-offs Λ and reduced actions needed to compute the pressure:

$$T >> \Lambda_E >> gT >> \Lambda_M >> g^2T$$

The pressure is normalized by p_0 and consists of three parts:

$$\frac{p}{p_0} = p_h + p_E + p_M$$

The hard modes are cut-off in the infrared by Λ_E and equal p_h. Schematically we get:

$$p_h = 1 + g^2 + g^4 \log \frac{T}{\Lambda_E} + g^4 + g^6 log \frac{T}{\Lambda_E} + g^6 + ..$$

All powers of the coupling are even. The short distance scales (larger than T) are absorbed in the running coupling, eq.(6.4). The cut-off Λ_E appears only in logarithms. The electric mode contributions are computed with $\mathcal{L}_E$ and give p_E:

$$p_E = g^3 + g^4 \log \frac{\Lambda_E}{m_E} + g^4 + g^5 + g^6 \log \frac{\Lambda_E}{m_E} + g^6 \log \frac{m_E}{\Lambda_M} + g^6 + ..$$

The dominant cubic term was computed in eq.(7.29). We can expect logarithms of the two ratios of the three scales m_E, λ_E and Λ_M in the electrostatic action.

Finally the magnetic contribution is computed with $\mathcal{L}_M$:

$$p_M = g^6 \log \frac{\Lambda_M}{g_M^2} + g^6 + ...$$

We only put in the obvious dependence on the parameters in the electrostatic and magnetostatic actions. There are three comments:

- All terms shown are perturbatively calculable, except the last one in p_M.
- All perturbatively calculable terms have been computed [40], except for the g^6 terms. In particular the log's are known by now [20].
- All dependence on the cut-offs cancels, as expected.

Clearly this is quite a calculational performance!

So at this point one expects that the miracle of the Debye mass may materialize: compute the non-perturbative term of $O(g^6)$ with $\mathcal{L}_M$ on the lattice, and the perturbative term of $O(g^6)$. Does the ensuing series stabilize down to $\sim 2T_c$?

To see whether this may work at all, values for the sum of the yet unknown non-perturbative and perturbative coefficients have been put in the right hand figure (13). Clearly there is a window where lattice data do connect smoothly to the series.

8 Flux of quasi-particles as seen by spatial Wilson and 't Hooft loops

The idea of quasi-particles is to leading order in the pressure embodied by the Stefan-Boltzmann term. It counts all the degrees of freedom of the gluons, colour and spin.

What we want to argue below is that specific degrees of freedom of the gluons, namely their flux, are clearly seen in the behaviour of the spatial 't Hooft loop, especially the scaling law in the strength of the loop. Then we will look for a similar effect in the Wilson loop, but now with quasi particles that carry magnetic flux.

8.1 Gluon flux and the 't Hooft loop scaling law

We found in perturbation theory, including two loop order:

$$\rho_k(T) = \frac{k(N-k)}{(N-1)}\rho_1(T). \tag{8.1}$$

Below we will show that the factor $k(N-k)$ is the number of gluons in the adjoint multiplet with a fixed value N for the charge characterizing the strength k of the loop. Essential is the physical meaning of the loop: it measures the colour electric flux in the plasma. The loop, eq.(7.10), can be rewritten as a sheet of electric dipoles:

$$V_k(C) = \exp \frac{i4\pi}{N} \int d\vec{S}.Tr\vec{E}Y_k \tag{8.2}$$

and measures the colour electric flux going through the loop.

Expression (8.2) is a dual Stokes law: remember the original definition of the 't Hooft loop is that of a loop of a Dirac magnetic flux. But there is no magic about (8.2)! Both definitions can be shown to have the same commutation relation with the Wilson loop

$$V_k(C)W(C')V_k(C)^\dagger = z_k^{n(C,C')}W(C') \tag{8.3}$$

$n(C,C')$ being the number of times C and C' loop each other.

Hence the product $\tilde{V}_k(C)V^\dagger(C)$ does commute with the Wilson loop.

So the product is a regular gauge transformation and acts as the unit operator in the physical Hilbert space. Both the representation eq.(7.10) and the flux representation eq.(8.2) are identical there.

- Show that $V_k(C)$ has the 't Hooft commutation relation eq.(8.3) with a spatial Wilson loop $W(C')$. (Hint: use the canonical commuation relations between field strength $\vec{E}$ and potential $\vec{A}$ in the Wilson loop.)

Consider a gluon close to the minimal area of the loop. As its flux is screened it needs to be within the screening length l_D to shine its flux through the loop [15]. The gluon is in the adjoint representation so its charge Y_k is either 0 or $\pm N$ as you can see from the differences of eigenvalues of Y_k. There are then obviously $2k(N-k)$ gluon species in the adjoint representation with this charge $\pm N$. Each one individually shines flux through the loop. This flux equals $\pm N/2$, the other half of the flux is lost on the loop. That means that such a gluon will give a contribution to the loop of

$$V_k(C)|_{one\ gluon} = \exp((i\frac{2\pi}{N}(\pm\frac{N}{2})) = -1. \tag{8.4}$$

Assume that the distribution $P(l)$ of gluons in the slab of thickness l_D containing the loop is independent of the species and say Poissonian:

$$P(l) = \frac{1}{l!}\bar{l}^{\,l}\exp -\bar{l}.$$

Here $\bar{l}$ is the average number of gluons of that species in the slab. Then the average of the loop over just one species is:

$$< V_k(C) > |_{one\ species} = \sum_l (-)^{\,l}\frac{1}{l!}\bar{l}^{\,l}\exp -\bar{l} = \exp -2\bar{l}. \tag{8.5}$$

If the contribution of all $2k(N-k)$ species is independent the result becomes:

$$< V_k(C) >= \exp -4k(N-k)\bar{l}. \tag{8.6}$$

Now $\bar{l} = 2l_D S(C)n(T)$. n(T) is the density of a specific gluon species and an area law results with the tension:

$$\rho_k(T) = 8k(N-k)l_D n(T) \tag{8.7}$$

This formula represents the physical raison d'être of the tension. It is due to the screening of the electric flux of the gluons, their density and their degeneracy with respect to the charge Y_k characterizing the strength of the loop. The Poisson distribution function is not essential to this result. Any thermodynamic distribution function, that is peaked around the average $\bar{l}$ with a width $\overline{l^2} - (\bar{l})^2$ proportional to $\bar{l}$, will give an area law. It is this proportionality constant that will appear in eq.(8.7). But it does not depend on the strength k and will drop out in ratios.

Hard gluons with momentum of order T, will have a photon like distribution function

$$P(l) \sim (\bar{l}/(1+\bar{l}))^l.$$

As the reader can easily check, it has a variance $\overline{l^2} - (\bar{l})^2 = (\bar{l})^2 + \bar{l}$, so fluctuations of order 1. Not surprisingly it does not give an area law as you can find out by plugging it in eq.(8.5). Small fluctuations are essential for the area law.

[15] This is a simplification. Also gluons farther out can contribute, giving a change in the overall factor of our final result.

8.2 Magnetic flux and k-scaling of the spatial Wilson loop

The question is now: what about similar ratios of the spatial Wilson loop?

We convinced ourselves in section (7) that the leading contribution to its tension came from the static magnetic sector. This sector is populated with static transverse gluons, i.e. gluons with a static magnetic field screened in some non-perturbative way with a screening mass m_M. The screening mass is inversely proportional to the coupling g_M^2 on dimensional grounds. More precisely, in our discussion of the magnetic screening in section (7) we noticed that the proportionality constant was large with respect to other scalar masses, because of its unnatural parity.

In the context of our model for the 't Hooft loop tension, it is interesting to see what the ratio of the tension of the Wilson loop to magnetic screening mass is according to the lattice [46]:

$$\frac{m_M}{\sqrt{\sigma}} = 8.15(15) \quad \text{SU(2)} \tag{8.8}$$

Suppose that the screening length $l_M = m_M^{-1}$ of our magnetic quasi-particles is much smaller than the mean distance between them:

$$n_M l_M^3 \ll 1. \tag{8.9}$$

Both are parametrically equal to g_M^6. There must be a dynamical reason for this ratio to be small. It supposes the magnetic screening of the quasi-particles is so efficient that they constitute a gas of approximately free lumps.

If the latter is true, then the reasoning in the previous section would, mutatis mutandis, give for the tension of the Wilon loop:

$$\sigma = c\, l_M n_M(T) \tag{8.10}$$

where c is some numerical constant depending on your preferred probability distribution. So this relation turns the numerical result for the screening length eq.(8.8) into a numerical result for the density of screened magnetic quasi-particles:

$$c\; n_M l_M^3 = 0.015.... \tag{8.11}$$

This means one quasi-particle in about seventyscreening volumes if c=1, an a posteriori justification for the model. For the Poison distribution this density is even smaller.

This is for SU(2).

Remember that for SU(3) and higher groups the magnetic screening was related to 0^{--} states. From Teper's work [46] the lowest 0^{--} mass in units of the string tension for SU(N) ($N \geq 3$)) is:

$$n_M l_M^3 = 0.028....(1 - (1.64...)/N^2 + O(N^{-4})) \tag{8.12}$$

Although both screening mass and square root of string tension are parametrically the same, their ratio is for dynamical reasons small! So we may say that the quasi-particles are to a good approximation free. How good will become more clear at the end of this section.

So having taken courage we now turn to a flux representation of the spatial Wilson loop [42] [43]. In fact, if we want to find out about the strength or N-allity k of the loop, we would expect in analogy with the 't Hooft loop, eq.(8.2), a magnetic dipole distribution projected on the k-charge:

$$W_k(C) \sim \exp i\frac{g}{N}\int_{S(C)} d\vec{S}.Tr\vec{B}Y_k \tag{8.13}$$

with the magnetic field strength $\vec{B}$ replacing its electric counterpart $\vec{E}$. This is incorrect for at least one reason: as discussed in section (7) the Wilsonloop with N-allity k defines a periodic tension σ_k only after averaging. Now the left hand side of the equation is periodic. So the equation must be considered in thermal-averaged form. There is much more to say about this formula!

But if we accept this flux representation on the basis of its analogy with the 't Hooft loop, then the same reasoning as for the 't Hooft loop applies and one concludes that the tension σ_k scales as its electric analogue:

$$\sigma_k(T) = \frac{k(N-k)}{(N-1)}\sigma_1(T) \tag{8.14}$$

All we have to do now is to hear the verdict of the lattice; the ratios found by simulation [28] are close- within a percent for the central value-:

$$SU(4) : \sigma_2/\sigma_1 = 1.3548 \pm 0.0064$$

$$SU(6) : \sigma_2/\sigma_1 = 1.6160 \pm 0.0086; \sigma_3/\sigma_1 = 1.808 \pm 0.025$$

The results are that precise, that you see a two standard deviation, except for the second ratio of SU(6). As we said, magnetic quasi-particles are dilute but only approximately free.

In conclusion, lattice results are quite encouraging. More simulations are under way [52].

9 Epilogue

We highlighted some of the salient problems in the critical region of deconfined and chirally restored QCD. Universality works well, when it is defined at all.

In the region above $2T_c$ we studied the indications that perturbation theory works well if the excitations of the magnetic sector are included. These excitations, the magnetic quasi-particles ("magnons"), may be considered as screened quasi-particles. 3d simulations suggest that their screening is so strong that they form approximately independent lumps. This idea is strikingly vindicated by the dynamically small ratio of magnetic screening over string tension.

The idea works quite well in the Wilson loop sector. And the result for the Debye mass again supports this idea; including the magnetic quasi-particles corrects a result which was off by more than O(1) without! The non-perturbative part of the pressure to order g^6 is the partial pressure of the magnetic gluon gas.

Unfortunately our arguments are at present only good enough for parametric statements. In cases like the Wilson loops we get rid of the unknown constants through ratios. A curious situation prevails: we are indirectly witness of the magnetic quasi-particles, but they remain for the time being the Poltergeists of the magnetic sector. Of course they are only approximately free: their density and pressure are subject to a logarithmic rescaling of the temperature $g^2(T)T$, and hence at asymptotic temperatures the Stefan Boltzmann limit prevails.

10 Suggestions for further reading

Textbooks:

- J. Kapusta, Finite Temperature Field Theory, CUP 1989.
- M. Lebellac, Thermal field Theory, CUP 1996.

Schools:

- M. Shaposhnikov (Erice 1996), High Temperature Effective Field Theory.
- M. Laine (Trieste 2002), Finite Temperature Field Theory.
- F. Karsch, Lect.Notes Phys. 583 (2002) 209-249, Lattice QCD at High Temperature and Density.
- K.Kanaya, An introduction to finite temperature QCD on the lattice, Prog.Theor.Phys.Suppl.131; 73, 1998; hep-lat/9804006.
- Robert D. Pisarski, Notes on the Deconfining Phasetransition. Proceedings of Cargese Summer School on QCD Perspectives on Hot and Dense Matter, Cargese, France, 6-18 Aug 2001; hep-ph/0203271.
- A.K. Rebhan, Hard thermal loops and QCD thermodynamics. Proceedings of Cargese Summer School on QCD Perspectives on Hot and Dense Matter, Cargese, France, 6-18 Aug 2001; hep-ph/0111341.

Acknowledgements

My thanks are due to Prof. Zichichi and Prof. 't Hooft for their invitation to this school. For discussions on the subject I am indebted to Dietrich Boedeker, Philippe de Forcrand, Pierre Giovannangeli, Christian Hoelbling, Oliver Jahn, Frithjof Karsch, Alex Kovner, Mikko Laine, Owe Philipsen, Robert Pisarski, Tony Rebhan, Dan Son, Misha Stephanov and Michael Teper.

Appendix A: Free energy of heavy source and Wilson line average

Below we prove eq.(3.9) in the main text. It relates the free energy of the gauge averaged heavy quark state to the thermal average of the Wilson line.

Start with the gauge average of a state with one heavy quark. The state with one quark, and eigenstate of the field operator $\vec{A}$ is

$$\psi_i^\dagger|\vec{A}\rangle. \tag{10.1}$$

We have suppressed reference to the position $\vec{x}$ of the quark. Gauss' operator is $G(A_0) = Tr\vec{E}.\vec{D}A_0 + \psi^\dagger A_0\psi$ for an infinitesimal transformation. A_0 is dimensionless and is up to a dimensionful multiplicative constant the scalar potential at Euclidean time τ, as in eq.(3.7). Integration over three-space is understood. To render the state, eq.(10.1), gauge invariant, we have to take the gauge projector $P_G \equiv \int DA_0 \exp iG(A_0)$ and act with it on the state. So the free energy excess due to the quark becomes:

$$\exp -F_\psi = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}_{\vec{A}}\langle\vec{A}|\psi_i P_G \exp -H/T P_G \psi_i^\dagger|\vec{A}\rangle / \mathrm{Tr}_{\vec{A}}\langle\vec{A}|P_G \exp -H/T P_G|\vec{A}\rangle \tag{10.2}$$

The gauge projector and the Hamiltonian commute. As in the usual transcription of the free energy into path integral language, we write for a partition of the Euclidean time interval into n bits $\delta\tau = \frac{1}{nT}$:

$$P_G \exp -H/T P_G = P_G \exp -\delta\tau H P_G \exp -\delta\tau H P_G \ldots\ldots P_G, \exp -\delta\tau H P_G \tag{10.3}$$

using that P_G is a projector. The transcription is identical to that of the free energy, except that after rewriting the n Hamiltonian factors as a path integral we are still left with the n fermionic projectors $\exp i\psi^\dagger A_0\psi$. Every single one of them has the effect - through the canonical anticommutation relations-:

$$\exp\,(i\psi^\dagger A_0\psi)\psi_i^\dagger|\vec{A}\rangle = \psi_k^\dagger(\exp iA_0)_{ki}|\vec{A}\rangle. \tag{10.4}$$

The reader will recognize the string bit appearing in eq.(3.7) Repeat the operation in eq.(10.4) n times and the net result is the familiar time ordered exponential $\mathcal{P}(A_0)$ multiplied on the left by the fermion operators:

$$\psi_i\psi_k^\dagger(\mathcal{P}(A_0))_{ki}.$$

Using once more the anti-commutation relations, one finds the anticipated relation between the thermal average of the Wilson line and the free energy excess:

$$\exp -F_\psi = \int DA \exp -\frac{1}{g^2}S(A)\frac{1}{N}\mathrm{Tr}\mathcal{P}(A_0)/\int DA \exp -\frac{1}{g^2}S(A) \equiv \langle P(A_0)\rangle. \tag{10.5}$$

as in the main text, eq.(3.9).

Corollary

What is true for one heavy quark can be shown along the same lines to be true for the correlator of heavy quarks: it is given by the correlator of two Wilson lines.

References

[1] For a review see: R. D. Pisarski, nucl-th/0212015; to appear in Proceedings Quark Matter 2002, Nantes.

[2] D. Bailin, A. Love; Physics. Rept. 107,325, (1984); M. G. Alford, K. Rajagopal, F. Wilczek, Nucl.Phys.B537:443-458,1999; hep-ph/9804403.

[3] A.M. Polyakov, Phys.Lett.B72,477 (1978); L. Susskind, Phys.Rev.D20,2610 (1979).

[4] K. Kanaya, An introduction to finite temperature QCD on the lattice, Prog.Theor.Phys.Suppl.131:73-105,1998; hep-lat/9804006.

[5] S. Hands, Contemp.Phys.42:209, 2001.

[6] F. Karsch, Lect.Notes Phys. 583 (2002) 209-249, Lattice QCD at High Temperature and Density.

[7] P. de Forcrand, O. Philipsen, Nucl.Phys.B642, 290, 2002; hep-lat/0205016.

[8] G.'t Hooft, Nucl.Phys.B138, 1 (1978); Nucl.Phys.B153:141,1979 .

[9] J. Kapusta, Finite Temperature Field Theory, Cambridge, 1989. M. Lebellac, Thermal Field Theory, CUP (1996).

[10] M. Shaposhnikov, Erice 1996, Effective theories and fundamental interactions, 360-383; hep-ph/9610247.

[11] G. 't Hooft, in High Energy Physics, ed. A. Zichichi (Editrice Compositori Bologna, 1976); S. Mandelstam, Phys. Rep. 23C (1976), 245 T. G. Kovács and E. T. Tombulis, Phys. Rev. D57 (1998) 4054; T. G. Kovács and E. T. Tombulis, Phys. Lett. B463 (1999) 104; J. M. Cornwall, Phys. Rev. D57 (1998) 7589; J. M. Cornwall, Phys. Rev. D58 (1998) 1250; R. Bertle, M. Faber, J. Greensite, S. Olejnik, Nucl. Phys. Proc.Suppl. 83 (2000) 425; M. Engelhardt, K. Langfeld, H. Reinhardt and O. Tennert, Phys. Lett. B431 (1998) 141.

[12] K. Kajantie, M. Laine, K. Rummukainen, M. Shaposhnikov, Phys.Rev.Lett.79, 3130 (1997); hep-ph/9704416.

[13] K. Kajantie, M. Laine, K. Rummukainen, M. Shaposhnikov, Nucl. Phys. B503,(1997),357; hep-ph/9704416.

[14] S. Chapman, Phys.Rev.D50:5308-5313,1994; hep-ph/9407313.

[15] S.Z.Huang, M. Lissia,Nucl.Phys.B438,54,1995.

[16] T. Applequist, R.D. Pisarski, Phys.Rev.D23,2305,(1981); P. Ginsparg, Nucl.Phys.B170,388, (1980).

[17] P. Giovannangeli, C.P. Korthals Altes, Nucl.Phys.B608, 203 (2001), hep-ph/0102022; hep-ph/0212298.

[18] S. Bronoff, Thèse, Faculté de Luminy, Université de la Méditerranée, 1998.

[19] A. D. Linde, Phys.Lett. B96, 289(1980).

[20] K. Kajantie, M. Laine, K. Rummukainen and Y. Schroeder, hep-ph/0211321, computes the 6th order log and reviews the lower order coefficients.

[21] A. Hart, O. Philipsen, Nucl.Phys.B572:243-265,2000; hep-lat/9908041.

[22] A. Hart, M. Laine, O. Philipsen, Nucl.Phys.B586:443-474,2000; hep-ph/0004060.

[23] F. Karsch, E. Laermann, M. Lutgemeier, Phys.Lett.B346:94-98,1995, hep-lat/9411020.

[24] J. Kuti, J. Polonyi, K. Szlachanyi, Phys.Lett.B98, 199 (1981);L. D. McLerran, B. Svetitsky, Phys. Lett.B98, 195 (1981); N. Weiss, Phys.Rev.D24, 475 (1981).

[25] S. Gavin, A. Gocksch, R. D. Pisarski, Phys.RevD49, 3079 (1994)

[26] R. D. Pisarski, F. Wilczek, Phys.Rev.D29, 338 (1984)

[27] G.'t Hooft, Phys.Rev.Lett.37, 8,1976.

[28] B. Lucini, M. Teper, Phys.Rev.D64,105019,2001; hep-lat/0107007.

[29] F. Karsch, M. Luetgemaier, Nucl.Phys.B550: 449-464, 1998; hep-lat/9812023; F. Karsch, in: Copenhagen 1998, Strong and Electroweak Matter 101-111; hep-lat/9903031

[30] A.K. Rebhan, Phys.Rev.D48, 3967,1993; hep-ph/9308232.

[31] R. D. Pisarski, hep-ph/0203271, to appear in the proceedings of Cargese Summer School on QCD Perspectives on Hot and Dense Matter, Cargese, France.

[32] E. Braaten and A. Nieto, Phys. Rev. D 51 (1995) 6990 [hep-ph/9501375]. E. Braaten and A. Nieto, Phys. Rev. D 53 (1996) 3421 [hep-ph/9510408].

[33] G. Boyd *et al.*, Nucl. Phys. B 469 (1996) 419 [hep-lat/9602007]; Nucl. Phys. B 478 (1996) 335 [hep-lat/9605004]; B. Beinlich, F. Karsch, E. Laermann and A. Peikert, Eur. Phys. J. C 6 (1999) 133 [hep-lat/9707023]; M. Okamoto *et al.* [CP-PACS Collaboration], Phys. Rev. D 60 (1999) 094510 [hep-lat/9905005].

[34] P. Bacilieri et al., Phys.Rev.Lett.61(1988),1545; F.R. Brown et al,ibid. 61 (1988) 2058; A. Ukawa, Nucl. Phys. Proc.Suppl.B17 (1990) 118, ibid. 53 (1997) 106; E. Laermann, ibid. 63 (1988) 114; F. Karsch, ibid. 83 (2000) 14.

[35] S. Ohta, M. Wingate, Nucl. Phys. Proc. Suppl. B73 (1999), 435; ibid. 83 (2000) 381; Phys.Rev.D 63 (2001)094502; R. V. Gavai, hep-lat/0110054.

[36] B. Lucini, M. Teper, U. Wenger, Phys.Lett.B545, 197 (2002) , hep-lat/0206029.

[37] S. Datta, S. Gupta, Phys.Lett.B471, 382 (2000);hep-lat/9906023; hep-lat/0208001

[38] T. Bhattacharya, A. Gocksch, C.P. Korthals Altes, R. D. Pisarski, Nucl.Phys.B 383, (1992),497; Phys.Rev.Lett.66,998 (1991); C.P. Korthals Altes, Nucl.Phys.B420 (1994), 637.

[39] J. Fingberg, U. Heller, F. Karsch, Nucl. Phys.B 392,(1993), 493;hep-lat/9208021. S. Gupta, Phys. Rev.D64 (2001) 034507.

[40] P. Arnold, C. Zhai, Phys.Rev.D51, 1906, 1995; hep-ph/9410360; C. Zhai, B. Kastening, Phys.Rev.D52, 7232, 1995; hep-th/9507380.

[41] P. Arnold, L. Yaffe, Phys.RevD52 (1995) 7208; hep-ph/9508280.

[42] D. Diakonov, V. Petrov, Phys.Lett.B224; 131, 1989; J.Exp.Theor.Phys.92, 905, 2001; hep-th/0008035.

[43] C. P. Korthals Altes, A. Kovner, Phys.Rev.D62, 096008, 2000; hep-ph/0004052.

[44] C. P. Korthals Altes, A. Kovner, M. Stephanov, Phys.Lett.B469, 205, 1999; hep-ph/9909516.

[45] A. Ukawa, P. Windey and A. Guth, Phys. Rev. D21 (1980) 1013; J. Groeneveld, J. Jurkiewicz, C.P. Korthals Altes, Phys.Scripta 23, 1022, 1981

[46] M. Teper, Phys.Rev.D59, 014512, 1999; hep-lat/9804008.

[47] Ph. de Forcrand, M.d'Elia, M. Pepe, Phys.Rev.Lett.86, 1438 (2001); hep-lat/0007034.

[48] C. Hoelbling, C. Rebbi, V.A. Rubakov, Phys.Rev.D63:034506,2001; hep-lat/0003010; C. Hoebling, in preparation.

[49] K. Farakos, P. de Forcrand, C.P. Korthals Altes, M. Laine, M. Vettorazzo, hep-ph/0207343, to appear in Nucl.Phys.B.

[50] C.P. Korthals Altes and M. Laine, Phys. Lett.B 511 (2001) 269; hep-ph/0104031; for the T=0 case: V. M. Belyaev, I.I. Kogan, Mod. Phys. Lett. A7 (1992),117.

[51] A. Kovner, Int.J.Mod.Phys.A17:2113, 2002; hep-th/0211248

[52] C. Hoelbling et al., in preparation.

CHAIRMAN: C.P. KORTHALS ALTES

Scientific Secretaries: R. Matyszkiewicz, M. Ramtohul

DISCUSSION

- *Ramtohul:*

The Lagrangian presented was in three dimensions and does not appear to be Lorentz invariant. Could you explain the origin of this?

- *Korthals Altes:*

It is obvious why Lorentz invariance is not there because the heat bath is a preferred reference frame and that has enormous consequences. First of all, the lack of Lorentz invariance allows the Coulomb potential to develop a mass and so behaves differently from its magnetic counterpart.

Secondly in gauge theories the electric sector is essentially still perturbative. The magnetic sector is non-perturbative despite the fact we have asymptotic freedom. This is very important and has probably very interesting consequences, one of which I tried to discuss in the last part of my talk.

- *Shuryak:*

In the last page of the talk, lumps of magnetic field were seen on the lattice. What are they and who found them?

- *Korthals Altes:*

Perhaps they have already been found. The idea is the following. In principle the Wilson loop should detect magnetic activity in the plasma. The Wilson loop has an area law from zero temperature onwards and so in that sense it is not much of an order parameter, it does not tell you anything dramatic at the onset of the transition. What happens is that it starts to grow with temperature at the onset of the transition. We understand on the basis of the dimension law that says if you look at the average of the Wilson loop and you are interested in the area law, then you can integrate out all the hard modes in the path integral. This is because the hard modes can only contribute to the perimeter because of the short distance effect. Once you have done that, you can integrate out the static electric modes because they do not appear at all in the spatial Wilson loop itself. This gives the effective 3-dimensional magnetostatic action and that determines what the surface tension of the fundamental Wilson loop will be. In other words, the surface tension looks back into the static magnetic sector. The static magnetic sector has only one dimensionful constant, i.e. g^2 T. By dimensional analysis you see that the surface tension will go like $(g^2 T)^2$ times some number and that number is not perturbative. If you look at four-dimensional simulations it is remarkable that this behaviour sets in at rather low temperatures, just

about $2T_c$. The data here is matched by this very simple formula and g^2 T scaling according to asymptotic freedom. Now the question is, can we understand how the tension is going to change when I go from a fundamental Wilson loop to a Wilson loop constructed with a multiquark representation, say k quarks?

There is a simple model which is adhoc but seems to work and goes as follows. You assume that the path integral for the three-dimensional Yang-Mills can be written as a partition function of a gas of lumps. These lumps have nothing to do with semi-classical objects, they are really quantum objects. Nothing is known of lumps in an unbroken 3-dimensional Yang-Mills which are classical, so these are really quantum lumps and we believe that they are there. Their screening length is assumed to be much smaller than the distance between lumps. It is obvious that the screening length and inter-particle distance are both expressed in the essential scale g^2 T of 3-dimensional Yang-Mills. The density of these objects is g^6 T^3 for dimensional reasons. The next thing to pursue is the analogy with what was said about the gluon gas of quasi particles: these lumps carry magnetic charge of plus or minus $2\pi/g$ times the number of colours. There are $N^2 -1$ species of lumps. Only k(N-k) species will give the disordering factor minus one in the way that the 't Hooft loop was disordered by the gluonic quasi-particles. If you believe these two things, then you get that sigma(k) should be proportional to k(N-k). This number is the number of species that has charge plus or minus N times $2\pi/g$ with respect to the magnetic charge. When you do this, then you ask your lattice friends and what they see is very encouraging: the predicted ratios are reproducible within a percent!

Michael Teper has been scanning the literature and has seen that in the beginning of the nineties there is a work by Bob Mawhinney and Johnny Duncan where they did 3-dimensional SU(2) and saw a lovely structure of 3-dimensional lumps for a 3-dimensional Yang-Mills. Mike has presumably a better understanding of what is going on, because they used cooling techniques. Whether this is appropriate for quantum lumps is in a way questionable but this is something I am not so familiar with.

- *Shuryak:*
  So they cut the flux out, like monopoles?

- *Korthals Altes:*
  The flux is screened.

- *Shuryak:*
  Screened by these magnetic scales?

- *Korthals Altes:*
  Absolutely, and of course you would say that cannot be, because if I think of screening, then normally statistical Boltzmann screening is such that within the screening length there are lots of particles. Magnetic screening is already there at zero

temperature as a consequence of electric confinement, as noted by 't Hooft long ago. In other words, magnetic screening is not really statistical screening, it is something different that we only partially understand.

- *Maas:*

People doing perturbation theory run into severe problems at finite temperature. The theory seems not even to be Borel summable and there have even been claims that the convergence radius of a perturbative expansion at finite temperature could be zero. Do you work on this fact by inclusion of the lattice coefficients? Secondly, recent lattice calculations showed that the Stefan-Boltzman limit only becomes asymptotically reached at temperatures like 10^{10} T_c. Is this limit suitable considering your results?

- *Korthals Altes:*

To illustrate convergence, we can look at the perturbation series for the pressure as it is known today. Include order g^2 corrections to Stefan-Boltzman. Then you can include g^3 and a disaster happens, all of a sudden it blows up. When you include g^4 and g^5, it becomes a little better. The question is what happens when we include g^6. The calculations for g^6 are under way, but will take quite some time; it is a 4-loop calculation plus a lattice determination of the non-perturbative part of the p_6 coefficient.

You cannot really convince anybody with a series having this type of behaviour. Recently Kajantie and collaborators have taken the following point of view. You saw that the bad behaviour starts with g^3, that is typically the soft electric sector. What they say is to take the hard part of the coefficients that you get by summing over the heavy mass parameters proportional to the temperature and calculate a series in terms of the hard contributions. This is a series in terms of g^2, g^4, g^6 and is a series that behaves well. What do you do with the rest? They propose to take the electric Lagrangian and treat it with lattice methods and so resum all the electric contributions, the first one of which is the bad guy g^3. What they get is a much better result. They assume a value for the non-perturbative part of p_6, the sixth order coefficient. If they assume it is about ten, then they get a reasonable connection with data, but below $4T_c$ it is not good any more. This is one attempt to try and resum the electric perturbative contributions for the series.

The other part of the question about Borel summability I cannot answer. I do not think anyone can say whether this series is Borel summable as far as I know.

- *Stamen:*

What happens at the phase transition point? You talk about confined and deconfined $\bar{q}q$ pairs. How do you link that to nature where we have π's, ρ's, J/Ψ's etc. which all have a different radius? I would expect that beyond the transition point there are still $\bar{q}q$ pairs (e.g. J/Ψ) bound together.

- *Korthals Altes:*

What you say is correct for a second order transition. Then the Debye mass starts out from zero at T_c, therefore the flux tube keeping the hadrons together is not affected. As you go up in T, the screening length becomes smaller and starts to unbind the largest hadrons, but please remember that near T_c it is mostly low-lying mesons that are thermally excited. For a first order transition, the screening mass can jump to a high enough value so that all hadrons deconfine at once.

- *Stamen:*

Is this for all the hadrons at the same time?

- *Korthals Altes:*

Yes, because the binding is the flux and once you screen the flux it is finished. It could also be second order then the screening mass starts very very small and so very very large screening.

- *Haidt:*

How much time do you need for thermalisation?

- *Korthals Altes:*

I have never thought about that, but somebody like Edward could answer that question.

- *Shuryak:*

In the regime where it is calculated in pQCD, what we are now discussing is very high temperature where we hope the perturbative formulas will apply. Then we are supposed to calculate the mean free path perturbatively for gluons and that is of the order of α_s^2 and T log α_s. The reciprocal of this is the mean free path.

- *Haidt:*

How much is that in Fermi?

- *Shuryak:*

That depends on T. We are discussing high T and have not put numbers in yet. Chris has explained that there are the following scales, T, gT, and g^2T where magnetic phenomena happen and the mean free path which depends on the scattering amplitude is an even larger scale in perturbation theory. It has α_s^2 at g^4 so particles move far and all these calculations are corrected at very high temperatures, so maybe at LHC, or whatever, will come to this region. What I said about RHIC is that we thought with momenta of the order of 1-2 GeV we are in this region. This is not true as the mean free path happens to be much shorter. Maybe at higher temperatures the perturbative road should work.

- *Haidt:*

Is there a minimum size of a quark gluon cluster? Or does it have to be very extended in space?

- *Korthals Altes:*

It depends on the equation of state. For thermodynamic behaviour we need a thousand particles or something.

- *Haidt:*

So for electron-proton scattering we will never come upon this state.

- *Korthals Altes:*

There are surely people more competent than I who can answer that question.

- *Haidt:*

This is a very dynamical question, I suppose you were discussing more from a different point of view.

- *Korthals Altes:*

Yes, I was talking about the equilibrium situation.

Physics of QCD Instantons

E.V.Shuryak
Department of Physics and Astronomy,
State University of New York,
Stony Brook NY 11794 USA
E-mail: shuryak@dau.physics.sunysb.edu

These lectures cover applications of the semiclassical theory to QCD. We start with tunneling through the topological barrier described by the *instantons*, and review their multiple applications in QCD. The *interacting instanton liquid model* describes the ensemble of instantons, in its multiple phases. The QCD vacuum happens to be a disordered "liquid", which breaks spontaneously chiral symmetry. The high-T phase is kind of a "molecular gas" which does not break any symmetries. High density phase is a color superconductor: here instantons form chains or diquarks. As recent applications we briefly discuss correlation functions and pion and nucleon form-factors. We discuss in some detail also the fate of instantons at large number of colors N_c. The second set of topics are applications of the semiclassical theory of high energy collisions: instead of tunneling through the barrier one has in this case excitation from under the barrier to it. These methods were recently applied to the old problem of soft Pomeron, and especially for explanations of puzzling behaviors of heavy ion collisions observed recently at Relativistic Heavy Ion Collider in Brookhaven.

1 Introduction

The opportunity to give 2.5 lectures at the 2002 Erice school was especially appropriate, because I tried to cover 2.5 different subjects. They are: (i) Topological tunneling phenomena of the gauge fields in the QCD vacuum and hadrons, (ii) Production of topological excitations of the glue in high energy collisions; plus a 0.5 lecture on (iii) a brief review of exciting first results from RHIC[a].

The first thing to note about those three subjects: their status is completely different. The instanton-based theory of tunneling is well established. Classical instanton solution and basic semiclassical methods were developed back in the 1970's, with important applications such as solution of the U(1) problem. The properties of the instanton ensemble and their role in *chiral symmetry breaking* were understood in the 1980's, producing in the 1990's a very detailed semiclassical instanton liquid model reproducing many correlations functions and hadronic phenomena. Lattice practitioners had tested many of its predictions, and especially the main idea – that the lowest Dirac eigenstates

[a]The final lecture of the school by T.a gives full description of the subject: I describe only some (most puzzling) aspects relevant for physics issues I discuss.

of QCD are dominated by subspace of instanton zero modes – recently passed several important tests. A rather complete summary of this material can be found in a review [1]: I am also currently writing a second edition of my book "QCD vacuum, hadrons..." where many aspects of the non-perturbative QCD will be discussed pedagogically, for beginners.

New developments continue to appear: and in these lectures I will specifically describe (i) comparison between vector and axial correlators obtained from tau lepton decay into hadrons to instanton predictions, (ii) the instanton contributions to pion form-factor; and (iii) the fate of instantons in the large N_c limit[b].

In contrast to the semiclassical theory of the QCD vacuum, applications of the semiclassical methods to high energy collisions is relatively new. Its main ideas originate from the early 1990's, when intense debate took place whether one can observe a baryon number violation in electroweak theory at a $\sim 10\,TeV$ scale. Important new analytic results - such as the internal structure and energy of pure gluonic clusters which describe the very topological potential in question, and which play the role of the turning points, as well as their subsequent Minkowski evolution [2] just appeared. Its applications to pp (the old soft Pomeron problem) and heavy ion high energy collisions have just started, and the role of these phenomena is still hotly debated.

RHIC data are very new indeed: the first full-setting run has just been completed a year ago. Needless to say, the interpretation of fresh data is always somewhat dangerous.

But, I am sure that school participants would forgive that. I am quite convinced that the lecturers must keep balance, providing both a summary of well established results and a list of puzzles they and their colleagues are thinking about as they speak. Only in this way we will convey to the students all the excitement of doing a living science, with all the intellectual challenges new discoveries constantly provide. Also, who else if not the participants of a school like that are suppose to solve eventually those puzzles?

2 Instantons and chiral symmetries

2.1 The topological barrier and tunneling

Schematically YM quantized field can be viewed as many coupled non-linear oscillators with simple potential of the type $\vec{B}^2 \sim O(A^2) + O(A^3) + O(A^4)$, with coordinates being the potentials A^a_m, a=1-3 (since we will limit ourselves

[b] Apart from being a famous problem by itself, in my previous 1998 Erice lectures I was specifically asked in a question period a set of questions about it by Prof.Witten. I am happy to be able to answer them now.

to SU(2)) and m=1-3. In the $A_0 = 0$ gauge $\frac{d\vec{A}}{dt} = \vec{E}$ and the electric part of the energy is identified as the kinetic term.

One combination of fields, the *Chern-Simons number*, have certain topological properties

$$N_{CS} = \int d^3x K_0, \;\; K_\mu = -\frac{1}{32\pi^2}\epsilon^{\mu\nu\rho\sigma}(\mathcal{G}^a_{\nu\rho}\mathcal{A}^a_\sigma - \frac{g}{3}\epsilon^{abc}\mathcal{A}^a_\nu\mathcal{A}^b_\rho\mathcal{A}^c_\sigma) \qquad (1)$$

The *potential energy* of Yang-Mills field versus this coordinate is schematically shown in Fig.1. It is a periodic function, with zeros at all *integer* points: those are "classical vacua", with zero field strengths but non-zero (and topologically distinct) A^a_m.

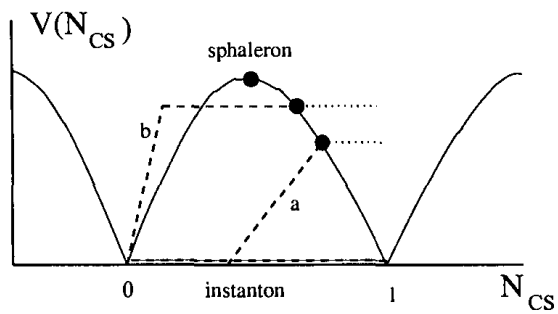


Figure 1: Potential energy of Yang-Mills field versus the Chern-Simons number N_{cs}.

Tunneling through classically impenetrable barrier is the most amazing consequence of quantum mechanics. Semiclassically it can be described by classical solutions of equations of motion – the Yang-Mills eqns in this case – in the Euclidean time with *zero* energy are the tunneling paths – called *instantons*. Those start at one minimum of the potential and end up in another. Anti-instantons make a tunneling back. In Euclidean space-time, such as used in lattice simulations, the instanton is a localized 4-dimensional soliton with a finite action found in 1975 by A.Polyakov and collaborators[3]. Its form in the so called singular gauge has the form

$$gA^a_\mu(x) = 2\frac{x_\nu}{x^2}\frac{\overline{\eta}_{a\mu\nu}\rho^2}{x^2+\rho^2}. \qquad (2)$$

where η is the so called 't Hooft symbol, a combination of zeros and ones with some algebraic background related to gamma matrices. Note that the instantons have variable sizes ρ: that would be very important later. The corresponding field strength is non-singular at x=0 and also very well localized

$$(gG^a_{\mu\nu})^2 = \frac{192\rho^4}{(x^2+\rho^2)^4} \qquad (3)$$

so the action is finite $S = 8\pi^2/g^2$ and the tunneling probability is thus $P \sim exp(-8\pi^2/g^2)$. The prexponent has been calculated by 't Hooft[4]. In the first half of these lectures we will discuss what the ensemble of those objects in vacuum looks like and what instantons can do in general.

Since tunneling probability is finite, the quantum vacuum state can be viewed as a superposition of all of those vacua. As we know from quantum mechanics, states in a periodic potential are characterized by a *quasi−momentum* called θ in field theory. From experiment we know that it is very close to zero, although we do not known why. Anyway, the quantum vacuum is thus an equal superposition of all topological vacua.

In the second half we will discuss high energy collisions. In this case *a sudden localization* of all quantum coordinates happens, including the topological one N_{CS}. The system suddenly finds itself *at or above* the barrier (see the dashed line (a) in Fig.1). Another possibility (shown by the dashed line (b) in Fig.1) is that a system at the collision moment is *not* under barrier, but becomes able to tunnel through it after it gets excited enough.

Whichever way the system is driven, it emerges out of the barrier via what we call *"a turning state"*, a relative of the sphaleron[c] solution of electroweak theory. This is a point where the path crosses the barrier and the total energy is equal to the potential one. Here the potential energy is equal to the total one, with zero momentum (the electric field). This is why those objects are born into our world as pure *gluomagnetic* clusters.

From there starts the real time motion outside the barrier (shown by horizontal dotted lines): here the action is real and $|e^{iS}| = 1$. That means that whatever happens at this Minkowski stage has the probability 1 and cannot affect the total cross section of the process: this part is only needed for predicting the properties of the final state.

I hope that existence of under-the-barrier and over-the-barrier parts, separated by turning (or stopping) points is a concept well familiar from quantum mechanics. Their field theory analogs we will discuss below.

2.2 Chiral symmetries and their breaking

The two main non-perturbative phenomena in gauge theories are *confinement* and *chiral symmetry breaking.* Discussion of the former at this school is done by Prof.Strassler, so I would concentrate on the latter. (This is very fortunate since the confinement problem resists solution for a quarter of a century while chiral symmetry breaking we now understand in fine detail.)

[c] "Ready to fall" in Greek, according to Klinkhammer and Manton.

Let me start the lecture by reminding a few well-known facts from the textbooks about massless fermions and related symmetries. If quark masses are ignored, the fermion part of the QCD Lagrangian becomes a sum of two independent terms, with left and right-handed quarks. The possibility to rotate these in flavor space *independently* generates *two* additional "chiral" symmetries, $U(1)_A$ and $SU(N_f)$, which have rather different fates.

The $U(1)_A$ one (generated by $exp(i\phi\gamma_5)$ rotation) is explicitly broken by the so called chiral anomaly, and so at quantum level it is simply *not* a symmetry of QCD. The strength of its violation can be seen from a deviation of the pseudoscalar singlet η' mass (959 MeV) from that of a pion/kaon/eta multiplet: note that it is surprisingly large.

The $SU(N_f)$ part of the chiral symmetry is *spontaneously broken*, the QCD vacuum is asymmetric. Its measure is the so called quark condensate $< \bar{q}q >$. By Goldstone theorem, massless modes (rotations to other equivalent vacua) appear, which are pions. General features of their interactions are described by chiral effective Lagrangians.

The main question we are going to discuss is *what is the underlying dynamics* of these phenomena in QCD.

The *violation of U(1) by instantons* could be understood in two ways. The first follows from chiral anomaly, which tells that the corresponding current is not conserved and its divergence is proportional to the topological charge $\sim \int d^4x G\tilde{G}$. Since instantons have unit topological charge, they contribute $2N_f$ units of the axial charge. But more detailed description of how it happens was provided in a classic paper by G.t'Hooft[4]. He found fermionic zero modes and new instanton-induced $2 * N_f$-fermion effective interaction of particular structure. The quantitative part of the U(1) problem was later related to the so called "topological susceptibility" by Witten and Veneziano, and recent lattice studies have left no doubts that these are indeed completely saturated by well-identified instantons. We return to this in the section 6.

The $SU(N_f)$ chiral symmetry is not violated by a single instanton: nevertheless their ensemble does it. But before we discuss it, it is appropriate to recall the earliest attempt to explain chiral symmetry breaking was made as early as 1961[5]. Driven by the analogy to a superconductivity and Cooper pairing, Nambu and Jona-Lasinio have shown that if there exists a sufficiently strong 4-fermion interaction generating attraction between quark and antiquark in the scalar channel, it can also create pairs, re-arrange the vacuum. If it happens, there appears a non-zero quark condensate $< \bar{q}q >$, which relates phases of the right- and left-handed quarks and thus breaks this symmetry. Also there appears a "gap" at the surface of the Dirac sea, the quark effective mass. It was essentially the right idea, although the understanding of the ori-

gin of that attractive interaction and its exact form in QCD was clarified only during the last two decades (see e.g. references in a review [1]). Nambu and his followers had significant freedom of selection of their hypothetical Lagrangian: one can write down many different 4-fermion operators. Chiral symmetry restricts it, but not too much. For example, based on a one-gluon exchange an operator of a vector-vector structure $\bar{q}_L q_L \bar{q}_L q_L + \bar{q}_R q_R \bar{q}_R q_R$ was suggested and used in NJL applications. As far as pions are concerned, they would do as well as scalar-scalar $\bar{q}_R q_L \bar{q}_R q_L + (L \leftrightarrow R)$. One may ask if all 4 quarks can be of the same flavor or not: again, gluon exchanges are flavor blind and would favor flavor independence.

However what was found (first empirically from properties of the correlation functions, then from the success of the so called instanton liquid models, and more recently from the lattice studies) is that there is no short-distance vectorial interaction at all, and no same-flavor (e.g. $\bar{u}u\bar{u}u$) interaction: at small distances one actually finds only the 't Hooft interaction, which is scalar and flavor-non-diagonal, $\bar{u}_R u_L \bar{d}_R d_L$ for 2 flavors. Thus the spontaneous breaking of $SU(N_f)$ $SU(N_f)$ chiral symmetries is also driven by instantons.

However the phenomenon is not seen at the level of one instanton, one needs some knowledge about their ensemble, and particular conditions should be met for it to occur. Otherwise (and this happens at high enough temperature T or large enough number of quark flavors N_f) the chiral symmetry remains unbroken. In 1982 I have found first evidences that the mean density and size of the instantons [6] are about

$$n = n_+ + n_- \approx 1 fm^{-4}; \;\; \rho \approx 1/3 fm \tag{4}$$

It is a rather dilute ensemble because $n\rho^4 \sim 1/81$, but it still is dense enough to be in the chirally broken phase.

A decade later lattice configurations were stripped of the "fog" of quantum fluctuations and their classical content was revealed. In Fig.2 from [7] one can see how it works, so that one can see and count instantons. The values given in (4) were confirmed: but we are still not quite able to evaluate them from first principles.

On the lattice one can study the wave function of the condensate on configuration-per-configuration basis, find the lowest Dirac eigenstates and see what they look like. Qualitative prediction is that all of them are just a superposition of instanton zero modes. And indeed, they turn out to be made of spherically-looking bumps of the same size .3 fm and same density 1 fm^{-4} as seen in Fig.2 and predicted 20 years ago[6]. Furthermore, those bumps are nearly 100 percent correlated with instantons in their size and positions, in each and every configuration. With the now available chiral fermions one was able to

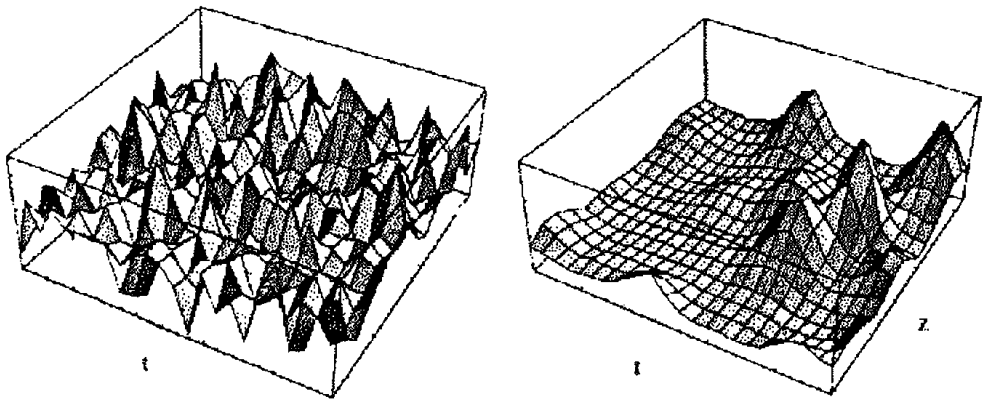


Figure 2: Sample of a gauge configuration before (left) and after (right) cooling. In the latter case instantons are clearly seen. The quantity shown is the action density, and its scale (not shown) is two orders of magnitude larger on the left figure.

test whether each bump is indeed purely chiral (left or right handed): several papers [18] have shown this to be the case, contradicting an earlier provocative paper [19].

2.3 The instanton liquid model

So, the instantons do all what the hypothetical NJL interactions were supposed to do: one can use the same mean field gap eqns, generate pions by Bethe-Solpeter equation etc.

Moreover, one can go much further than the NJL-type approach had ever been able to do. Instead of selecting some subset of diagrams, one can actually solve to *all orders on 't Hooft interaction.* This approach is called interacting instanton liquid model, or IILM for short, it is defined by a partition function

$$Z = \int \Pi_{I,A} d\Gamma(\alpha_{I,A}) exp(-S_{int}) det^{N_f}[i\hat{D} + im] \tag{5}$$

There are basically three parts of it:

1.Semiclassical tunneling probability, which I will denote $d\Gamma(\alpha)$, depending on the collective coordinates of the instanton

2.Gluonic interaction between instantons S_{int}

3.Quark-induced interaction, related to fermionic determinant.

In the SU(3) color group there is 12 collective variables per instanton: 4 coordinates of the center position, the size and 7 color rotation angles. (In SU(3) there are 8 angles, but one is not acting on the instanton solution.) Thus, each "atom" is characterized by these 12 variables. Evaluating the determinant exactly is a quite formidable problem. In practice we factorize the determinant into a low and a high momentum part. The latter is approximated by the product of contributions from individual instantons calculated in Gaussian approximation, whereas the low momentum part associated with the fermionic zero modes of individual instantons is calculated explicitly as the determinant of $N_+ \times N_-$ matrix of the so called overlap matrix elements between the zero modes of all $\bar{I}I$ pairs

$$T_{IA} = \int d^x \psi_A^+(x)(iD_\mu\gamma_\mu)\psi_I(x) \tag{6}$$

Its determinant contains vacuum loop diagrams with the 't Hooft interaction to all orders: this becomes evident if one writes it as a sum of products of its matrix elements. With about $N_+ + N_- \sim$ few hundreds of instantons we typically use in a simulation, the number of vacuum diagrams is huge, its factorial, but computer evaluation of the determinants takes microseconds and allows to keep track of determinant modification during statistical updates.

A simplified version of this model, known as Random Instanton Liquid Model (RILM) ignores the determinant and distributes instantons randomly in color and ordinary space: the size and density are given by (4). It shares all problems with quenched lattice QCD, to a large extent.

Studies of this model revealed that instantons are responsible for a quite significant fraction of non-perturbative phenomena associated with light quarks, their propagation in vacuum and bound states. Their masses are mostly the masses of the "constituent quarks" we already mentioned, and even their spin-dependent forces seem to be instanton-generated as well.

3 What instantons can do: a brief summary

Hadronic spectroscopy and correlators

As explained above, instantons describe chiral symmetry breaking, but they do *not* explain color confinement. In fact their contribution to heavy quark potential is not growing with distance but saturates. Moreover it is surprisingly small numerically[d] and in the $\bar{c}c, \bar{b}b$ spectroscopy one can simply ignore it.

[d]Standard instanton parameters mentioned lead to effective mass of a heavy quark of about 30-40 MeV.

One might think that without confinement there is no hope to have any success in description of hadronic states. And yet, it turns out that many hadrons actually do exist in the "instanton vacuum", as ordinary bound states of the constituent quarks. Moreover, the masses of those states are in much better agreement with data than of most quark models, such as the non-relativistic quark model or the MIT bag[e].

When one takes a look at mesonic masses squared, one finds a very simple splitting pattern

$$m_\pi^2 \approx 0; \qquad m_\rho^2 \approx 0.5; \qquad m_{\eta'}^2 \approx 1\ GeV^2 \tag{7}$$

suggesting that the first and last case have comparable attraction and repulsion, relative to our "anchor", the ρ, with the mass unmodified. This is precisely what one finds in the first order in 't Hooft interaction due to diagrams shown in Fig.3: attraction for pions, repulsion for η' and no effect for ρ. The splitting for these three channels is well seen even at small distances, see Fig.9(b) below, where the relevant correlation functions are calculated in the instanton ensemble.

For readers who are not convinced that there can exist such huge instanton-induced splittings, consider another pair of particles, the I=0 scalar f_0 (known as a σ meson for a long time) and I=1 one a_0. Excluding a pair of states at 980 MeV widely believed to be a $\bar{K}K$ molecule, the lowest states in those channels display even larger splitting

$$m_{f_0}^2 \approx .3\, GeV^2, \qquad m_{a_0}^2 \approx 2\, GeV^2 \tag{8}$$

And again, instantons show attraction and repulsion in those two cases, and explain the splitting *quantitatively*. (Practitioners of the various quark models usually denied the existence of the σ, which did not fit into those models. It however continue to show nicely in many more decay channels and, after two decades of fights, it made its way back to the Particle Tables.)

As we are on the subject, let me make another comment about spectroscopy of light quark mesons. There is a simple generic argument indicating that all those attraction/repulsion effects are quasi-local, when $\bar{q}q$ are close to each other. The *primed* excitations in all those channels repeat the splitting pattern, which is however much weaker, $\delta m/m \sim$ few percent, scaled down approximately as the squared wave function at the origin.

Let me explain a bit how the effect should be calculated: in the first order it can be done rather simply. In the original 't Hooft paper he considered a

[e]It is not apparent in most papers devoted to these models simply because the model builders tend to omit from consideration the most "difficult" hadrons such as π, σ, η' etc.

background field of a single instanton. In the diagrams shown above all lines passing through the instanton (the shadowed circles) the propagators contain the zero mode term

$$S(x,y) = \frac{\psi_0(x)^+\psi_0(y)}{m} + S_{NZM} \tag{9}$$

where ψ_0 is 't Hooft zero mode. The light quark mass in denominator goes to zero in the chiral limit. But before concluding it is infinite in the chiral limit, one should recall that the product of light quark masses appears in the weight (the fermionic determinant) for a single-instanton configuration and can cancel each other, yieldings finite result – the 't Hooft vertex.

However, in a real vacuum there are many instantons and anti-instantons, and so one may wonder how one should change the argumentation in order to account for the effect of a single instanton *out of the ensemble.* A very short answer to that is that in both the fermionic determinant and the fermionic propagators some *effective quark masses* substitute for the bare masses; all of those remain finite in the chiral limit. So, there would be no compensation of infinitely large factors present: otherwise numerical simulations would not be feasible.

In a realistic vacuum of QCD one can ignore configurations with a non-zero net topological charge (those are highly disfavored by the small value of quark masses). Therefore, we will focus at an ensemble with equal number of instanton and anti-instantons, without exact zero modes.

The propagator in such a background field has special features in a subspace spanned by all zero modes. Incorporating diagrams in all orders in 't Hooft interaction, the propagator can be written as follows

$$S(x,y) \;=\; \sum_{IJ} \psi_I(x) \left(\frac{1}{T+im}\right)_{IJ} \psi_J^\dagger(y) + S_{NZM} \tag{10}$$

and we would ignore the last term representing the contribution of all non-zero modes. The capital letters number all instantons and anti-instantons, $\psi_I(x)$ is the zero mode of the I-th instanton and T_{IJ} denotes the overlap matrix mentioned above. The inverse "hopping matrix" is the amplitude to jump from one instanton to another. The *inverse* of matrix $T+im$ appears because the propagator is the inverse of the Dirac operator.

Here comes the main idea. Suppose we calculate some diagram with several propagators, and all their arguments are either the same point, or points which are sufficiently close to each other, for all $x, y, \ldots$ $|x-y| << R$ ($R \sim 1\,fm$ is the typical instanton separation). If so, the biggest term in this sum is likely

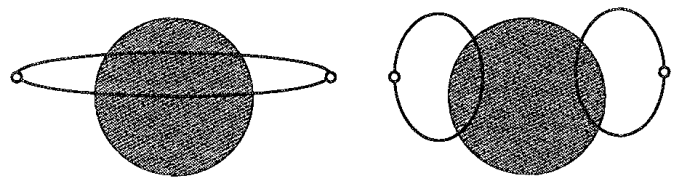


Figure 3: Two diagrams for mesonic correlators in a single-instanton background (the dashed circle). The single loop (a) is the only one for flavored currents, while the unflavored currents have also two-loop contributions of the type (b).

to be associated to the closest instanton, I^*, while all the terms in the sum representing instantons far away from the points x and y will be sub-leading.

$$S(x,y) \approx \psi_{I_*}(x) \left(\frac{1}{T+im}\right)_{I_* I_*} \psi^\dagger_{I_*}(y) \tag{11}$$

So a proper definition of the effective mass involves the inverse of the matrix T_{IA}

$$S(x,y) = \frac{\psi_{I_*}(x)\psi^\dagger_{I_*}(y)}{m^*} \tag{12}$$

containing all the information about the particular configuration of the instanton ensemble. Furthermore, in the diagrams on Fig.3 there are 2 such propagators, so the appropriate mass is determined by averaging a *square* of $(T+im)^{-1}_{I_* I_*} \simeq N^{-1}\sum_I (T+im)^{-1}_{II}$ over ensemble. Numerical value for m^* defined in such way is rather small $60-80\,MeV$, comparable or smaller than the strange quark mass. It is very different from "consituent quark mass" and many other definitions one can invent: therefore the so called mean field approximation with the same mass for all ocasions is just a very crude approximation.

QCD correlation functions

Correlation functions are the main tools used in studies of structure of the QCD vacuum. They can be obtained in several ways. First, they can in many cases be deduced phenomenologically, using a vast set of data accumulated in hadronic physics. Second, they can be directly calculated *ab initio* using quantum field theory methods, such as lattice gauge theory, or semiclassical methods. A significant amount of work has also been done in order to understand their small-distance behavior, based on the Operator Product Expansion

(OPE). The large distance limit can also be understood using effective hadronic approaches or various quark models of hadronic structure.

Large-distance behavior of the correlators yields the hadronic masses. Short-distance instanton effects contribute to the form-factors we will discuss below. The behavior at medium $x \sim 1fm$ is so a large subject that I would not go into it here, but just say the agreement between instanton predictions and data or lattice results is truly amazing. In dozens of channels studied we have not yet seen a single statistically significant difference between them. As a recent "anchor" example in which the first order instanton effects are absent, we will discuss below vector and axial I=1 correlators, for which detailed data exist from tau lepton decays.

Diquarks and baryons

The existence of a strongly attractive interaction in the pseudoscalar quark-antiquark (pion) channel also implies an attractive interaction in the scalar quark-quark (diquark) channel. This can be seen by just Fiertz transformation of 't Hooft interaction into $(\bar{q}\bar{q})(qq)$ form.

Another argument for deeply bound diquarks came from the bi-color ($N_c = 2$) theory: in it the scalar-isoscalar diquark is degenerate with pions and vector-isovector with ρ. By continuity from $N_c = 2$ to 3, a trace of it should be seen in real QCD as well. Instanton-induced interaction strength in diquark channel is $1/(N_c - 1)$ of that for $\bar{q}\gamma_5 q$ one. It is the same at $N_c = 2$, zero for large N_c, and is exactly in between for $N_c = 3$.

I became aware of that when we were doing instanton liquid evaluation of baryonic correlation functions in the early 90's: we found that contrary to OPE predictions, the correlation functions for the nucleon and Δ were very different, even at small distances. The instanton-induced qq interaction is phenomenologically very desirable, because it immediately explains why the nucleon and lambda are light, while the delta and sigma are heavy: they have scalar and vector diquarks, respectively.

One testable consequence of that is different form-factors: one may think that at $Q \sim$ few GeV the nucleon form-factor is much large than that for Δ. Of course, the latter can only be done on the lattice, not experimentally. What we know so far is that instanton-based evaluations of the nucleon form-factor [8] agrees with data, and that the measured transitional $N \to \Delta$ form-factor is indeed much softer.

Another consequence is that scalar diquarks drive the transition to Color Superconductivity at high density, since they are the best Cooper pairs available: we return to this issue below.

Parton Distribution Functions (PDFs)

Unfortunately, there is no direct instanton-based calculation of the structure functions and PDFs. The reason is the Euclidean approach can only do moments of the structure functions as expectation values of some operators, which is technically involved. Let me only mention that lattice people do so, and at least one calculation (by MIT group[9]) have tested that their "cooled" configurations (basically the instanton liquid) give the same values of the moments, within a few percent.

There is however a very interesting qualitative idea[10] relating to the puzzling properties of the nucleon sea quarks to the 't Hooft vertex. What we learned from "spin crisis" is that sea quarks are polarized *negatively* relative to valence ones. We also learned with amazement that the sea at $x \sim .01$ is strongly flavor polarized, and it is also *negative* in the sense that the proton has more d sea than u. Both is impossible to explain if the sea is produced by gluons, in a pQCD evolution. However both agree with the features of the 't Hooft vertex: say u_L produces the $\bar{d}_L d_R$ pair. The opposite flavor is enforced by the Pauli principle at the zero mode level[f].

Small x physics and the Pomeron problem

High energy or small x limit of QCD is of course a long standing problem, and deep inelastic scattering such as done at HERA added very interesting information about the large-Q aspect of it. Opinions differ at the moment whether a pQCD prediction in form of BFKL or its modification, or the "saturation" scenario works. What is clear is that we seem to have a power growth $1/x^p$ of PDFs which is different at large and small Q: so let me separate it into soft and hard Pomeron problems, and focus on the former one.

It refers to the specific behaviors of all hadronic (e.g. pp or the γp) cross sections, $\sigma \sim s^{0.08}$. A number of authors[20] have suggested that it can be related with excitation from the under-the-barrier we discussed at the beginning. More specifically, a small index 0.08 is in this case proportional to the cross section of the production of the turning states, the topological clusters. Basically, this small number stems from small probability in the QCD vacuum to be under barrier at the beginning, or the instanton diluteness parameter $n\rho^4 \sim 10^{-2}$.

This power, as well as the pomeron size have been evaluated and gave reasonable numbers. An interesting byproduct found was the *absence of odderon*

[f] One may ask if the effect is thus a trivial consequence of fermi statistics. The answer is that instanton field locks chirality and color of the quarks, making Pauli principle much more robust. Account for blocking in hadronic models like MIT bag leads to tiny assymmetry much smaller than observed.

in this theory, which is due to the SU(2) color nature of fields. As discussed at this meeting, it is in agreement with all the data including recent HERA results, but in disagreement with many models.

Semiclassical soft Pomeron can be viewed as a ladder-type diagram similar to the perturbative BFKL ones. The difference is that the so called Lipatov vertex – 2 virtual gluons into the "physical" gluon – is substituted by a new vertex, with a topological cluster produced in a gg and other collisions instead of a gluon. As we argued above, it is the most natural state which can be excited from *under the barrier*.

Experimentally, high amount of clustering in pp collisions is known for a very long time. Unfortunately, these clusters have not been studied and identified in detail. They may be related with the gluonic clusters the semiclassical theory predicts. One hint for that is very large abundance of "scalar glueball" resonance, which has no place in the usual string fragmentation picture.

The issue is very interesting and we will briefly address it below. It will be studied experimentally at RHIC in its pp and pA mode.

Different phases of QCD

The QCD under normal conditions is in the *chirally asymmetric and confining phase* we are so familiar with: but in the so called extreme conditions it turns to quite different phases. There are at least three different directions in which one expects three different phase transitions. (i) At high temperature $T = T_c \approx 150 MeV$ it undergoes transition to the so called *Quark-Gluon Plasma* (QGP) phase, in which there are no condensates and color interaction is screened[21] rather than confined. (ii) At high density and low T it is believed to be *Color Superconductor* (CSC), in which color symmetry is broken by diquark condensates induced by instantons[22,23] or, at high density, by gluons[24]. (iii) At a sufficiently large number of flavors $N_f > N_f^c$ it should eventually become a *chirally symmetric deconfined* phase.

The map (for the first two) is shown in Fig.4. Most of this diagram can and partly already is spanned in heavy ion collisions. However the CSC phases are unfortunately not relevant for them, and so we will discuss mostly the high-T direction in this section. Let me only add a few words about the others.

All transitions are believed to be driven by instantons which create the following amusing *triality*. In general, there are three attractive channels which compete: (i) the instanton-induced attraction in $\bar{q}q$ channel leading to χ-symmetry breaking. (ii) The instanton-induced attraction in qq which leads to color superconductivity. (iii) The *light-quark-induced* attraction of $\bar{I}I$, which leads to pairing of instantons into "molecules" and a Quark-Gluon

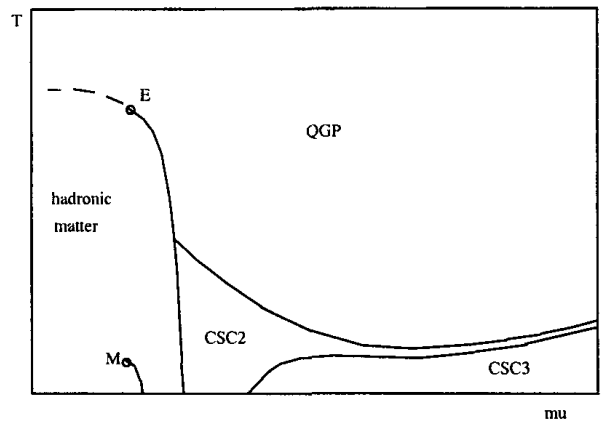


Figure 4: Schematic phase diagram of the QCD phases in coordinates temperature T and baryonic chemical potential μ. The phases denoted by CSC2 and CSC3 are two different color superconductor phases characteristic for 2 and 3 flavors, respectively. The latter is also known as Color Flavor Locked (CFL) phase. E is the endpoint of the 1-st order transition, M is the endpoint of another 1-st order transition, between liquid and gas phases of nuclear matter.

Plasma (QGP) phase without *any* condensates.

The large N_f direction is less studied. At least one reason for that transition is a tendency of instantons and anti-instantons to be bound by the increasing number of fermion lines connecting them, till finally the "instanton liquid" is gone and only finite pieces with zero topology (or neutral "molecules") are left. Calculations in IILM [25] have found that $N_f^c = 5$.

Now we return to the non-zero T case and provide some details. Recall that it can be incorporated in quantum field theory in a very simple way: the Euclidean time τ is limited by a period $1/T$, the so called Matsubara time. The instanton solution with periodic boundary conditions, called caloron, is well known. Fermionic *anti*−periodic zero modes are also known. It is important that they have exponential decay $\exp(-\pi r T)$ in the spatial direction, but just oscillate in τ. So if instantons are like atoms with the quark zero mode as a wave function, the finite T compresses their special extension and enhances the temporal one. (It looks like "pencil-like" atoms in a very strong magnetic field.) That radically changes their interactions, which are only strong if instantons are interacting along the time direction. In particular, a pair of such type can be formed, connected to themselves by periodicity. So the chiral phase transition is actually driven by a *rearrangement* of the ensemble into a gas of instanton anti-instanton "molecules"[g]. In a series of IILM numerical

[g] Note a similarity to the Kosterlitz-Thouless transition in the $O(2)$ spin model in two

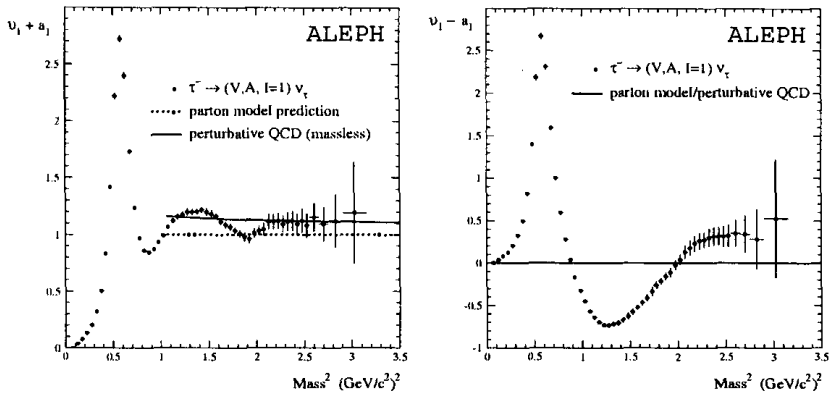


Figure 5: Spectral functions $v(s) \pm a(s) = 4\pi^2(\rho_V(s) + \rho_A(s))$ extracted by the ALEPH collaboration.

simulations[25] it was found that this transition indeed goes as expected, with molecules driving the transition. Furthermore, many thermodynamic parameters, the spectra of the Dirac operator, the evolution of the quark condensate and susceptibilities were calculated[25], with results surprisingly consistent with available lattice data.

4 Two-point correlation functions: τ decays

The available information on hadronic correlation functions, both from experimental data, the OPE and other exact results was reviewed in [11]. Since then, the high statistics measurement of hadronic τ decays $\tau \to \nu_\tau +$ hadrons by the ALEPH experiment at CERN [26] has significantly improved the experimental situation in the vector and axial-vector channel. The purpose of this section is to compare these results with theoretical predictions, both from the OPE and instanton models. In particular, we would like to assess the range of applicability of the two approaches and put improved constraints on the parameters that enter. Translating the spectral functions measured by the ALEPH collaboration into Euclidean coordinate space correlation functions will also allow precise comparison of the experimental data with improved lattice calculations.

Consider the vector and axial-vector correlation functions $\Pi_V(x) = \langle j^a_\mu(x) j^a_\mu(0)\rangle$ and $\Pi_A(x) = \langle j^{5\,a}_\mu(x) j^{5\,a}_\mu(0)\rangle$. Here, $j^a_\mu(x) = \bar{q}\gamma_\mu \frac{\tau^a}{2} q$, $j^{5\,a}_\mu(x) = \bar{q}\gamma_\mu\gamma_5 \frac{\tau^a}{2} q$ are the isotriplet vector and axial-vector currents. The correlation functions have

dimensions: again one has paired topological objects, vortices in one phase and random liquid in another. The high and low-temperature phase change places, though.

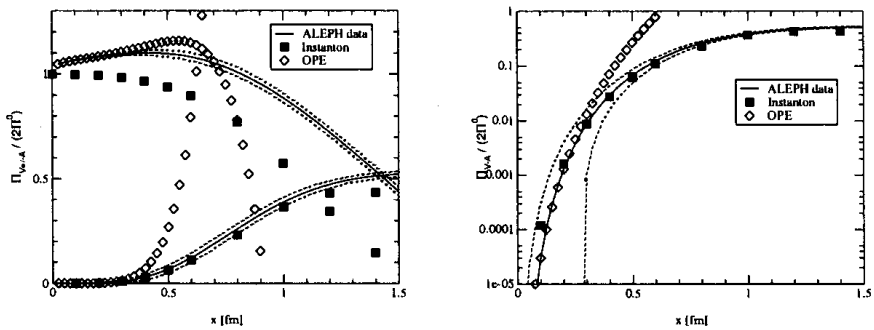


Figure 6: (a) Euclidean coordinate space correlation functions $\Pi_V(x) \pm \Pi_A(x)$ normalized to free field behavior. The solid lines show the correlation functions reconstructed from the ALEPH spectral functions and the dotted lines are the corresponding error band. The squares show the result of a random instanton liquid model and the diamonds the OPE fit described in the text. (b) The same for V-A, but in logarithmic plot.

the spectral representation[11]

$$\Pi_{V,A}(x) = \int ds\, \rho_{V,A}(s) D(\sqrt{s}, x), \tag{13}$$

where $D(m,x) = m/(4\pi^2 x) K_1(mx)$ is the Euclidean coordinate space propagator of a scalar particle with mass m. We shall focus on the linear combinations $\Pi_V + \Pi_A$ and $\Pi_V - \Pi_A$. These combinations allows for a clearer separation of different non-perturbative effects. The corresponding spectral functions $\rho_V \pm \rho_A$ measured by the ALEPH collaboration are shown in Fig. 5.

In Fig. 6 we compare the measured correlation functions with predictions from the instanton model. We begin our analysis with the combination $\Pi_V - \Pi_A$. This combination is produced by chiral symmetry breaking, while all the perturbative diagrams, as well as gluonic operators cancel out. The agreement of the instanton prediction with the measured $V - A$ correlation is impressive and extends all the way from short to large distances. At distances $x > 1.25$ fm both combinations are dominated by the pion contribution while at intermediate x the ρ, ρ' and a_1 resonances contribute.

We shall now focus our attention on the $V + A$ correlation function. The unique feature of this function is the fact that the correlator remains close to free field behavior for distances as large as 1 fm. This phenomenon was referred

to as "super-duality" in [11]. The instanton model reproduces this feature of the $V + A$ correlator. We also notice that for small x the deviation of the correlator in the instanton model from free field behavior is small compared to the perturbative $O(\alpha_2)$ correction. This opens the possibility of precision studies of the pQCD contribution.

5 Three-point correlators: the pion form-factor

In principle, 3-point correlation functions are used to measure the expectation values of some operators over a particular current, say magnetic moment of a nucleon. One of the operators is then a "probe" and two others a "source" and a "sink". As an example of important physical quantity with a long-standing controversy we will discuss the electromagnetic form-factor of a pion.

A well-defined pQCD asymptotics of $F_\pi(Q)$ is known at large Q , and a comparison of the asymptotic behavior with the data is crucial for understanding of the momentum scale at which the perturbative regime of QCD is reached. Recently, the charged pion form-factor has been measured very accurately at momentum transfers $0.6\ GeV^2 < Q^2 < 1.6\ GeV^2$ by the Jefferson Laboratory (JLAB) F_π collaboration [13] and lead to quite surprising results. Not only are the data at highest experimentally accessible momenta still very far from the asymptotic limit (the shaded area at the right), but the trend away from the pQCD prediction is still continuing (see Fig. 7).

In the single-instanton approximation the problem was first addressed in [16]. Unlike the previous works (done in the QCD sum rule framework), these authors have started with a 3-point correlator including electromagnetic and two *pseudoscalar* (rather than axial) currents. In such correlator (as well as other scalar and pseudoscalar channels) the instanton contribution with maximal number of the zero-mode terms in the quark propagators is possible, resulting in enhanced (relative to e.g. the vector or axial channels) contribution, by a factor $1/(m^\star\rho)^2$, where $m^\star$ denotes the effective quark mass defined above. Numerically, the enhancement factor is about 30, and parametrically it is the inverse diluteness of the instanton ensemble.

This feature, however, is not generically related to the pion itself and depends on the particular three-point function under investigation. For example, the enhancement is absent, when one considers the pion contribution to the axial correlator. Similarly, there is no such enhancement of the $\gamma\gamma^\star\pi^0$ neutral pion transition form-factor. The relevant instanton effects for this process are not due to (enhanced) zero modes, but are either related to non-zero mode propagators in the instanton background or to multi-instanton effects, which are suppressed by the instanton diluteness. This conclusion is nicely supported

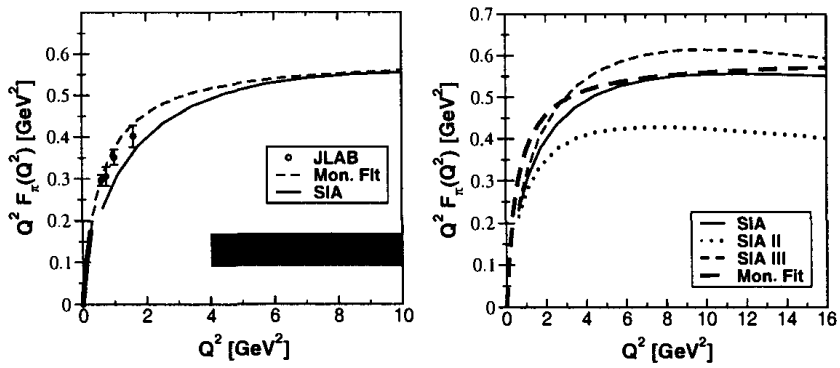


Figure 7: (a) The recent JLAB data for $Q^2\ F_\pi(Q^2)$ in comparison with the asymptotic pQCD prediction (thick bar, for a typical $\alpha_s \approx 0.2-0.4$, the monopole fit (dashed line), and our SIA calculation (solid line). The SIA calculation is not reliable below $Q^2 \sim 1\ GeV^2$. The solid circles denote the SLAC data. (b)The dependence of the pion form-factor $Q^2 F_\pi(Q^2)$ on the instanton size distribution. The SIA (solid) curve represents a small-size 't Hooft distribution with a Fermi distribution cutoff and lattice QCD parameters. The SIA II (dotted) curve has a different mean instanton size $\bar{\rho} = 0.47\,fm$ (same width) and the SIA III (dashed) curve is obtained with the simplest delta distribution $n(\rho) = \bar{n}\,\delta(\rho - 1/3\,fm)$.

by recent CLEO measurements of this form-factor, which indeed show that the asymptotic pQCD regime is reached much earlier, at $Q^2 \sim 2\ GeV$ [14]. The instanton-induced calculation of such form-factor also has been recently done [15], in good agreement with data. In this case there is no interaction enhancement discussed above, and the effect comes from the momentum-dependent mass of each quark.

Let us select points as follows

$$\Gamma_\mu(x,y) =< j_5^+(-x/2) j_\mu(y) j_5(x/2) > \tag{14}$$

and, for simplicity, think of a charged pion, so that axial currents made of quarks of different flavor and therefore only the "triangular" diagram has to be considered. This diagram for small x and y was calculated in [16], and (after plugging in standard RILM parameters) and relatively complicated numerical analysis (removing the non-pion contribution), the pion form-factor for $Q^2 \sim 1\,GeV$ was obtained, in good agreement with data.

In [17] calculation of the 3-point correlator was extended to multi-instanton background (RILM), which has allowed to take $x, y \sim 1fm$ and therefore to ensure dominance of the pion pole. Furthermore, one can obtain the pion mass and coupling constant from 2-point correlators for the same ensemble and check its consistence with the measured 3-point one. After that, the pion form-

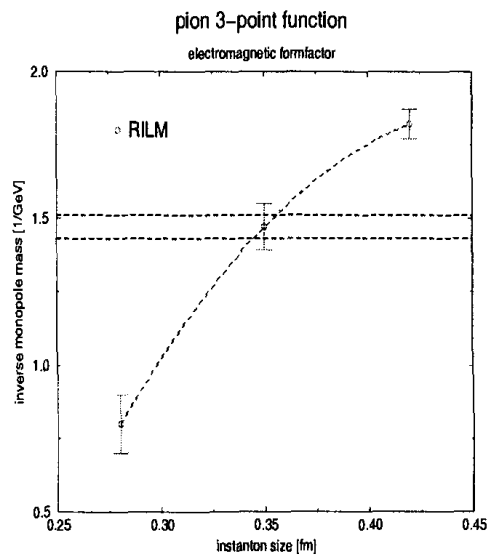


Figure 8: The fitted mass parameter from the pion form-factor versus the instanton size in RILM. The horizontal lines display experimental uncertainty.

factor can be rather accurately determined. Its standard parametrization in a "monopole form" $F_\pi(Q^2) = (\frac{M^2}{Q^2+M^2})$ was found to work well, and dependence of the parameter M on the instanton size was specially studied. The results are shown in Fig.7. One can see that the pion size is directly related with the instanton size, and experimentally observed value is indeed obtained for $\rho \approx .35 fm$. So, in a way, *the pion form-factor is basically the instanton form-factor.*

Another important step has been made in a recent paper [12], which calculated pion form-factor at intermediate momentum transfers, $2\ GeV^2 < Q^2 < 10\ GeV^2$. The central prediction following from analytic SIA calculation is shown in Fig. 7 in comparison to the recent JLAB measurements. We find the very intriguing result that the instanton contribution to the form-factor is completely consistent with the monopole fit at intermediate momentum transfers, even where the vector dominance model has no justification. For larger momenta transfer, $Q^2 > 20\ GeV^2$, the SIA breaks down, as it is necessary to increase the distances in order to isolate the pion ground state. At these needed distances, however, the correlation functions will become sensitive to multi-instanton effects.

We consider the spatial Fourier transforms of the Euclidean three-point

function and two-point function,

$$G_\mu(t, \mathbf{p}+\mathbf{q}; -t, \mathbf{p}) = \int d^3\mathbf{x}\, d^3\mathbf{y}\, e^{-i\,\mathbf{p}\cdot\mathbf{x}+i\,(\mathbf{p}+\mathbf{q})\cdot\mathbf{y}} \times \langle 0|\, j_5(t,\mathbf{y})\, J_\mu(0,\mathbf{0})\, j_5^\dagger(-t,\mathbf{x})\, |0\rangle, \tag{15}$$

$$G(2t, \mathbf{p}) = \int d^3\mathbf{x}\, e^{i\,\mathbf{p}\cdot\mathbf{x}}\, \langle 0|\, j_5(t,\mathbf{x})\, j_5^\dagger(-t,\mathbf{0})\, |0\rangle, \tag{16}$$

where the pseudo-scalar current $j_5(x) = \bar{u}(x)\,\gamma_5\, d(x)$ excites states with the quantum numbers of the pion and $J_\mu(0)$ denotes the electro-magnetic current operator. In the large t limit (at fixed momenta), both correlation functions are dominated by the pion pole contribution and the ratio of the three-point function to the two-point function becomes proportional to the pion form-factor. In the Breit frame, $\mathbf{p} = -\mathbf{q}/2$ and $Q^2 = \mathbf{q}^2$, one has simply

$$\frac{G_4(t, \mathbf{q}/2; -t, -\mathbf{q}/2)}{G(2t, \mathbf{q}/2)} \to F_\pi(Q^2). \tag{17}$$

Notice that the LHS of Eq. (17) should not depend on t, for t large enough: for the pion, this is achieved already for $t \sim 0.6\, fm$.

Since the pion form-factor may experimentally be rather accurately measured, it is instructive to ask whether such data may shed some light on the instanton size distribution. In Fig. 7(b), we have plotted the results of our theoretical predictions for $Q^2\, F_\pi(Q^2)$ obtained for different cases of $n(\rho)$. We contrast the simplest size distribution [6], $n(\rho) \sim \delta(\rho - 1/3\, fm)$, to the results obtained from a lattice QCD parameterization. We notice that the presence of tails in the size distribution introduces only small corrections to the form-factor. Therefore, we conclude that the simplest distribution $n(\rho) \sim \delta(\rho - 1/3\, fm)$ indeed captures the relevant features for the pion form-factor at intermediate momentum transfer. We observe that our result becomes closer to the perturbative limit, if the average instanton size is larger or possibly if there is an asymmetric tail toward larger-sized instantons in the distribution.

Next, in Fig.1(b) we compare the results from distributions which have the same small-size limit, but cut off at different instanton sizes, $\bar{\rho} = 0.37\, fm$ and $0.47\, fm$. We observe that, throughout the entire kinematic region we have considered, these predictions are quite sensitive to the average instanton size but not yet to their size distribution.

6 Instantons at large N_c

Let me remind that in the one-loop order the instanton action is given by $S_0 = (8\pi^2)/g^2 = -b\log(\rho\Lambda)$ where $b = (11N_c)/3$ is the first coefficient of the beta function in pure gauge QCD. In the 't Hooft limit $N_c \to \infty$ with $g^2N_c = const$ we expect $S_0 = O(N_c)$ and $\rho = O(1)$. This leads to Witten's suggestion that instantons do not survive the large-N_c limit which was very important historically. However the argument has loopholes: one should keep in mind that (i) instantons in QCD come in all sizes, and (ii) their interaction is $O(N_c)$ also. We will discuss the effects of both subsequently.

What was shown by T. Schafer [27] (whose work we follow in this section) is that assuming that the semi-classical approximation remains valid, one arrives at a self-consistent picture, which also remarkably well agrees with pre-existing theoretical expectations .

In particular, the density of instantons grows as N_c whereas the typical instanton size remains finite. Furthermore, the size distribution becomes the delta-function like, as was used for the instanton liquid at $N_c = 3$ already in [6]. Interactions between instantons are important and suppress fluctuations of the density and of the topological charge. Using mean field approximation and then numerical IILM simulations one finds that this scenario does not require fine tuning but arises naturally if the instanton ensemble is stabilized by a classical repulsive core. Although the total instanton density is large but the instanton liquid remains effectively dilute because instantons are not strongly overlapping in color space.

6.1 Naive counting and expectations

The total density

Let us first consider existing theoretical prejudice about the total instanton density. Since it is related to the non-perturbative gluon condensate

$$\frac{N}{V} = \frac{1}{32\pi^2}\langle g^2 G^a_{\mu\nu} G^a_{\mu\nu}\rangle \tag{18}$$

the N_c counting suggests that $\langle g^2G^2\rangle = O(N_c)$ and we are led to the conclusion that $(N/V) = O(N_c)$. This is also consistent with the expected scaling of the vacuum energy. Using equ. (18) and the trace anomaly relation

$$\langle T_{\mu\mu}\rangle = -\frac{b}{32\pi^2}\langle g^2 G^a_{\mu\nu} G^a_{\mu\nu}\rangle, \tag{19}$$

the vacuum energy density is given by

$$\epsilon = -\frac{b}{4}\left(\frac{N}{V}\right). \tag{20}$$

Using $(N/V) = O(N_c)$ we find that the vacuum energy scales as $\epsilon = O(N_c^2)$ which agrees with our expectations for a system with N_c^2 gluonic degrees of freedom.

Note that $(N/V) = O(N_c)$ implies that the effective diluteness of instantons remains constant in the large N_c limit. Indeed, in spite of large density most instantons do not see each other: the number of mutually commuting $SU(2)$ subgroups of $SU(N_c)$ scales as N_c.

Density and charge fluctuations

If instantons are distributed randomly then fluctuations in the number of instantons and anti-instantons are expected to be Poissonian. This leads to the predictions

$$\langle N^2\rangle - \langle N\rangle^2 = \langle N\rangle, \tag{21}$$

$$\langle Q^2\rangle = \langle N\rangle, \tag{22}$$

where $N = N_I + N_A$ is the total number of instantons and $Q = N_I - N_A$ is the topological charge. Equ. (22) implies that

$$\chi_{top} = \frac{\langle Q^2\rangle}{V} = \left(\frac{N}{V}\right). \tag{23}$$

Using $(N/V) = O(N_c)$ we observe that $\chi_{top} = O(N_c)$ which is in contradiction to Witten's assumption $\chi_{top} = O(1)$. However, as we shall see in the next section, interactions between instantons cannot be ignored in the large N_c limit and the fluctuations are suppressed.

6.2 Mean field arguments and the Chiral condensate

We now include fermion related dynamics in, and ask how the chiral condensate scales with N_c, using first analytic mean field approximation[h]. Instantons induce an effective $2N_f$-fermion Lagrangian. After averaging over the color

[h] For definiteness, we will consider the case $N_f = 2$ but the conclusions are of course independent of the number of flavors.

orientation of the instanton the effective Lagrangian is given by

$$\mathcal{L} = \int n(\rho) d\rho \, \frac{2(2\pi\rho)^4 \rho^2}{4(N_c^2-1)} \epsilon_{f_1 f_2} \epsilon_{g_1 g_2} \left(\frac{2N_c - 1}{2N_c} (\bar{\psi}_{L,f_1} \psi_{R,g_1})(\bar{\psi}_{L,f_2} \psi_{R,g_2}) \right. \quad (24)$$
$$\left. - \frac{1}{8N_c} (\bar{\psi}_{L,f_1} \sigma_{\mu\nu} \psi_{R,g_1})(\bar{\psi}_{L,f_2} \sigma_{\mu\nu} \psi_{R,g_2}) + (L \leftrightarrow R) \right).$$

We observe that the explicit N_c dependence is given by $1/N_c^2$. This is again related to the fact that instantons are $SU(2)$ objects. Quarks can only interact via instanton zero modes if they overlap with the color wave function of the instanton. As a result, the probability that two quarks with arbitrary color propagating in the background field of an instanton interact is $O(1/N_c^2)$.

Chiral symmetry breaking can be studied in the mean field approximation: one can use a gap equation for the spontaneously generated constituent quark mass

$$M = GN_c \int \frac{d^4k}{(2\pi)^4} \frac{M}{M^2 + k^2}, \quad (25)$$

where M is the constituent mass and G is the effective coupling constant in equ. (24). The factor N_c comes from doing the trace over the quark propagator. The coupling constant G scales as $1/N_c$ because the density of instantons is $O(N_c)$ and the effective Lagrangian contains an explicit factor $1/N_c^2$. We conclude that the coefficient in the gap equation is $O(1)$ and that the dynamically generated quark mass is $O(1)$ also. This also implies that the quark condensate, which involves an extra sum over color, is $O(N_c)$.

The results in the mean field approximation and simulations show that for $N_c > 4$ the average instanton size is essentially constant while the instanton density grows linearly with N_c. Expanding $\log(N/V)$ in powers of N_c and $\log(N_c)$ we observe that independent of the details of the interaction the instanton density scales at most as a power, not an exponential, in N_c.

Size distribution

One generic argument is that the instanton size distribution $dN/d\rho \sim exp(-N_c * F(\rho))$ goes to zero where $F(\rho) > 0$ and makes no sense where $F(\rho) < 0$: so it must settle down to a delta function at the "fixed size" given by $F(\rho^*) = 0$. This is indeed what happens.

One can further see why the instanton density scales as the number of colors is as follows: the size distribution is regularized by the interaction between instantons. This means that there has to be a balance between the average

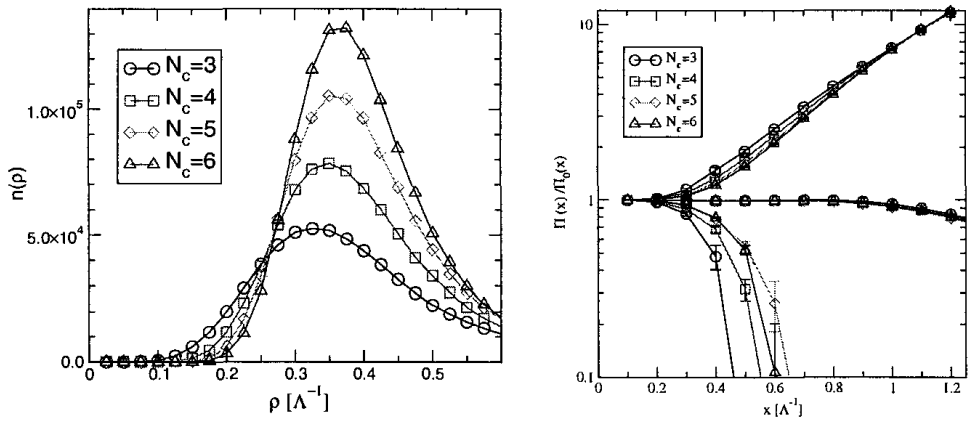


Figure 9: (a) Instanton size distribution in a pure gauge instanton ensemble for different numbers of colors. The results were obtained using numerical simulations with $N = 128$ instantons. (b) Meson correlation function in a pure gauge instanton ensemble for different numbers of colors. We show the correlation function of the pion, the rho meson, and the η' meson normalized to the corresponding free correlation functions.

single instanton action and the average interaction between instantons. If the average instanton action satisfies $S_0 = O(N_c)$ we expect that $\langle S_{int}^{tot} \rangle = O(N_c)$ also. Using $\langle S_{int}^{tot} \rangle = (N/V)\langle S_{int} \rangle$ and the fact that the average interaction between any two instantons satisfies $\langle S_{int} \rangle = O(1)$ we expect that the density grows as N_c.

Fig. 9(a) shows the instanton size distribution for different numbers of colors. We observe that the number of small instantons is strongly suppressed as $N_c \to \infty$ but the average size stabilizes at a finite value $\bar{\rho} < \Lambda^{-1}$. We also note that there is critical size ρ^* for which the number of instantons does not change as $N_c \to \infty$. The value of ρ^* is easy to determine analytically. We write $n(\rho) = \exp(N_c F(\rho))$ with $F(\rho) = a\log(\rho) + b\rho^2 + c$ where the coefficients a, b, c are independent of N_c in the large N_c limit. The critical value of ρ is given by the zero of $F(\rho)$. We find $\rho^* = 0.49\Lambda^{-1}$. The existence of a critical instanton size for which $n(\rho^*)$ is independent of N_c was suggested long ago. The problem was studied on the lattice by Lucini and Teper[28], who find distributions which are very similar to Fig.9(a) and have the fixed point size $\rho^* = 6a = 0.43$ fm.

6.3 *Fluctuations in the interacting instanton liquid*

Fluctuations in the net instanton number are related to the second derivative of the free energy with respect to N. We find

$$\langle N^2\rangle - \langle N\rangle^2 = \frac{4}{b}\langle N\rangle. \tag{26}$$

This result is in agreement with a low energy theorem based on broken scale invariance

$$\frac{1}{(32\pi^2)^2}\int d^4x\ \{\langle g^2G^2(0)g^2G^2(x)\rangle - \langle g^2G^2(0)\rangle^2\} = \frac{4}{b}\frac{1}{32\pi^2}\langle g^2G^2\rangle. \tag{27}$$

For a dilute system of instantons equ. (27) reduces to equ. (26). The result (26) shows that fluctuations of the instanton ensemble are suppressed by $1/N_c$. This is agreement with general arguments showing that fluctuations are suppressed in the large N_c limit.

Interaction is important: If r denotes the ratio of the average interaction between instantons of opposite charge and instanton of the same charge, $r = \langle S_{IA}\rangle/\langle S_{II}\rangle$, then in the mean field analysis

$$\langle Q^2\rangle = \frac{4}{b - r(b-4)}\langle N\rangle. \tag{28}$$

This result shows that for any value of $r \neq 1$ fluctuations in the topological charge are suppressed as $N_c \to \infty$. The same conclusions follow from numerical simulations of IILM. We specially note that we got $\chi_{top} = O(1)$, in agreement with Witten's hypothesis. Eventually, at $N_c > 20$ or so, η' does become much lighter and resembles a partner of pions, as Witten predicted. Interestingly, σ on the contrary is becoming heavier in this limit.

7 The RHIC puzzles

7.1 *Two main puzzles*

The very first RHIC data on particle spectra had confirmed that we observe a *Bang*, explosive behavior, not a *fizzle* with little re-scattering predicted by extrapolation of pQCD to $p \sim 1 - 2\,GeV$. Both radial and elliptic collective flows are very well reproduced by ideal hydrodynamics [30], with the equation of state close to that derived from the lattice. Somewhat surprisingly it also describes well the tails of particle spectra at $p_t \sim 2\ GeV$, including the unexpected crossing of proton and pion spectra in this region. It also works for rather peripheral collisions (although of course extremely peripheral ones

agree with pp). This implies that QGP is produced promptly enough, at time $\sim 1/2\, fm/c$. How that happens is *the first RHIC puzzle*, that of very short mean free path.

The second RHIC puzzle is that of very strong *jet quenching*, combined with *large azimuthal asymmetry*. The expected (and measured in pp or peripheral AA) spectrum of hard jets from parton processes is suppressed in central AA collisions by a significant factor. The measured values of azimuthal asymmetry v2 at large $p_t > 2\ GeV$ exceed even the strong quenching limit worked out in [31]. Events triggered on high-p_t hadrons have seen nice forward jet fragments, but not the backward ones, directly confirming strong jet quenching.

7.2 Theory Overview

Why is equilibration so rapid, the viscous corrections to ideal hydro are so small, and nuclei are so black even for high-p_t jets? Trying to answer those questions, we will compare predictions of pQCD (the *weak* coupling) with recently available results for $\mathcal{N} = \triangle$ SYSY YM in *strong* coupling limit.

Naturally, pQCD approach goes back to the 1970's. Re-scattering of gluons with momenta about 1-2 GeV (later called "mini-jets") has been considered already in the 1970's. Recent development[32] included account for the Landau-Pomeranchuck-Migdal (LPM) effect and other refinements: although these authors have found that perturbative equilibration is still possible, its rate is suppressed by powers of small coupling constant α_s.

Now we definitely know that extrapolation of pQCD to momenta $Q \sim 1\ GeV$ *fails miserably*[i]. Probably it has to be so, because both at $Q \gg 1\, GeV$ (pQCD) and $Q \ll 1\ GeV$ (pion Lagrangians) strong interaction is weak, and we also know it must be strong somewhere, to make all this non-perturbative phenomena like confinement and chiral symmetry breaking.

The most direct way to compare weak and strong coupling regimes is in the hydrodynamic context, where it appears as viscous corrections. pQCD predictions are well summarized by P. Arnold et al[34]. The high temperature shear viscosity in a gauge theory with a simple gauge group (either Abelian or non-Abelian) has the leading-log form

$$\eta = \kappa \frac{T^3}{g^4 \ln g^{-1}}\,, \tag{29}$$

where g is the gauge coupling. For the case of $SU(3)$ gauge theory the leading-log shear viscosity coefficient $\kappa \sim 100$, specific values for various numbers of

[i] In [33] it is nicely quantified by how much pQCD-based scenario with 1-2 GeV cutoff misses what is needed: in order to reproduce elliptic flow by a parton cascade the product of the *gluon density times the cross section* should be increased by a factor of about 80.

light $m << T$ fermion species are shown in the table of that paper.

These results are to be contrasted to a new calculation developed via the famous AdS/CFT correspondence for $\mathcal{N} = \Delta$ supersymmetric Yang-Mills theory. In it the coupling does not run and one can study the *strong coupling limit* of the theory. Let me mention only one paper of the kind [35], focused on shear viscosity at high T. Their main result is the following shear viscosity coefficient

$$\eta = \frac{\pi}{8} N_c^2 T^3 \tag{30}$$

Let me also mention that in this theory the strong coupling limit of thermodynamical quantities is 3/4 of its ideal gas (Stephan-Boltzmann) values: those are not far from lattice results. So, it seem now likely that QGP produced at RHIC is, unlike the asymptotic one at high T, is actually close to the strong coupling regime. If so, the hydrodynamics becomes justified, but *not* the pQCD parton cascades.

Although HERA data on deep inelastic scattering can still be described by the DGLAP evolution, it is not the only possible explanation: the *saturation-based* models have created very good fits to data and vast theoretical literature. It has been argued [36] that random classical glue, or Color Glass Condensate (CGC) can be a better description at sufficiently small x.

In this language the point we would like to make now is that a significant part of this classical glue should be non-random but highly structured, due to intrinsic topology related to tunneling. A big difference between these kinds of classical glue appears for quark pair production: the topological clusters creates[j] $\bar{u}u\bar{d}d\bar{s}s$ [38]. It helps to solve the rapid entropy production puzzle.

8 Quantum mechanics of the glue, in vacuum and high energy collisions

8.1 The main idea

As we already discussed above, a promptly excited glue is not just gluons: it should also appear in form of specific gluomagnetic *topological clusters*. The idea itself originates from two major sources: studies of the non-perturbative tunneling phenomena in the QCD vacuum [1], and insights obtained a decade ago in the discussion of high energy collisions in electroweak theory [37]. Recent development [2,40,38] has been very rapid: obviously full presentation of these results at technical level is done elsewhere and here I only try to present an

[j] In a way, those are simply materialized quarks from the vacuum chiral condensate.

emerging physical picture, more details (and applications to event-by-event fluctuations and J/ψ suppression) can be found in my preprint[41].

In very general terms, in high energy collisions *sudden localization of quantum fields* makes virtual fields (partons) real. In QED we have Weizsacker-Williams approximation, suggesting that boosted virtual Coulomb field may easily become real photons. In pQCD we have the so called Lipatov vertex, describing how two virtual gluons make the real (on-shell) one: it is the basis of celebrated BFKL re-summation. Sudden excitation of the part of vacuum wave function *under the barrier* also produces certain real objects[k]: those are states *on* the barrier. Surprisingly, only recently they have been studied in QCD[2].

8.2 Turning states as a conditional minimum

The shape of the potential and the corresponding turning states can be obtained[2] from the *minimization of the potential energy of a static Yang-Mills fields*, consistent with two appropriate constraints: *(i) fixed value of (corrected) Chern-Simons number N_{CS} (1). (ii) fixed value of the r.m.s. size R* . To find those one should search for the minimum of the following functional

$$E_{eff} = \frac{1}{2}\int B_m^2 d^3x + R(A_\mu)/\rho^2 + \kappa N_{CS}(A_\mu) \tag{31}$$

where $1/\rho^2, \kappa$ are two Lagrange multipliers. Although these two terms append the YM equations and make them more complicated, an *analytical solution is found.* Skipping the details, let me only say that the energy and (corrected) Chern-Simons number are

$$E_{stat} = 3\pi^2(1-\kappa^2)^2/(g^2\rho), \;\; \tilde{N}_{CS} = \mathrm{sign}(\kappa)(1-|\kappa|)^2(2+|\kappa|)/4 \tag{32}$$

Eliminating κ, we find the shape of the potential plotted above. The maximum – the YM sphaleron – corresponds to $\kappa = 0$, and its energy is $3\pi^2/(g^2\rho)$. If the size is determined by the mean radii of vacuum instantons $\rho = 1/3$ fm[6], the sphaleron mass[l] is about 3 GeV. One can view it as a magnetic ball with field lines for gluons with 1-3 colors, rotating around x,y and z axes, respectively, while all gluon fields with colors 4-8 are absent.

[k] They are real in the same sense as gluons: namely at times shorter than the one at which confinement sets in.

[l] In electroweak theory this mass is about 10 TeV, so it is made of about 100 gauge bosons.

8.3 *Explosive behavior of the turning states*

Solution of the classical YM equations describes what happens to these turning states after they are produced[m] In QCD there are no Higgs scalar and its non-zero VEVs, so gluons are massless. This makes the process even more *explosive* because all harmonics with different momenta move together, with the speed of light. In [2] the problem was solved both *numerically* and *analytically*. Omitting all details, let us go directly to large times, when the promised spherical shell has the following energy density profile

$$4\pi r^2 e(r,t) = \frac{8\pi}{g^2\rho^2}(1-\kappa^2)^2 \left(\frac{\rho^2}{\rho^2+(r-t)^2}\right)^3 \tag{33}$$

Of course, at large times the fields becomes weak and can be decomposed into gluons: the Fourier transform of the fields provides the energy distribution of the resulting gluons. One gets about 3 gluons out of it, if the size is 1/3 fm. Furthermore, as shown by explicit solution of the Dirac equation [38], a pair of each light quark flavor is taken from vacuum and accelerated to the physical energies.

8.4 *Are there topological clusters in pp collisions?*

A number of authors [20] have suggested that the growth of pp cross section, $\sigma \sim s^{0.08}$ (soft pomeron) is related with tunneling and that 0.08 is proportional to cross section of the production of topological clusters. This power, as well as pomeron size have been calculated: an interesting byproduct found was the *absence of odderon* due to SU(2) nature of fields, in agreement with data but in disagreement with most other models.

Why have these clusters not been already seen and identified long ago? When a topological cluster is produced in hadronic collisions, as a colored object it is still connected by strings to receding partons, and is easily lost in debris of fragmenting strings. It can in principle be located from correlation measurements, but it is difficult to do.

One may think about the so called double-diffractive (or Pomeron-Pomeron) processes in which colorless clusters are produced. Let me mention one recent paper including such analysis of old data [46] in which one sees that such collisions indeed result in clusters of few GeV mass. Also intriguing is the fact that clusters with mass up to 5 GeV decay isotropically in their rest frame.

[m] A similar study has been made a decade ago in electroweak theory [43] for the sphaleron, where it has been found that it decays in about 51 W,Z,H: half of the energy thus goes to acceleration.

Unfortunately, the UA2 detector used was just a simple calorimeter, and we do not know anything about the structure of these clusters. RHIC detectors and especially STAR can do a lot of clarification in pp mode. There are first indications that a scalar glueball production is seen, in $K+K-$ mode.

When a cluster is produced into a vacuum, its expansion is affected by confining forces and also other instantons (chiral condensates). This is above the mean energy of the spectrum just discussed: it means that it is impossible to use classical YM description in this case, unless the outgoing quarks and gluons are projected into particular hadronic final states[n].

Finally, let me mention the paper [45] in which parton-model-style phenomenology of various hadronic collisions is developed. The main idea was to identify two components of the hh collisions, the color exchanges and the "color objects production", and deduce the corresponding cross sections at the partonic level. We looked at high energy NN, πN, γN, and $\gamma\gamma$ cross sections which all *increase* with energy logarithmically for $\sqrt{s} \sim 100\ GeV$

$$\sigma_{hh'}(s) = \sigma_{hh'}(s_0) + X_{hh'} \ln(s/s_0) \tag{34}$$

We identified the two components mentioned above with these two terms, respectively, and concentrated on the last (growing) terms. We studied whether some universal *semi-hard* parton-parton collisions can explain all known $X_{hh'}$. Using fitted structure functions of N, π, γ and simple scaling – each gluon can be counted as 2 quarks[o] – we have expressed all of those with only one parameter, the value of the *qq cross section.* With the fitted value[p] $\sigma_{qq} = 1.69 \times 10^{-3} fm^2$ we got correct rising part of cross sections for 4 hadronic reactions. In [45] we have also looked at the *shadowing* corrections, of the second (growing) component by the first.

8.5 Toward the semiclassical description of high energy collisions

To find the semiclassical description for this process was known in 1990's as the so called *holy grail* problem. Three methods toward its solution have been proposed

(i) *Unitarization* of the multi-gluon amplitude when it becomes strong was first suggested by Zakharov and worked out by Shifman and Maggiore [37]. Basically one can treat a sphaleron as a resonance.

[n] The observed decrease of HBT correlation strength with multiplicity in pp collisions can be related with large fraction of η', η, K_s long-lived sources of pions, associated with instanton-induced production.

[o] Corresponding to SU(2) Casimir scaling, appropriate for instanton-induced reactions.

[p] Note that simple parametric estimate for this cross section, namely $\pi\rho^2 n_{inst}\rho^4$ gives the right magnitude.

(ii) *Landau method with singular instantons* was applied by Diakonov and Petrov [37] (following some earlier works which are cited there) who were able to find the opposite limit of high energies. It follows from the comparison of the two limits, that the peak is indeed very close to the sphaleron mass, and the cross section is very close to be *first order* in instanton diluteness.

(iii) *Classical solution on the complex time plane* [47] is another possible direction, in which a zig-zag shaped path in complex time includes classical evolution and tunneling in one common solution.

There is progress along all those lines, although we are still far from a complete semiclassical theory of high energy collisions.

9 Acknowledgments

I am grateful to the organizers, Profs.G' t Hooft and A. Zichichi, for the invitation to this school, for their deep interest in the content and quality of all lectures, and, in particular, for their critical and helpful remarks during my lectures and question periods. As before, I am very impressed by the high scientific level of the school and the unbeatable level of hospitality in Erice.

1. T. Schafer and E. V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998) [hep-ph/9610451].
2. D. M. Ostrovsky, G. W. Carter and E. V. Shuryak, hep-ph/0204224.
3. A. A. Belavin, A. M. Polyakov, A. A. Schwartz, and Yu. S. Tyupkin, *Phys. Lett.*, 59B:85, 1975.
4. G. 't Hooft, *Phys. Rev.*, D14:3432, 1976.
5. Y. Nambu and G. Jona-Lasinio, *Phys. Rev.*, 122:345, 1961.
6. E. V. Shuryak, *Nucl. Phys.*,, B198:83, 1982.
7. M. C. Chu, J. M. Grandy, S. Huang, and J. W. Negele, *Phys. Rev.*, D 48:3340, 1993.
8. P. Faccioli, A. Schwenk and E. V. Shuryak, Phys. Lett. B **549**, 93 (2002) [arXiv:hep-ph/0205307].
9. D. Dolgov *et al.* [LHPC collaboration], Phys. Rev. D **66**, 034506 (2002) [arXiv:hep-lat/0201021].
10. N. I. Kochelev, Phys. Rev. D **57**, 5539 (1998) [arXiv:hep-ph/9711226].
11. E. V. Shuryak, *Rev. Mod. Phys.*, 65:1, 1993.
12. P. Faccioli, A. Schwenk and E. V. Shuryak, arXiv:hep-ph/0202027.
13. J. Volmer *et al.* [The Jefferson Laboratory F_π Collaboration], Phys. Rev. Lett. **86** (2001) 1713.
14. J. Gronberg *et al.*, Phys. Rev. **D57** (1998) 33.
15. A.E. Dorokhov, hep-ph/0212156
16. H. Forkel and M. Nielsen, Phys. Lett. **B345** (1997) 55.

17. A. Blotz and E.V. Shuryak, Phys. Rev. **D**55 (1997) 4055.
18. C. Gattringer, Phys.Rev.Lett.88:221601,2002; hep-lat/0202002; T. Blum et al, Phys.Rev.D65:014504,2002; hep-lat/0105006; T. DeGrand, A. Hasenfratz, Phys.Rev.D64:034512,2001; hep-lat/0012021
19. I. Horvath *et al.*, Phys. Rev. D **66**, 034501 (2002) [arXiv:hep-lat/0201008].
20. D. E. Kharzeev, Y. V. Kovchegov and E. Levin, Nucl. Phys. A **690**, 621 (2001) [hep-ph/0007182]. M. A. Nowak, E. V. Shuryak and I. Zahed, Phys. Rev. D **64**, 034008 (2001) [hep-ph/0012232].
21. E. Shuryak, Sov.Phys. JETP 74 (1978) 408, Phys.Rep.61 (1980) 71.
22. R. Rapp, T. Schaefer, E.V. Shuryak, M. Velkovsky, Phys. Rev. Lett. **81** (1998) 53.), hep-ph/9711396.
23. M. Alford, K. Rajagopal, F. Wilczek, Phys. Lett. **B422** 247 (1998), hep-ph/9708255.
24. D. T. Son, Phys. Rev. D **59**, 094019 (1999) [arXiv:hep-ph/9812287].
25. T. Schäfer and E. V. Shuryak, *Phys. Rev*, D53:6522, 1996.
26. R. Barate *et al.* [ALEPH Collaboration], Z. Phys. **C76**, 15 (1997). Eur. Phys. J. **C4**, 409 (1998).
27. T. Schafer, Phys. Rev. D **66**, 076009 (2002) [arXiv:hep-ph/0206062].
28. B. Lucini and M. Teper, JHEP **0106**, 050 (2001) [hep-lat/0103027].
29. D. Diakonov and V. Y. Petrov, Sov. Phys. JETP **62** 204 (1985).
30. D. Teaney, J. Lauret and E. V. Shuryak, Phys. Rev. Lett. **86**,ace-str 4783 (2001) [arXiv:nucl-th/0011058], nucl-th/0110037. P. F. Kolb, P. Huovinen, U. W. Heinz and H. Heiselberg, Phys. Lett. B **500**, 232 (2001) [hep-ph/0012137]. P. Huovinen, P. F. Kolb, U. W. Heinz, P. V. Ruuskanen and S. A. Voloshin, Phys. Lett. B **503**, 58 (2001) [hep-ph/0101136].
31. E. V. Shuryak, nucl-th/0112042, Phys.Rev.C.in press.
32. R. Baier, A. H. Mueller, D. Schiff and D. T. Son, Phys. Lett. B **502**, 51 (2001) [hep-ph/0009237]. hep-ph/0204211.
33. D. Molnar and M. Gyulassy, Nucl. Phys. A **697**, 495 (2002) [nucl-th/0104073].
34. P. Arnold, G.D. Moore, L.G. Yaffe JHEP 0011 (2000) 001
35. G. Policastro, D. T. Son and A. O. Starinets, Phys. Rev. Lett. **87**, 081601 (2001) [hep-th/0104066].
36. L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 2233 (1994) [hep-ph/9309289].
37. A. Ringwald, Nucl.Phys. B330 (1990) 1, O. Espinosa, Nucl.Phys. B343 (1990) 310; V.V. Khoze, A. Ringwald, Phys.Lett. B259:106-112, 1991 V.I.Zakharov, Nucl.Phys. B353 (1991) 683: M. Maggiore and M. Shifman, Phys.Rev.D46:3550-3564,1992 D. Diakonov and V. Petrov, Phys.

Rev. D **50**, 266 (1994) [hep-ph/9307356].
38. E. Shuryak and I. Zahed, "Prompt quark production by exploding sphalerons," hep-ph/0206022.
39. A. Krasnitz, Y. Nara and R. Venugopalan, Phys. Rev. Lett. **87**, 192302 (2001) [hep-ph/0108092].
40. R. A. Janik, E. Shuryak and I. Zahed, "Prompt multi-gluon production in high energy collisions from singular Yang-Mills solutions," hep-ph/0206005.
41. E. Shuryak, hep-ph/0205031.
42. N. Manton, Phys.Rev.D 28 (1983) 2019; F.R. Klinkhamer and N. Manton, Phys.Rev.D30 (1984) 2212.
43. J. Zadrozny, Phys. Lett. B **284**, 88 (1992). M. Hellmund and J. Kripfganz, Nucl. Phys. B **373**, 749 (1992).
44. P. Levai and U. W. Heinz, Phys. Rev. C **57**, 1879 (1998) [hep-ph/9710463].
45. G. W. Carter, D. M. Ostrovsky and E. V. Shuryak, Phys. Rev. D **65**, 074034 (2002) [hep-ph/0112036].
46. A. Brandt et al (UA8), hep-ex/0205037.
47. V. A. Rubakov and D. T. Son, Nucl. Phys. B **424**, 55 (1994) [hep-ph/9401257].

CHAIRMAN: E. SHURYAK

Scientific Secretaries: A. Maas, G. Marandella

DISCUSSION I

- *Tschirhart:*

For the (g-2) theoretical calculation, how well are the uncertainties controlled in the dispersion relation that relates it to $e^+e^- \rightarrow$ Hadrons?

- *Shuryak:*

The contribution to this particular problem, the muon g-2, is really that the integral which enters is well convergent at small masses. Here I show the plot of the schematic resonances. It has some shape and it goes to the left and it ends at two times the pion mass when there are no lightest states in the vector channel. That particular integral completely concentrates in the left corner. And the models I speak of are too crude to be used for really quantitative estimates needed for the muon g-2.

In another language, if I continue this exponential tail of the correlation function proportional to the exponential of minus the ρ mass to large distances, eventually the ρ meson-signal dies and the two times pion mass signal proportional to the exponential of minus two times the pion mass will remain up to the large x.

Restated in the x-language: What I should know in order to do something meaningful for the g-2 experiment is to know this function at much larger x than I actually can obtain from my numerical calculations. So I cannot help g-2. Experimentalists can measure this quantity better than we can calculate it. Unfortunately, this is a good example: more than 20 years ago, at the beginning of the e+e- physics, people had machines with a total energy of the order of 200-300 MeV and they could very accurately measure this piece of the cross-section. Now there is only NovosibirskVEPP2 left of such low energy machines and it is too late, since nobody is going to make a new machine for this measurement. I doubt that the Frascati φ factory can go that low.

So any measurements should be done well in their time, since you never know when you will need them later.

- *Levy:*

Why are instantons needed? Can one consider the success of the instanton calculations to explain the V-A measurements of ALEPH as a "proof" of their existence?

- *Shuryak:*

This is a complicated proof. It is not only a proof that you need instantons but it is a proof that they are in a particular ensemble and of particular sizes.

They do explain very well the violation of chiral symmetry in general: this correlator V-A is a violation of chiral symmetry also. Historically the question "Why we need instantons" started from the mass of the η' or the $U(1)_A$ problem. There were no diagrams which can kill the degeneracy between the pion and the η' unless you have the 't Hooft interaction. So when 't Hooft derived this effective interaction, it was really a new and welcome vertex you cannot get from any gluonic diagrams. That is very important. So qualitatively we knew we needed instantons. Moreover, because the mass-difference between η' and the pion is very large, it was clear from the onset in the seventies that this effect has to be large.

Now it explains ALEPH's data and I repeat it explains plenty of other things. I will not write down a table which can be found in literature. The model explains many masses like the nucleon mass and hadronic coupling constants.

The first application was the splitting between the pion and the η' and I repeat that this interaction simultaneously pushes η' up and the pion down. It just enters correlation functions at small distances very symmetrically, so I understand the fact that η' is heavy is not separated from the fact that the pion is light. And the fact that the pion is light cannot be separated from chiral symmetry breaking. So if you get a massless pion, it means that we do create a quark-antiquark condensate and all these things. Then constituent quarks and hadronic masses also follow.

This model explains many things and I present one other recent example, the pion form-factor: If I am saying instantons are created by diagrams with iterated instanton vertices, they produce the pion. I should then be able to calculate the pion form factor times q^2. There is a lot of agreement between models at the moment in the region of small q^2, where good measurements exist. In the asymptotic region of large q^2, perturbative QCD leads to the form factor of the pion as a function of q^2 being proportional to alpha_s over q^2. Recent data from Jefferson-Lab in Virginia added new points, which continue to move against the perturbative QCD asymptotic, but will eventually go down to it. So the question is where it happens. Recently my student Pietro Faccioli has calculated what instanton model predicts, and found that up to very large q^2 of 10 GeV^2, this form factor remains very different from perturbative QCD.

This should not happen to other normal hadrons like ρ which exist due to confinement, not instantons. There is no 't Hooft interaction at the centre of the ρ. However, there is a strong attraction at the very centre of the pion because of small-sized instantons. This is one example for a prediction which is different from any others. It was obtained in the following manner: The form factor is roughly speaking a ratio of a three-point function to a two-point function. And Pietro was clever enough to find a kinematical region in momentum space in an analytical calculation, where q is not too small and not too large and in which the one instanton diagram dominates the three-point function and the two-point function. And so if you are in this region,

the density of the instantons that enters in this vertex drops and many other quantities also drop – in other words in this region the form factor of the pion is just determined by the mean instanton size. And so this is the region where he was able to calculate the result and it was in agreement with the data.

It would be interesting to see if different distributions of instanton sizes can be excluded, since you then get different predictions. In principle I told you that we have good understanding of what happens with instantons at small size, but not at large size. So a spread of predictions is possible.

In general we do not invent instantons on phenomenological grounds. They are a path in Hilbert space from one minimum to another. They exist because the tunnelling exists. It is not invented to fit anything. It is just a feature of Yang-Mills theory.

- *Markov:*

The instantons do not allow particle interpretation. Does it make some sense to speak about radii of instantons?

- *Shuryak:*

It is a four-dimensional soliton. Just one of the coordinates is the euclidean time. Suppose you do not want to do euclidean theory, what then would be an instanton? That question is very often asked. I will address it partly in my second lecture.

What do we do in ordinary quantum mechanics if we have a potential which has some type of a barrier? In ordinary quantum mechanics we do not use euclidean formulation and the euclidean path. We use a wave-function and the Schroedinger equation. But the Schroedinger equation tells us that the wave-function is non-zero under the barrier. So there is a probability that we are sitting in the classically forbidden region.

So in QCD also some part of the wave function of the vacuum is in a classically forbidden region, in between the classical vacua, which are classically allowed.

So suppose you have a very classical problem: The double-well potential, which was done in the same year as quantum mechanics was invented, by Hund. The point is that there is a wave-function which is non-zero under the barrier. You want to know whether the particle really sits here. There are many ways of finding out. For example you may try to look in the microscope: You make a beam of particles which you focus at x around 0, where you want to localise the particle and see if it is there. If there is enough energy in your microscope beam particles, they will find the wave-function here and excite the quantum system into an excited state. If you plot the excitation probability versus the excitation energy you find a maximum which corresponds to the maximum of the potential, called V_0. In other words: If you localise a particle under the barrier, you observe a particle that sits on the barrier.

What it means for Yang-Mills dynamics is that an instanton is a tunnelling path from one classical vacuum to another. But in the language of wave-functions, it means that there is a probability to find something between the vacua. If you make a

high energy collision and the Yang-Mills field happens to be under the barrier, this a topological barrier, with the Chern-Simmons number as the coordinate. If one hits it very quickly, there is no time to change the coordinate, it just jumps up. So you produce a particular gluonic state which corresponds to the maximum of the potential and is called a sphaleron. In QCD these objects are clusters of glue. And they happen to be purely magnetic without electric field. Pure magnetic topological clusters of glue, with a mass about 3 GeV.

So if you do not want to think about euclidean space and correlation functions, this is the point of view to take. This means that if you collide two hadrons, you must produce some amount of these objects. That happens if the system was under the barrier. And what happens to them later we see in the experiment.

By the way: in the electroweak theory you have exactly the same excitation, it has electroweak sphalerons and instantons. In electroweak theory, the height of the potential is of the order of 15 TeV and the object which we produce is some ball made of Ws and Zs, hundreds of them, bound together. Everybody knows that Ws and Zs interact very weakly with each other because weak coupling is small, so how do we produce a ball made of hundreds of them? The answer is that you have coherent superposition of a hundred of particles compensating a factor of a hundred in the coupling constant. So you cannot bind together two Ws and Zs by weak interaction. They just do not bind, the interaction is too weak. But a hundred can bind. And this is actually a bound state, which can be produced.

In other words: Both in QCD and electroweak theory there are processes in real Minkovski time which start when the field is sitting under the barrier. This barrier is physical and the states are physical, we can excite them and see them in principle.

- *Markov:*

They are physical only in temporal gauge? Or is this interpretation valid in only one gauge or is it natural?

- *Shuryak:*

Physics should not depend on the gauge whatever happens, but the word "physical" should indeed be in quotation marks. In electroweak theory we create an object, a sphaleron, and this object decays into some number of Ws and Zs, you see them in your detector, you can calculate the probability of finding them. Nobody has seen that phenomenon and the calculation shows that there are many orders of magnitude missing, so nobody will see it. In QCD, one can produce such objects as well, but they are coloured objects, so they exist in the same sense as people use when they calculate a three-jet production from diagrams with three gluons emitted. After production of gluons, confinement takes care of making them colourless and it will go to some multi-hadronic state. But we can calculate the probability, ignoring the formation of strings etc., because it happens with probability 1. So in the same sense we can calculate cross-sections of production of sphalerons and they are in that sense

physical, although the final state is hadronic and not gluonic. This possibility is very exciting and unfortunately only very recently people started thinking about it in the context of QCD. In electroweak theory this topic was very hot 10 years ago but when people realised that it is really not possible to see it, it disappeared from Phys. Rev.

- *Polosa:*

The strongest argument in favour of the phenomenological relevance of instantons is that they explicitly break $U(1)_A$ symmetry. Anyway, according to Witten, DiVechia, Veneziano the $U(1)_A$ puzzle can be solved without instantons. Could you please comment on this?

- *Shuryak:*

They did not solve it. They thought they should try to solve it without instantons because of difficulties with instantons. This is a very good question, which brings me directly to the issue which I actually prepared. I was speaking here, in this same room, four years ago about basically the same subject and when the same question came up Witten, sitting in the first row, asked me: "What about large N_c limit of this theory?" First of all, large N_c of course was invented by Prof. 't Hooft. I can repeat what everybody knows. He considered the particular limit in which $N_c \rightarrow$ infinity and $g^2 \rightarrow 0$ so that some combination remains constant. He used arguments based on perturbation theory. The arguments are that, in this limit, physical quantities, like hadronic masses and string tension, exhibit a very simple behaviour. For the instantons it would mean that their size should not depend on N_c while the density is proportional to N_c. The two questions which were posed by Witten are as follows.

The first one was: "How is it possible?" Naively, small size instantons have action $(-8\,\pi^2)/(g^2)$. In the action there is no N_c but only g^2 because instantons are SU(2) objects. So $(1/g^2)$ is of the order N_c and small size instantons are exponentially suppressed. In the large N_c limit it is not clear how this exponential behaviour can be reconciled with the power behaviour of total physical observables, like the hadronic masses.

The other question deals with the fluctuations of the topological charge $\langle$(n° of instantons – n° of anti-instantons)$^2\rangle = \langle Q^2\rangle$, where Q is the topological charge. This is called the topological susceptibility, which Witten introduced . For a random poissonian distribution $\langle Q^2\rangle \sim \langle N\rangle \sim (N_c)^1$. However Witten argued: this is wrong. $\langle Q^2\rangle \sim (N_c)^0$ he claimed, and he wrote a paper about it in 1997 based on some limit of a string theory. Recently a paper by T. Schaefer appeared studying this large N_c limit with interacting instantons. Schaefer found random formulae were wrong; in the large N_c limit the interacting instantons become liquid-like and so the fluctuations are not proportional to N_c but to 1, as Witten predicted.

Finally he found what happens to η'. Witten said that at large N_c limit the η' is supposed to become light again, like a pion. They also become degenerate because the splitting between η' and π is suppressed by a factor $1/N_c$. This is indeed what happens

in the instanton liquid. Looking at the correlation function, you can see that the pion is very light. It is an attractive channel so the correlation function $K(x)/K_0(x)$ as a function of x goes up, while the η' channel goes down. In the large N_c limit η' becomes lighter. However, there is a very large coefficient. What I am saying is that every mathematical statement of Witten, that the fluctuations are supposed to disappear in the large N_c limit, and the η' mass becomes smaller, are all reproduced. Instanton dominance in $< Q^2>$ is very easy to verify on the lattice. For example you can smooth the fields, throw away basically all except instantons and then measure $<Q^2>$ and the η' mass in the remaining ensemble. People did it and they got the same value as in the original lattice configurations. They measured the topological susceptibility and concluded it is about 90% given by instantons: this is an established fact.

- *Polosa:*

At some point you quoted the f_0 and a_0 scalar mesons. Are they $f_0(980)$ and $a_0(980)$? If your f_0 is what is usually called $\sigma(500)$, then it is not clear if it is a $(u\bar{u}+d\bar{d})/\sqrt{2}$ state. It could be also a dynamical effect, like π-π strongly interacting in s-wave. How does this affect your considerations ?

- *Shuryak*

No, I do not mean $f_0(980)$ and $a_0(980)$. Indeed my f_0 is the σ. Whatever PDG decides or does not decide is a matter of policy of PDG. There is attraction in that channel and if I calculate correlation function for the σ it goes up. One also can see attraction from lattice calculations. So quark-antiquark are attracted in this channel and they form some wide resonance structure. While in the large N_c limit, η' becomes as light as a pion, σ mass goes up and becomes as heavy as $a_0(1400)$. An explanation is that in the large N_c limit we have a very rigid liquid with small fluctuations and so σ becomes heavy.

CHAIRMAN: E. SHURYAK

Scientific Secretaries: S. Hillert, H. Ohnishi

DISCUSSION II

- *Zichichi*:

On what basis can you exclude the existence of a baryon-meson-plasma (BMP) before the existence of QGP? The experimental results are perfectly consistent with a phase transition from the QCD-vacuum to BMP before the transition to QGP. In other words, it could be that in the nuclei-nuclei collisions what happens is:

$$\begin{array}{ccccc} \text{QCD-vacuum} & \rightarrow & \text{BMP} & \rightarrow & \text{QGP} \\ (1) & \rightarrow & (2) & \rightarrow & (3) \end{array}$$

On what basis can you exclude the sequence

$$(1) \rightarrow (2) \rightarrow (3)$$

and be sure that what happens is

$$(1) \rightarrow (3) \ ?$$

- *Shuryak*:

The transition from QCD vacuum to QGP takes place almost instantaneously, within a time scale of 0.5 fm. At the moment of the collision, the nuclei at RHIC are Lorentz contracted by a factor of 100. Hence they are extremely thin discs that quickly give rise to very dense matter. In the T-μ-phase diagram, the transition corresponds to a jump to a region, where the chemical potential μ is very small and temperature T is very high, about 300 MeV. The plasma has very large pressure divided by energy density $p/\varepsilon \sim 1/3$. It exists for a time of the order of 5 – 6 fm.

The transition (1)→ (2) → (3) would probably correspond to a different equation of state. Remember that the temperature during the collision is of the order of 100 – 200 MeV. At such a temperature, baryons of 1 – 2 GeV mass do not move very much. Hence no kind of matter made up of baryons can, at such temperature, have the large pressure that is observed in the big explosion at RHIC, while matter made up of massless gluons, which move relativistically, can hence have high pressure. The quark-gluon plasma scenario managed to explain the big difference between observations at CERN at $\sqrt{s}$ = 17 GeV and at RHIC at $\sqrt{s}$ = 130, 200 GeV. Alternative scenarios are possible, but they should be able to do it as well.

- *Polosa:*

How can you make use of equilibrium statistical mechanics and hydrodynamics for a system which is at equilibrium for a duration of only 1 fm?

- *Shuryak*:

We can always use statistical mechanics and hydrodynamics if a system is large enough compared to the correlation length in that system. The data seem to indicate that the correlation length is of order of a fraction of a fm, while the diameter of Au nuclei is about 14 fm. In ideal hydrodynamics, entropy is conserved. However, in reality, non-ideal terms also contribute, for example the viscosity term. Consider matter, which consists of particles that propagate according to perturbative QCD. The viscosity coefficient of such matter can be calculated from pQCD to be

$$\eta \propto \frac{T^3}{\alpha_s^2} \cdot \log \frac{1}{\alpha_s} .$$

That result seems to be qualitatively wrong, as it differs from the data by a big factor. In a recent paper by D. T. Son and collaborators, the viscosity of N=4 supersymmetric CFT-theory matter is calculated in the strong coupling regime, where the 't Hooft coupling $g^2 N_c >> 1$. Using the Maldacena conjecture and AdS-CFT correspondence, viscosity is found to be $\eta \propto T^3$, below the pQCD result by a factor of 30. This theory is an example of how one may arrive at an ideal liquid with a very small mean free path and very small viscosity. So the question is how can it be so non-viscous? Efforts are made to invent other models in which QGP is more liquid-like and less gas-like. The topological cluster model I spoke about is one of them.

- *Polosa*:

You had a plot, in which the mass of the pion grows with N_c. Does this mean that chiral symmetry is restored at high N_c?

- *Shuryak*:

It was not a mass going up: it was the correlation function as a function of distance x going up. The pion is a Goldstone boson. Chiral symmetry is violated at any N_c. Therefore, the pion is massless for any N_c. What was illustrated in the plot you mention is a correlation function and what I wanted to show is that ε', which is a heavy particle in our world – it has a mass of about 1 GeV – becomes lighter and lighter as N_c increases. It means its correlation function decays less rapidly with distance.

- *'t Hooft*:

You explained that the effects of instantons do not disappear in the large N_c limit. Now one would expect that instanton effects, going like e^{-C/g^2}, will go like $e^{-C'N_c}$. They then look like soliton solutions in a N_c theory. But C′ must be zero, somehow. Is there a way to understand such "solitons" in a $1/N_c$ expansion?

- *Shuryak*:

What happens to the instanton in the large N_c limit? The action is indeed $e/g^2 \propto e^{-N_c}$. Witten argued that because in the 't Hooft limit $g^2 N_c$ = const, the expression e^{-1/g^2} approaches 0 very fast. New instanton liquid solutions by Thomas Schäfer for larger N_c demonstrated that because g runs and the term e^{-N_c} should be replaced by $e^{-N_c f(\rho\Lambda_{QCD})}$ with a function $f(\rho\Lambda_{QCD})$ depending on the instanton size ρ. In the large N_c limit, instantons concentrate at $\rho = \rho^*$, where that function is zero, $f(\rho^*\Lambda_{QCD}) = 0$. One can see that as follows: for $f(\rho\Lambda_{QCD})$ positive, the action is exponentially suppressed, for $f(\rho\Lambda_{QCD})$ negative, it is exponentially enhanced and is stabilized by repulsion. In the large N_c limit, the total number n of instantons is proportional to the first power of N_c, $n \propto N_c^1$ and ρ approaches a constant value of about $\rho^* \approx 0.45$ fm. The $n \propto N_c^1$ dependence complies with the general counting rules which you established for perturbative diagrams. Also the masses of mesons, as well as the instanton contributions to these masses, are independent of N_c in the large N_c limit. The exception is η', its mass is decreasing to zero, as Witten suggested long ago.

- *Korthals-Altes*:

So the function f comes from the group entropy – the number of ways you can embed the instanton in the SU(N) group – and interactions?

- *Shuryak:*

The function $f(\rho\Lambda_{QCD})$ comes from three quantities: the action itself – this is just a logarithmic dependence, the entropy factor you mentioned, and the interaction between instantons. The combination of these contributions ensures that the factor $f \rightarrow 0$ as $N_c \rightarrow \infty$. The instanton densities for all N_c values intersect at the instanton size $\rho*$. In other words: all physical quantities are smooth in N_c except for the instanton size, which approaches a delta function in the large N_c limit, $dn/d\rho \rightarrow \delta(\rho-\rho*)$.

- *White*:

You argued that things changed dramatically when we increased $\sqrt{s}$ by more than an order of magnitude going from SPS to RHIC, and that we do not yet have a dynamical explanation. Could you speculate on what will happen at LHC energies?

- *Shuryak*:

The multiplicity, and hence the total entropy production, is not changing dramatically: the total number of pions is increasing by less than a factor of 2 at RHIC compared to SPS. We do not have much more matter in the end, but observe a much more rapid onset of this collective behaviour. This shows that the mean free path is very sensitive to the local density. We attribute it to crossing into the QGP domain, the phase transition. Going further toward LHC, no phase transition is predicted. The region at very high energies, like large p_t, might again be described by the parton model. At the moment I would not speculate about LHC: we have to understand the RHIC data first.

- *Hallman*:

Hydrodynamics reproduces the elliptic flow for particle identified spectra below 2 GeV, but fails to reproduce R_{out} / R_{side} from HBT. How is this resolved?

- *Shuryak*:

The first comment is very trivial. In hydrodynamics, the equation of motion is integrated. For a given force, the velocity is found by integration of acceleration. Now, HBT radii are positions at the end of the expansion. Position is found by integrating velocity over time. In order to do a double-integral I have to know my function much better, so HBT radii are sensitive to details of the freeze-out, i.e. to where exactly one should stop the hydrodynamical expansion. What can currently be reliably predicted is the velocity distribution and in particular the azimuthal distribution for different types of secondaries, as a function of centrality and of the collision energy. The description of the exact shape of the matter at the moment when it stops to interact is not yet accurate. The error of the radii is 30 – 40 %. I expect the theory of hydrodynamics to improve in the future. We know a couple of effects, from which a change of radii by ≈ 20% in the right direction can be expected, but this has not yet been quantitatively worked out. There will be adjustments, but the gross picture is clearly a "Bang", with multiple rescatterings and approximate hydrodynamical behaviour.

CHAIRMAN: E. SHURYAK

Scientific Secretaries: S. Hillert, A. Maas, G. Marandella, H. Ohnishi

DISCUSSION III

- *Ohnishi:*

In the previous discussion, when S. White asked you about LHC perspectives, you answered that before we need to understand the RHIC data. My question is: do we have a clear explanation to describe CERN/SPS data? What is the conclusion about SPS energy? How to explain anomalous J/ψ suppression and low mass dilepton enhancement?

- *Shuryak:*

We have a reasonably good explanation of what is happening at SPS. It is also an explosion but less energetic than at RHIC. Its general space-time development is under control. We know there is J/ψ suppression, we know how it depends on centrality. The suppression pattern is non-trivial and does not show a smooth behaviour. There is no clear answer whether it is a fluctuation in the experimental data or a real signature of the phase transition. This will be ensured if RHIC measures the same result at the same density but, of course, at higher energy and more peripheral collisions. In principle, RHIC can go down in energy to the SPS energy range. RHIC has a power to explore the whole energy region between SPS to RHIC. I remind you that the collision energy at SPS, $\sqrt{s}$ = 17 GeV, and RHIC is $\sqrt{s}$ = 200 GeV. We will know more physics results from RHIC with time. If you are asking whether SPS physics are finished, my answer is: no, it should not be finished. But according to CERN plans it looks like it is finished. So the SPS physics will be continued at lower collision energies at RHIC.

- *Stamen:*

You get a very good prediction for the proton mass but you cannot say anything about confinement. Is this a principle problem of the theory or a lack of calculation?

- *Shuryak:*

We know how to use the instanton model. We know the criteria of confinement, for example the area law of the Wilson loop. If I take a configuration which is a simple super position of instantons and I do not have anything else, I can either calculate the Wilson loop by myself or ask lattice friends to do it for me. We did both. Then we compared the results in great detail: they are the same. The answer is that there is no area-law and so an ensemble of instantons does not confine. Furthermore, I can calculate a potential between two heavy quarks. The calculation does not

resemble the real potential between heavy quarks. It is several times smaller and is not growing at large distances. Instantons do not contribute much to heavy quark physics. On the other hand, if you go to light quark physics, you can do the opposite. You can switch off Coulomb, switch off confinement, this is reached on the lattice theory by smoothing configurations. When one does Feynman diagrams, you do the integral over the variable "k", where "k" is the momentum, one makes integral over many harmonic oscillations. When one makes lattice configuration smooth, one kills all these harmonics, so there are no Coulomb, no perturbative QCD at all. One also seems to be killing confinement by this operation. Lattice people have found that you can go, more or less, to the instanton vacuum. Heavy hadron masses and wave functions would be completely wrong after this procedure because of the difference of the calculated potential from the instanton model; they will not be confined. All mass levels for hadrons containing heavy quarks except the Upsilon would be absent, while in reality, we have many mass states there. Now for light quarks, it turns out if I switch off confinement, very little happens, the lower mass levels stay where they are. I told you that next levels, the prime levels, are much less sensitive to instantons. While the prime-prime levels are insensitive to instantons to a degree which we were really surprised about. The highly excited states of hadrons are bounded because of confinement. However, it is interesting that instanton forces are strong enough to make bound states of constituent quarks. I repeat that these bound states resemble very closely what we observe experimentally. So I conclude that confinement is not very relevant for these particles, while instantons are not very relevant for the highly excited ones. It is not a lack of calculation; we can calculate everything for any configuration.

- *Maillard:*

If I have Yang-Mills or gravitational theory in which I put a scalar field, I want to change the vacuum energy of the scalar potential by the introduction of an instanton. What is the consistency condition or restriction for the introduction of such an instanton?

- *Shuryak:*

First of all, we have to understand whether there is a non-trivial topological structure in a theory like gravity, like the structure in the field of Yang-Mills theory (YM). YM has topologically distinct classical vacua and tunnelling between them. In gravity I can also imagine there are configurations which are topologically different between each other and then there could be quantum tunnelling from one configuration to another. By the way, even in quantum mechanics, we know that we account for tunnelling between two wells, the lowest symmetric level goes down. In QCD, instantons also give energy shift which is negative, after subtraction of perturbative infinities, the energy of vacuum goes negative, it is roughly minus 1 GeV/fm^3, a huge value. So, I presume in any theory with a scalar field and gravity,

there would be some tunnelling and, like each other tunnelling, it will automatically shift the ground state energy and contribute to the cosmological term. The problem is that the cosmological term which you get from QCD or the standard model is so many orders of magnitude above what we actually observed. We have no clue how the real cosmological constant is formed and why so many orders of magnitude are cancelled and what is left is non-zero.

- *Chakrabarti:*

Is it possible to get instanton solution without zero mode? Charles Thorn has shown that in (1+1)D φ^4 theory, one can have soliton-antisoliton solution without zero mode.

- *Shuryak:*

If you can make a scalar theory which has some kind of a soliton and then declare that one of the axes is time, Euclidean time, then that is the instanton. So there is no real technical distinction between the instanton and solitons. You can always go to the physics of lower dimensions and take one of the coordinates to be time. So that would then be the instanton.

I am confused which zero-modes you mean. So let us take the most simple example, the so-called domain wall with, for example, sine-Gordon, that you have a one-dimensional soliton which changes the phase by 2π. Then there is a bosonic zero-mode which means that the wall can be shifted and there is a fermionic zero-mode which means that the fermion can live on this wall and has exactly zero energy. That is the mode which is used for so-called domain wall fermions by lattice people. Any topological solution, whether you interpret it as a soliton or instanton etc,. has exact zero-modes for fermions.

- *Bechtle :*

What would be necessary to measure the pion form-factor which differs for high Q^2 from perturbative QCD prediction experimentally?

- *Shuryak:*

The Form-Factor by definition comes from elastic scattering of pion. One scatters virtual photon on pion and you get back your pion. That is the Form-Factor. At larger and larger Q^2, above say 2 GeV^2 to 10 GeV^2, the problems in experiments exist because it is difficult to make a target made by pions. You have to extrapolate some very peripheral scattering of the proton. So, there is some ambiguity with the elastic Form-Factor of the pion. The difference between models with instantons and without instantons is as follows. If you imagine non-relativistic quantum mechanical description of the pion wave function, you have two models for a particle like a pion. The first model is that of a square wall. The particle is bound because there are walls. The other model is that the particle is bound because there is a Δ function at the origin. You understand, from quantum mechanics, that the wave function would have

a different shape in each case. In the first example, there would be a very smooth sinusoidal wave function at the origin. In the second example there would be a cusp at the origin. So I am saying that the instanton produces a cusp in the pion wave function. This is what makes a harder form-factor and makes it alive when you shake the pion very hard. And I am saying that the ρ would not behave like this. I am claiming that the nucleon has also a hard form-factor, much larger than the Δ for example would have, because there are so-called diquark clusters, which is also very compact for the same reason as the pion is.

- *Markov:*
  Do instantons violate the baryon number conservation law?

- *Shuryak:*
  There are two kinds of instantons. Instanton in electroweak theory is a tunnelling in the space of W and Z fields, it indeed violates baryon number conservation. That supposes to give very spectacular events in which about 50 Z and W and 12 fermions are produced, in such a way that the nucleon number is changed by 1. It would be fantastic to observe one of it, but unfortunately, the probability is something like 10^{-80}. The original factor which follows from 't Hooft calculations was 10^{-170}, then people realized how to increase it and then increased it 90 orders of magnitude, but still 80 is left there. So, one cannot see such an event.

  On the other hand, in QCD the baryon number is not violated. QCD instantons create $\bar{u}$, $\bar{d}d$, $\bar{s}s$. The reason for this is very simple. Quark zero-mode is left-handed and antiquark is right-handed. In weak interaction, right-handed are absent. Therefore it can have only one mode.

  In QCD, quarks are both left-handed and right-handed. Baryon number is not violated. The only violated quantum number, which is called the axial charge, is the total number of left-handed minus right-handed particles in the universe. When one converts left into right-handed quarks, it is much less spectacular. Experimentalists did not figure out how to measure axial charge and we cannot check its violation in an event. We can study some correlators, like these triangles where I flip chirality from right to left and then back, or I do not. It was observed experimentally, that when I have a small distance triangle with two photons and one pion, there are small corrections. And when I have two pions plus one photon, there is a big effect. Therefore I conclude that instanton induced chirality flipping interaction does exist. The same effect gives mass to the ε'. So there are ways to see it, but you do not have a detector which measures violation of the total axial charge.

- *Ramtohul:*
  Have diquarks been seen in experiment?

- *Shuryak:*

No, of course not as free particles. They are coloured objects. There is confinement in QCD. You may ask if there are diquark clusters inside a nucleon, whether the two quarks in one configuration, say spin 0 and isospin 0, are more compact and have a lower distance to each other than to the third one. This is very important for the difference between the nucleon and the Δ structure. Δ is not supposed to have a compact diquark structure. There were some conferences on diquarks which published some kind of summary papers in Rev. Mod. Phys., you can find there 20 different arguments why diquarks of spin-isospin = 0 exist as compact clusters. These are phenomenological statements. In reactions like jet physics or fragmentation of strings, some people very routinely use these diquark clusters in phenomenological models. In this case the single quark goes one way, and diquark goes the other way. After that there is a colour string that breaks. That explains the data. Another reason why we speak about diquark is that this specific argument leads to the discovery of colour superconductivity, by the way. Because we realized that there is a strong pairing force between two quarks, not only quark and antiquark, which is the same force which creates the quark condensate and eventually all the masses. One more reason I speak about diquark theory is because of the QCD with two colours, in which one does not need the third quark. Then the diquarks are physical hadrons which can be calculated on the lattice and have a spectrum. Then one finds indeed that for different quantum numbers of these diquarks there are very different masses. You can deduce from that what is the ratio of the force between two quarks to the force between quark and antiquark. This is the main point I want to emphasize and the ratio is supposed to give a simple formula, $1/(N_c-1)$. It vanishes at $N_c \rightarrow \infty$, but gives 1 when $N_c = 2$. The real world, $N_c = 3$, is exactly in between.

Erice Lectures on Confinement and Duality

Matthew J. Strassler[†*]

[†] *University of Washington, Seattle, Washington, USA*

February 27, 2003

*strassler@phys.washington.edu

Contents

1 Confinement: An Overview **156**
1.1 Some preliminary questions 156
1.2 Confinement in pure Yang-Mills 158

2 Confinement of Magnetic Flux **163**

3 Duality, Sources and Fluxes **168**
3.1 Electric sources and fluxes 168
3.2 Magnetic sources and fluxes 168
3.3 Conclusion 170

4 Duality and Confinement **171**
4.1 Duality and QCD 171
4.2 Seiberg duality for $SO(N)$ gauge theories 172
4.3 Confinement or not? 174
4.4 Spinor sources from the dual point of view 175
4.5 Conclusion 177

These lectures deal with the issues of and the relationship between confinement and duality. Let us look at them in turn.

Confinement: in $SU(N)$ QCD (quarks and gluons), and in $SU(N)$ pure Yang-Mills (gluons only), and related theories, understanding how and why confinement occurs is difficult and confusing. In order to get some insight into confinement, it is important to have some simpler "toy models" which display the same phenomenon but are easier to study. We could try confinement in two dimensions, or in three; but these are kinematically very different, and are probably misleading. In these dimensions, even classical electrodynamics shows a sort of confinement; the potential energy between an electron and a positron diverges as the distance between them is taken to infinity!

Instead, we might consider four dimensional $\mathcal{N} = 1$ supersymmetric field theories with confinement. These at least have the right kinematics and the right spacetime, and the classical theory is scale invariant (except perhaps for quark masses), just as in QCD and/or Yang-Mills theory.

Now, what is the connection between supersymmetry and confinement? There *isn't* any, as far as I know. And that's good! We want to study the latter, not the former, and the less the confinement process depends on the supersymmetry, the better. For our purposes, we're going to think of supersymmetry as a purely technical device, one which does not change the basic issues surrounding confinement, but makes calculations much easier.

The questions we will address in this first lecture are these: what is confinement? How do we test for confinement? We will look at an magnetic version of confinement — the Meissner effect in superconductors — and this will lead us to consider the possibility that we can understand confinement using electric-magnetic duality.

Duality: in $SU(N)$ QCD (quarks and gluons), and in $SU(N)$ pure Yang-Mills (gluons only), and related theories, understanding how and why duality works is difficult and confusing. In order to get some insight into duality, it is important to have some simpler "toy models" which display the same phenomenon but are easier to study. We could try duality in two dimensions, which actually does provide some general insights; the three dimensional case is too hard already. But with supersymmetry, both in three and in four dimensions, duality is somewhat better understood, and we can use supersymmetric theories, of the same sort mentioned above, to investigate duality, and in turn use duality to gain insight into confinement — our main goal.

In our second lecture, we will ask "what is duality?" From there we will explore an example of Seiberg duality, and use it to learn about the possible phases of gauge theories. Finally, we will see how we can test a theory, using another Seiberg duality, to find out whether it exhibits confinement.

1 Confinement: an overview

1.1 Some preliminary questions

Let us begin with a set of questions, to test your understanding?

- *Is QCD a confining theory?* Well, what do we mean by this? Let's try to refine it...

- *Are quarks confined in QCD?* This is clearly ambiguous. We can think of two possible meanings to this question:

  - If I create a quark-antiquark pair a small distance apart, can I pull them infinitely far apart using a finite amount of energy?
  - Are there finite-energy states in the theory containing isolated quarks?

  These questions do not obviously have the same answer. In a state without isolated quarks, it might be impossible to create isolated quarks using pair production; but still there might exist states with isolated quarks, and we might happen to live in one. But the answer to the first question is clearly "no". If you pull on a quark living inside a proton, you will have no trouble extracting it. You'll generate a region of large flux density when you do this, and a quark-antiquark pair will be created, leaving you with a baryon and a meson (Figure 1). But you will have quite successfully separated the original quark from its brethren. So clearly QCD is not a confining theory unless we ask the second question. And how will we answer it? Maybe this is too complicated. At least in Yang-Mills theory we all know there is confinement...

- Or do we? *Is Yang-Mills a confining theory?* What do we mean by this?

- *Are gluons confined in Yang-Mills theory?* Well, if you have a glueball, and you try to grab hold of a gluon inside, you will face the same problem as with the quarks inside a proton: gluon pair production will occur, and you'll be left with two glueballs.

- *What is confined in Yang-Mills theory?* It is **Electromagnetic Flux** that is confined. This is the key point! In electromagnetism, or even classical Yang-Mills theory, flux spreads out uniformly. Not in Yang-Mills theory! or QCD. You will never find yourself in a region of weak color fields analogous to the extremely weak electromagnetic that cause dust to cling to your clothing. Instead, the flux of Yang-Mills theory binds together into tubes (Figure 2), with energy densities of order Λ^4, where Λ is what we will call the confinement scale.

- *How can we demonstrate that flux is confined in Yang-Mills?* We can't use the gluons; as we already discussed, the potential energy of two gluons does not go to infinity as we pull them apart. Let's approach it from a different angle.

- *What happens to QCD if all quarks have mass m_q much greater than the strong-coupling scale Λ_{QCD}?* Well, in this case pair-production of quarks is exponentially suppressed, because the energy densities associated to the field strengths which we can obtain are of order Λ^4_{QCD}, so a quark pair of mass m_q cannot be formed without a large upward fluctuation in the energy. So if we build a proton from heavy quarks and we try to pull a heavy quark out of this proton, we now get a tube of flux that, although it can eventually break, is *metastable*; it lives an exponentially long time. Eventually the tube will snap, and the picture will be as in Figure 1; but for a long time the picture in Figure 2 will apply. This is approximate confinement, and it becomes exact as $m_q \to \infty$.

- *Does flux confinement imply quark confinement?* Well, almost; if the quarks are heavy, and the flux tubes are long-lived, then Gauss's law implies that the potential between a quark and antiquark, which is generated by the flux tube between them, grows *linearly with distance* Figure 3.
- *Does quark confinement imply flux confinement?* No. The potential between a quark and antiquark could grow like a logarithm of distance, for instance; in this case flux would still spread out, though not as rapidly as in electromagnetism.

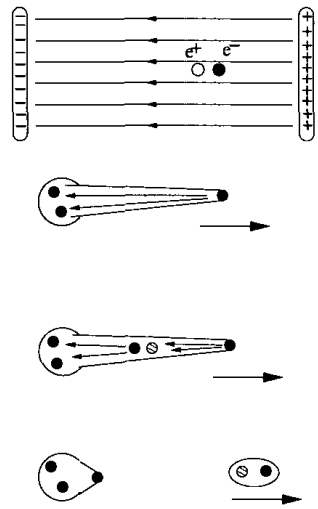


Figure 1: As with pair production of electrons in a strong electric field, pair production occurs as a quark tries to escape from a proton.

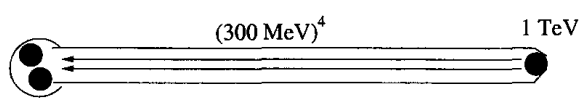


Figure 2: If all quarks were heavy, then flux tubes would break much less readily.

1.2 Confinement in pure Yang-Mills

Let's turn our attention to Yang-Mills theory, which has only a gauge boson A_μ in the adjoint of $SU(N)$. The group $SU(N)$ consists of $N \times N$ matrices $U^\alpha_{\bar\beta}$ (with row indices α and column indices $\bar\beta$.) which are special ($\det U = 1$) and unitary ($U^\dagger = U^{-1}$). The "gluon" field A_μ takes values in the algebra of $SU(N)$,

$$(A_\mu)^\alpha_{\bar\beta} = A^a_\mu (T^a)^\alpha_{\bar\beta} \ .$$

Here T^a is a generator of the group $SU(N)$, also an $N \times N$ matrix (normalized to tr $T^a T^b = \frac{1}{2}\delta^{ab}$,) and the group index a runs from 1 to the dimension of $SU(N)$, namely $N^2 - 1$. The

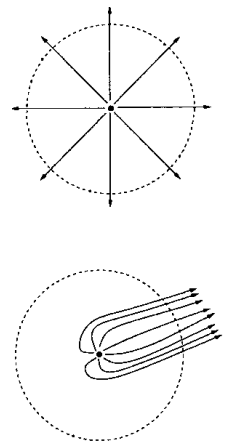


Figure 3: Gauss's law for unconfined and confined flux.

theory has the simplest possible Lagrangian; defining $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu]$ (here F and A are matrices and the brackets indicate a matrix commutator), we write the Lagrangian as

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{tr}\ F_{\mu\nu}F^{\mu\nu}\ .$$

This normalization of the field A_μ differs from the one in standard textbooks on perturbation theory. In perturbative calculations, it is more convenient to absorb the $1/g$ into A_μ; then $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu]$. The quadratic terms in the Lagrangian are then the free Maxwell equations, and do not depend on g. We may then think of the theory as a set of free fields — simply $(N^2 - 1)$ independent photons — coupled together by interactions of order g. However, in these lectures we will not assume small g, and will rarely expand in powers of g. The normalization chosen here is more profound; it puts the coupling constant in its proper place, multiplying $\hbar$ and therefore determining the size of all quantum effects. Most nonperturbative properties of the theory will involve either $1/g^2$ or e^{-1/g^2}.

Pure Yang-Mills theory is weakly coupled at high energy, like QCD, and becomes strongly coupled at a scale Λ. And at this scale, it is believed (on the basis of several arguments and numerical simulation) to have confinement of flux. The strong coupling dynamics makes it impossible to talk about gluons at low energies. Instead, we have only bound states, whose name "glueballs" is reasonably accurate, in that these gluey states do not really consist of a fixed number of gluons, but rather of a shifting mass of chromoelectric flux lines. There are a large number of these states.

This theory has only gluons, and as we discussed earlier, they are not themselves confined. In Yang-Mills theory, *"confinement" means that chromoelectric field is confined; it cannot spread out in space over regions larger than about* Λ^{-1} *in radius.*

Now, how can we detect the presence of the tubes — "strings" — which contain the chromoelectric flux? Ideally we would like to find a long and straight flux tube and find its tension (energy per unit length) but we might have trouble convincing one to stay straight long enough to do this measurement. So here we need a new idea. Recall how the heavy

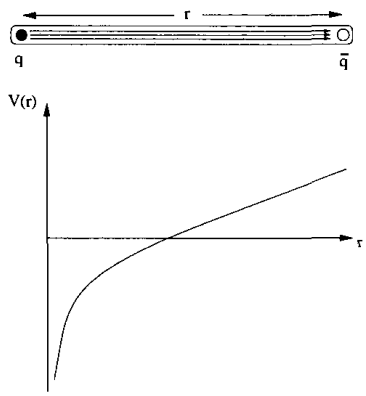


Figure 4: The confined field lines between a heavy quark and antiquark form a tube; the potential energy of the system goes as $1/r$ at distances short compared to Λ^{-1} but becomes linear at larger distances.

quarks of QCD-with-no-light-quarks were truly confined. This suggests that the way to detect confinement of flux in Yang-Mills theory is to put some extremely heavy quarks in it — so heavy that they can't affect the dynamics of the Yang-Mills theory — and see that these quarks are confined! That is, we can compute the quark-antiquark potential $V(r)$ and see that it grows without bound (indicating confinement) and more specifically is linear in r (indicating confinement by flux tube.) Why is the linear potential characteristic of a flux tube? Well, consider Gauss's law. In an unconfined theory, the electric flux is uniformly distributed over a sphere surrounding a charge, and therefore falls off as $1/r^2$. In a confining theory with flux tubes, the flux tube has a fixed cross-sectional area of order Λ^{-2} no matter how long it is; and thus, for any sphere of radius $r \gg \Lambda^{-1}$ surrounding a charge, the flux on the sphere is zero everywhere except in a region of area Λ^{-2} where the flux tube passes through the sphere. From this we conclude that the electric field in that region has a magnitude which is r-independent! In turn, this implies the force that it generates on a test charge is also r-independent, and finally, that the potential between charges grows linearly with r.

So, let us add a charged fermion (or scalar) to the Yang-Mills theory, one whose mass M is so much larger than Λ that it cannot play a role in the strong-coupling physics. Adding a quark ψ we make the Lagrangian

$$\mathcal{L} = -\frac{1}{2g^2}\text{tr } F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi - M\bar{\psi}\psi \ .$$

The quark ψ is charged under $SU(N)$, but for the moment let us not specify the representation R of $SU(N)$ under which it transforms. Now let us consider the potential $V(r)$ between ψ, placed at one position, and $\bar{\psi}$, placed a distance r away. Since the quarks are very heavy, we can expect that they can be placed at rest and will move only very slowly, allowing us to do this computation. Confinement means that when r is large, a string — a tube of chromoelectric flux — stretches between ψ and $\bar{\psi}$, of constant tension T_R, such that the potential $V(r) = T_R r$ [1]. The force between two such fermions goes to a constant, and never drops off to zero. (That these facts are true in Yang-Mills theory does not follow from

any direct theoretical calculation. Highly quantum mechanical in nature, they have only been checked using direct numerical simulation of Yang-Mills theory.)

In the limit where $M \to \infty$, the quarks become completely non-dynamical [1]; they are what we may call "chromoelectrostatic sources", probes which never appear in any loop diagram and thus are purely classical. What remains dynamical is the flux tube. Thus, we didn't really need the quarks as physical particles; using nondynamical chromoelectric sources, we could have detected the confinement of chromoelectric field, which is a property of the Yang-Mills theory without the added quarks. (An equivalent way to make this statement, without introducing the quarks, is to talk about Wilson loops in various representations R; in a confining theory the value of the Wilson loop is proportional to the exponential of minus its area, with proportionality constant T_R .)

In general, the string tension, and the corresponding force, between quark and antiquark can depend on the representation R. After all, why not? In particular, for R the adjoint, we already know $T_{adjoint} = 0$: any fermion in the adjoint can combine with a light gluon to make something gauge neutral, so two such fermions will each cloak themselves with a gluon and will feel no long-range force as we pull them apart. So clearly we need to think about how things depend on the representation R. Clearly the map from representations to flux tubes cannot be one-to-one (since both the trivial representation and the adjoint representation have $T_R = 0$.) Lie groups have an infinite number of representations, but the stable flux tubes number at most $\dim C_G$, the dimension of the center of the gauge group. Let us see why this is so.

What is the center of $SU(N)$? A matrix $U^{\alpha}_{\bar{\beta}} = e^{2\pi ik/N}\delta^{\alpha}_{\bar{\beta}}$, $k = 0, \ldots, N-1$, is an element of $SU(N)$. Being proportional to the identity, it obviously commutes with everything in $SU(N)$; in short, U is in the center $C_{SU(N)}$. The elements of the center are thus labelled by the integer k, which from the definition of U is only determined modulo N, so the labels form the group $\mathbf{Z}_N$, the additive integers mod N. Now consider any representation R. An element ρ of this representation is labelled by a certain number n of unbarred (upper) and $\bar{n}$ of barred (lower) indices; that is, it takes the form $\rho^{\alpha_1\alpha_2\cdots\alpha_n}_{\bar{\beta}_1\bar{\beta}_2\cdots\bar{\beta}_{\bar{n}}}$. Under a group transformation, each unbarred index is rotated by the matrix U, while each barred index is rotated by $U^{\dagger}$. Consequently, the transformation of the representation R under the center C_G is by the phase $e^{2\pi ik(n-\bar{n})/N}$, where $n - \bar{n}$ is called the "N-ality" of the representation. The adjoint representation, with one upper and one lower index, is invariant under the center. The fundamental $\mathbf{N}$ representation (one unbarred index) rotates by $e^{2\pi ik/N}$; the antifundamental $\bar{\mathbf{N}}$ rotates by $e^{-2\pi ik/N}$. Both the antisymmetric-tensor and symmetric-tensor representations $\mathbf{N(N \pm 1)/2}$, which have two unbarred indices, rotate by $e^{2\pi i(2k)/N}$. Indeed, all p-upper-index tensors carry charge p under $\mathbf{Z}_N$ — that is, they rotate by $e^{2\pi ipk/N}$ under the kth element of $\mathbf{Z}_N$. In short, the representations R of $SU(N)$ break up into equivalence classes under the center, and can be labelled by their "N-ality" charge p [2, 3]. Note that the conjugate representation of R has "N-ality" $N - p$, since the number of barred and unbarred indices is exchanged.

Why is this interesting? First consider, for example, adding a heavy quark ψ_A, in the antisymmetric representation, to Yang-Mills theory; the potential between $\bar{\psi}_A$ and ψ_A is $V(r) = T_A r$. Now consider instead adding a heavy quark ψ_S in the symmetric representation; the quark-antiquark potential between $\bar{\psi}_S$ and ψ_S is now $V(r) = T_S r$. Suppose that $T_S > T_A$

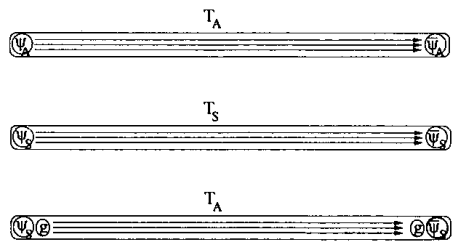


Figure 5: The flux between different quarks, or combinations of quarks and gluons, all with N-ality 2.

in Yang-Mills theory. (This is probably true, but what I'm about to say won't depend on the specific assumption.) Nothing prevents the theory from taking one of its light gluons (remember their number is not conserved so it need not be pair-produced) and putting it very near ψ_S. The combination of the gluon A_μ and the fermion ψ_S looks, from a distance, as though it were a single object. What is its charge? Well, we must consider the group theory of $SU(N)$; what is (adjoint)$\otimes$(symmetric)? It is a direct sum of a number of representations, *all of which have the same "N-ality" as the symmetric representation, namely 2.* Said another way, the product of $(A_\mu)^\alpha_{\bar\beta}$ and $\psi_S^{\gamma\delta}$ can lead, no matter how the indices are contracted, only to representations with two more upper indices than lower indices. Among these representations is the antisymmetric representation. (In $SU(3)$, for instance, the symmetric tensor is $\mathbf{6}$, the antisymmetric tensor is $\bar{\mathbf{3}}$, and $\mathbf{8}\otimes\mathbf{6} = \bar{\mathbf{3}}+\mathbf{6}+\overline{\mathbf{15}}+\mathbf{24}$.) But then, since we assumed $T_A < T_S$, there exists a dynamical process by which the theory may lower its energy! By popping a gluon out of the vacuum and putting it near ψ_S, the theory can make ψ_S look more like a fermion in the antisymmetric representation. The same goes for $\bar\psi_S$. Then, instead of a string of tension T_S, a string of tension T_A can link these two fermion-gluon combinations. The energy cost is that of making two extra gluons — at most of order Λ — while the energy gain is $(T_S - T_A)r$, which for r sufficiently large always wins. The reverse process will hold if $T_A > T_S$.

More generally, the fact that gluons are in the "N-ality"-zero adjoint representation implies that *the presence of nearby gluons can change one representation to another but only in a way that conserves N-ality.* Thus in Yang-Mills, the representation R of a chromoelectric source is not a conserved quantum number; only its "N-ality" is actually conserved. Consequently, we should expect that for the entire class of representations with the same N-ality charge, there will be only one stable configuration of strings (which might involve one or more tubes — for "N-ality"=2 there might be one tube with two units of flux or two tubes with one unit each.) *The tensions of the stable strings, or combinations of strings, are labelled not by R but by the N-ality p of R.* Charge conjugation symmetry also ensures that $T_p = T_{N-p}$; thus we have of order $N/2$ stable flux tube configurations in $SU(N)$ Yang-Mills theory.

Can we see this in $SU(3)$ Yang-Mills? Yes and no. There is N-ality 0,1, and 2; but $T_0 = 0$ while $T_2 = T_1$, so only one confining string is predicted. The nontrivial statements are then only that, for example, the symmetric $\mathbf{6}$ representation of $SU(3)$ is confined by the same string tension as the antisymmetric tensor, the $\bar{\mathbf{3}}$; this in turn has the same tension as the fundamental $\mathbf{3}$. To have a nontrivial set of strings we must go to $SU(4)$; here the

antisymmetric tensor **6** should have a tension T_6 different from that of the 4 and $\bar{4}$, T_4. There is still a question as to whether $T_6 < 2T_4$; if not, the flux between two **6** fields may be carried by two strings of N-ality 1 rather than a single string of N-ality 2. Theoretical arguments [4] and lattice calculations [5, 6, 7] support the view that $T_6 < 2T_4$ (and similarly in other theories) so that there really are two independent stable flux tubes, of N-ality 1 and 2 (and again $T_3 = T_1$.)

To summarize, we expect that Yang-Mills theories have stable flux tubes labelled by a charge in the center of the group [2]; for $SU(N)$ this is its N-ality, a charge under the $C_{SU(N)} = \mathbf{Z}_N$ group action. While the gluons are not confined by these strings, any heavy quark with nonzero N-ality will experience a linear potential energy and a constant force which will confine it to an antiquark, or more generally, to some combination of quarks and antiquarks which have the opposite N-ality. (For example, it could combine with $N-1$ other quarks to form a baryon. As another example, a **6** of $SU(4)$ could combine with two $\bar{4}$ quarks to form an exotic object not found in real-world $SU(3)$ QCD.)

In our next lecture, we will find it interesting to study the group $SO(N)$. What happens to this story there? A typical $SO(N)$ group [more precisely, its double cover, $Spin(N)$] has "vector" representations of dimension N and "spinor" representations of dimension $2^{[N-2]/2}$ for N even and $2^{[N-1]/2}$ for N odd. The center of the group depends on N, but one element h_0 of the center which is always present is the one which is **1** acting on vectors and adjoints and $-\mathbf{1}$ acting on spinors (and antispinors, if they are distinct from spinors.) Thus spinors are odd under this group element while pairs of spinors, and vectors, are even. Suppose we consider theories with light quarks in the N representation. In this case, when we choose a source in representation R, both the gluons and the light quarks can be used to to change the charge of a source. However, neither gluons nor the light quarks can change the spinor nature of a source; if the source is a spinor it may be changed to an antispinor, but it cannot be changed to a vector, or any other representation even under h_0. Thus we can test for confinement in a theory with light, or even massless, quarks in the N representation by using sources in the spinor representation.

2 Confinement of Magnetic Flux

Now let us try to understand why and how confinement occurs. In Yang-Mills theory it occurs through a process requiring strong coupling; detailed investigations have revealed no small parameter in which we can do perturbation theory, and no simple calculation that we can perform. From where can we gain some insight? We might ask: where we have seen tubes of confined flux before?

In Type I superconductors, magnetic flux is excluded from the material. This occurs through the appearance of surface currents, which can exist without energy cost due to the absence of any resistance in the material. These currents generate an exactly-compensating magnetic field which cancels any external magnetic field trying to penetrate the material, and instead produces some additional magnetic field outside. This makes it appear that all external magnetic fields are "expelled" from the superconductor. This famous piece of physics is called the "Meissner effect."

In Type II superconductors, however, the situation is a bit more complicated. Flux can indeed penetrate the superconductor in this case, although only in a very specific way. The material becomes nonsuperconducting in a narrow tube running from one side of the material to another, and the magnetic flux threads that tube. The magnetic field, which would have been free to roam in a normal material, is trapped inside "Abrikosov vortices" [9] traversing the superconductor. These vortices carry one or more quanta of flux; in short, they carry an integer charge, $q \in \mathbf{Z}$. *Superconductors confine magnetic flux into quantized vortices.* This looks vaguely similar to the confining flux tubes we have been discussing, so we had better examine this more closely.

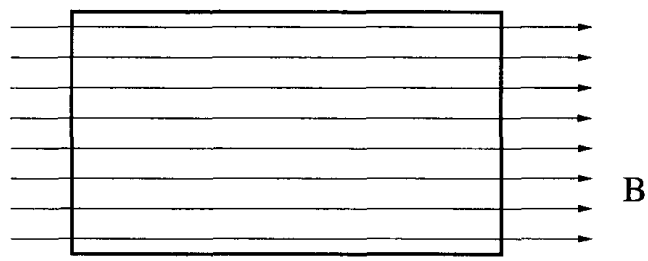


Figure 6: Normal materials can sustain magnetic fields.

How does a superconductor accomplish this? The superconductivity occurs because electrons form Cooper pairs, which are bosons. Let us call the density of these pairs ϕ. Since the pairs carry electric charge 2, ϕ must be complex, and couples to the photon. More specifically, the photon must couple to a conserved current, namely

$$J^\mu = \phi^\dagger \partial^\mu \phi - (\partial^\mu \phi^\dagger)\phi \tag{1}$$

Now suppose that there were a magnetic field attempting to pass through the material. Since the Cooper pairs can flow without resistance, they can respond by creating a compensating current. For instance, suppose we have a long cylinder of material of radius R; let us use cylindrical coordinates r, θ, z. Suppose we attempt to apply a uniform magnetic field $B_z > 0$ along the axis of the cylinder. The Cooper pairs can respond by generating a current J^θ, which can propagate without resistance, at the surface of the cylinder $r = R$. This completely cancels the applied field, reducing the energy density inside the superconductor. It also generates a dipole field outside the cylinder. The field appears to have been "expelled" from the material.

However, the material could also respond in an additional way, and does so in the type II case. In addition to generating a current at $r = R$, it could also generate a current in the opposite direction at $r = r_0 \ll R$, deep within the material. This current, like the current in a solenoid, generates a field in the positive B_z direction, all confined within the region $r < r_0$. This is a magnetic flux tube.

What does ϕ do near this flux tube? Consider a circle of radius $r_1 > r_0$. The integral of the magnetic flux inside this circle, $\int_{r<r_1} B_z \; r \; dr \; d\theta$, should be independent of r_1 if flux is indeed confined. On the other hand, it is also equal to $\oint_{r=r_1} d\theta \; A_\theta$. By cylindrical symmetry, A_θ can be only a function of r. From this we learn that A_θ is a constant for large r. But this

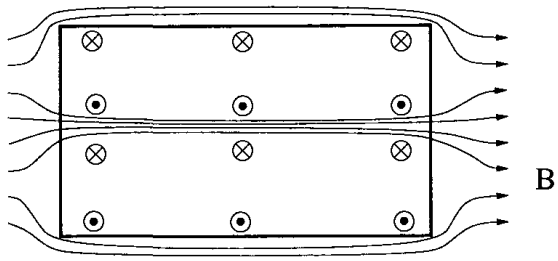


Figure 7: In superconductors, Cooper pair currents (shown into and out of the plane of the paper) are induced, causing the magnetic flux to be expelled or trapped in vortices.

poses problems. The kinetic terms for ϕ itself surely include $\vec{\nabla}\phi \cdot \vec{\nabla}\phi$, where $\nabla_i = \partial_i + iA_i$, and thus $A_\theta^2|\phi^2|/r^2$. If ϕ is a constant v at infinity, then the integral of such a term in the Hamiltonian density is divergent! So this cannot give a finite energy solution. The only way out is to have $\partial_\theta\phi = -iA_\theta\phi$, which can be accomplished if $\phi(r) = ve^{is\theta}$ at large r, where s a real constant. Furthermore, we can avoid a divergent potential energy only if v is at a potential minimum; and at the minimum $v \neq 0$ (or we would not have superconductivity!) But then single-valuedness of ϕ requires that s is an integer. Therefore this approach only works if $A_\theta = s \in \mathbf{Z}$, and thus if $\int B_z\, rdrd\theta$ is an integral multiple of a fundamental flux quantum.

From Eq. (1), we see that J^θ is now nonzero; as advertised, the flux is of necessity enclosed by a current of Cooper pairs. Furthermore, because the phase of ϕ is winding as we go once around in θ, the radial derivatives of ϕ will be ill-defined at $r = 0$ unless ϕ has a zero there. Thus we have $\phi = ve^{is\theta}f(r)$, where $f(0) = 0$ and $f(r \to \infty) \to 1$, and s an integer. The material becomes nonsuperconducting at the vortex core, paving the way for the magnetic field to pass through unobstructed.

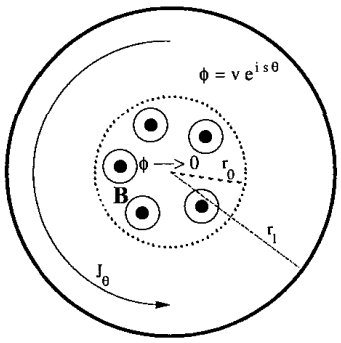


Figure 8: A flux tube of radius r_0; the phase of ϕ winds as one circles the core, in which the magnetic flux is mapped and $|\phi| < v$.

This configuration, with quantized magnetic flux and a zero for ϕ at its center, and a winding of A_θ and a corresponding winding of the phase of ϕ outside its core, is the Abrikosov vortex. Let us consider the topology associated with this vortex. We have a $U(1)$ gauge group, under which ϕ is charged. When the vacuum expectation value of ϕ is nonzero,

the $U(1)$ group is broken spontaneously; gauge transformations will rotate the phase of $\langle\phi\rangle$. [However, remember that gauge transformations are not real symmetries! Therefore, unlike the case of spontaneously broken global symmetries, we do not have a continuous set of physically distinct vacua and associated Nambu-Goldstone bosons. Instead we will get a massive photon!] To make a magnetic flux tube, it must be that as we traverse a circle around the flux tube in space, the phase of the field ϕ makes a closed loop inside the $U(1)$ group. We may think of this as a map from a circle in space to a closed loop in the broken gauge group. Such a map may wind s times around the $U(1)$ as we make a single circle in space. In short, the topology of such maps, given by the first homotopy group of $U(1)$, is the group $\pi_1[U(1)] = \mathbf{Z}$. Every element in the group is labelled by an integer, the winding number s.

To round out the story, it is a bit more convenient to look at a slightly different system. Instead of studying superconductors — three-dimensional nonrelativistic systems — I will take us on a quick tour of the relativistic version, the "abelian Higgs model". This model has Nielsen-Olesen vortices [10], magnetic flux tubes very similar to those of Abrikosov.

Let us take a photon — a $U(1)$ gauge field — and a charged scalar field ϕ. The action for ϕ must be invariant under local $U(1)$ rotations $\phi \to \phi e^{i\alpha(x)}$, which can only happen if all derivatives of ϕ are covariant, that is, of the form $D_\mu\phi \equiv (\partial_\mu + iA_\mu)\phi$, where A_μ is the photon vector potential. In particular, the kinetic term for ϕ must be of the form

$$(D_\mu\phi)^\dagger D^\mu\phi \ .$$

There can also be a potential for ϕ, but gauge invariance requires it be a function only of $\phi^\dagger\phi$. In addition we should add the action for the photon. The action is thus of the form

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D^\mu\phi - V(\phi^\dagger\phi) \ .$$

The potential V may have its minimum at $\phi^\dagger\phi = 0$. In this case the vacuum of the theory is much like the one we live in; the photon is massless, propagates at maximum speed, and generates a long-range force. Magnetic and electric fields are related by a symmetry; both fall off as $1/r^2$ from magnetic and electric point charges.

However, the potential might instead have its minimum at $\phi^\dagger\phi = |v|^2 \neq 0$. Now the physics is very different. First, the photon is now massive. To see this, consider small fluctuations of electric fields A_μ for fixed $\phi = ve^{i\sigma}$. The Lagrangian for these modes is

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} - |v|^2(A_\mu A^\mu).$$

A massive photon, which can be brought to rest, must have three polarization states ($J_3 = 1, 0, -1$) unlike a photon which has only two, $J_3 = \pm 1$. Where does this extra state come from? It comes from σ, the phase of ϕ! Let us see this; if we write $\phi = ve^{i\sigma(x)}$ the Langrangian density now becomes

$$-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} - |v|^2(\partial_\mu\sigma + A_\mu)^\dagger(\partial^\mu\sigma + A_\mu)$$

from which we see that σ and A_μ mix. We cannot think of them any longer as separate fields, and thus σ and A_μ together form a massive, three-polarization-state spin-one particle. (If we like, we can use a gauge transformation to set $\sigma = 0$ and absorb it into A_μ, but this merely puts the degree of freedom of σ into A_μ. It will not always be useful to do this.) This is the Higgs mechanism, discovered by Anderson (always remember that condensed matter physicists have much to teach us) and then rediscovered by many others independently, including Higgs.

Finally, we still have the magnitude of ϕ. Writing $\phi = v + \delta\phi$, we can quickly see from the Lagrangian that $\delta\phi$ acts as a neutral, massive field. I will leave this as an exercise. This means *the theory has a mass gap!* There are no massless modes and no long-range forces.

Now, what happens to electric fields in this context? Suppose I put an electric charge at the origin. The equation of free electrostatics

$$\nabla^2 A^0 = g^2 \delta(x)$$

whose solution is the usual $1/r$ electrostatic potential, is now modified. The new equation is

$$[\nabla^2 + (gv)^2] A^0 = g^2 \delta(x).$$

The solution to this equation is the Yukawa potential for a massive field with mass $m_\gamma = gv$, $V(r) \propto e^{-m_\gamma r}/r$. The electrostatic field falls off exponentially rapidly at distances larger than the inverse of m_γ. *Electric fields are screened!.*

What about magnetic fields? We cannot expel magnetic fields from an infinite system, but we can make currents, just as in superconductors, from the charged scalar ϕ, and use them to confine magnetic flux. Since the photon is massive, it is energetically preferable for the magnetic field to be localized in tubes where ϕ shrinks to zero and the photon is lighter than m_γ. On the other hand, the presence of the magnetic field in a confined region requires, as we saw, that the phase of ϕ wind an integer number of times around the center of the vortex. Classical solutions to the above equations satisfying these conditions can be found; they are called Nielsen-Olesen vortices. Their tensions can be calculated, and are proportional to $1/g^2$. Thus, *magnetic flux is confined!* The topological analysis that we did for the Abrikosov vortex — that the charges of these vortices is given by the first homotopy group of $U(1)$, the group $\pi_1[U(1)] = \mathbf{Z}$ — goes through here as well, without alteration.

Magnetic flux tubes can arise in other gauge groups as well when they are broken via the Higgs mechanism. If we have a gauge group G broken down to a smaller gauge group H (which might be the identity, as in the example above) we will get magnetic flux tubes if $\pi_1(G/H)$ is not trivial. For example, if we have the group $SU(N)$, and it is broken down to nothing, then there are no flux tubes; $SU(N)$ is simply connected, so all closed curves on it can be shrunk down to nothing, and all of its homotopy groups are trivial. However, if we break $SU(N)$ down to its center $\mathbf{Z}_N$, then since $\pi_1(G/H) = \pi_0(H)$ if G is simply connected, and since $\pi_0(H)$ is the number of distinct components of H, we have simply $\pi_1(SU(N)/\mathbf{Z}_N) = \mathbf{Z}_N$. Magnetic flux tubes are generated, and they carry a charge in $\mathbf{Z}_N$, the integers modulo N [2]. [As an example, consider $SU(2)$. The matrices $\mathrm{diag}(e^{i\alpha}, e^{-i\alpha})$ are in $SU(2)$; for $\alpha = 0$ and π they are in the center. The path from $\alpha = 0$ to $\alpha = 2\pi$ is a closed path in $SU(2)$, but the path from $\alpha = 0$ to $\alpha = \pi$ is not closed. However, in $SU(2)/\mathbf{Z}_2$, the matrices with $\alpha = 0$ and $\alpha = \pi$ are identified, so the second path is also closed and forms the nontrivial element of $\pi_1(SU(2)/\mathbf{Z}_2) = \mathbf{Z}_2$.]

3 Duality, Sources and Fluxes

We have now seen that there is a beautiful example of confinement of magnetic flux, caused by condensation of an electrically charged field. The Maxwell equations are symmetric under exchange of electric and magnetic fields, and so we might naively hope that there are examples of electric flux confinement caused by condensation of magnetically charged fields. But the Yang-Mills equations do not have this symmetry classically, so this seems somewhat problematic for the nonabelian gauge theories which we know confine. Nonetheless, something like this idea does indeed work quantum mechanically, at least in part.

Before we begin to study it, we need to complete our classification of fluxes and non-dynamical sources, of both electric and magnetic type. This is necessary so that we can put electric-magnetic duality to proper use.

3.1 Electric Sources and Fluxes

Let us review what we have learned up to now about electric fluxes and sources, a bit more formally. Consider a pure gauge theory with gauge group G. Suppose we have a source — an infinitely massive, static, electrically charged particle — in a representation R of G. If we surround the source with a large sphere, what characterizes the flux passing through the sphere? If G is $U(1)$, the flux measures the electric charge directly. However, in non-abelian gauge theories the gauge bosons carry charge. Since there may be a number (varying over time) of gauge bosons inside the sphere, the representation under which the charged objects in the sphere transform is not an invariant. But, by definition, the gauge bosons are neutral under the discrete group C_G, the center of G. It follows that the charge of R under the center *is* a conserved quantity, and that the total flux exiting the sphere carries a conserved quantum number under C_G.

Electric sources and fluxes in pure gauge theories carry a conserved C_G quantum number. If the gauge group confines, then the confining electric flux tubes will also carry this quantum number.

If the theory also contains light matter charged under C_G but neutral under a subgroup C_m of C_G, then the above statements are still true with C_G replaced with C_m. For example, if we take $SU(N)$ with light fields in the $\mathbf{N}$ representation, then C_m is just the identity, reflecting the fact that all sources can be screened and all flux tubes break. If we take $SO(10)$ with fields in the $\mathbf{10}$, then the center $\mathbf{Z}_4$ is replaced with spinor-number $\mathbf{Z}_2$. Sources in the $\mathbf{10}$ will be screened and have no flux tube between them, while sources in the $\mathbf{16}$ or $\overline{\mathbf{16}}$ will be confined by a single type of flux tube.

3.2 Magnetic Sources and Fluxes

Before discussing the magnetic case, I review some basic topology. [The presentation which follows is overly naive, though it serves for present purposes. A more rigorous story requires a study of the relevant fiber bundles.] The p-th homotopy group of a manifold $\mathcal{M}$, $\pi_p(\mathcal{M})$, is the group of maps from the p-sphere into $\mathcal{M}$, where we identify maps as equivalent if they are homotopic (can be continuously deformed into one another) in $\mathcal{M}$. All we will need for

present purposes are the following examples. Suppose a Lie group G has rank r, so that its maximal abelian subgroup is $U(1)^r$; then

$$\pi_2[G] = \mathbf{1} \quad \Rightarrow \quad \pi_2\left[G/U(1)^r\right] = \pi_1[U(1)^r] = \mathbf{Z} \times \mathbf{Z} \times \cdots \times \mathbf{Z} \equiv [\mathbf{Z}]^r \ . \tag{2}$$

Similarly,

$$\pi_1[G] = \mathbf{1} \quad \Rightarrow \quad \pi_1\left[G/C_G\right] = \pi_0[C_G] = C_G \ . \tag{3}$$

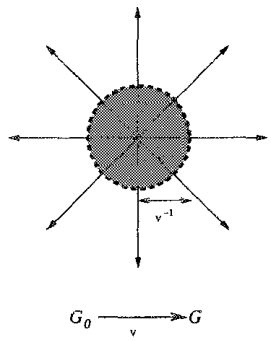


Figure 9: A magnetic monopole soliton of size v^{-1}.

We will need to investigate both monopole solitons and string solitons below. The classic monopole soliton is that of 't Hooft and of Polyakov, which arises in $SU(2)$ broken to $U(1)$; in this case the important topological relation is $\pi_2[SU(2)/U(1)] = \pi_1[U(1)] = \mathbf{Z}$. This leads to a set of monopole solutions carrying integer charge. Note that the stability of, for example, a single monopole which has charge two against decay to two monopoles, each of charge one, is determined not by topology but by dynamics. The situation is similar for the Nielsen-Olesen magnetic flux tube of the abelian Higgs model; here the relevant topological relation is $\pi_1[U(1)] = \mathbf{Z}$. This again leads to solutions with an integer charge, whose stability against decay to minimally charged vortices is determined dynamically.

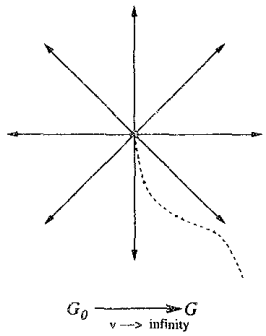


Figure 10: A pointlike Dirac monopole, with its unphysical Dirac string.

More generally, if we have a *simply connected* gauge group G_0 which breaks to a group G at a scale v, there will be solutions to the classical equations in the form of magnetic

monopoles carrying a quantum number in $\pi_2[G_0/G]$ (see, for example, [11].) These will have mass [radius] proportional to v $[1/v]$. Now imagine that we take $v \to \infty$. In this limit the gauge group G_0 disappears from the system. The monopoles become pointlike and infinitely massive; their only non-pointlike feature is their (nonphysical) Dirac string, which stems from our having discarded G_0, and which carries a quantum number in $\pi_1[G]$. In short, the solitonic monopoles become fundamental Dirac monopoles in this limit. Note that since $\pi_2[G_0/G] = \pi_1[G]$, the charges carried by the solitonic monopoles and their Dirac monopole remnants are the same. At this point, we can forget about G_0, which is only relevant at infinitely high energies. Since the Dirac monopoles are heavy, we may use them as magnetic sources in a theory with gauge group G.

Let's further suppose that the gauge group G is broken completely at some scale v'. In this case no Dirac strings can exist in the low-energy theory, and so the monopoles allowed previously have seemingly vanished. However, solitonic magnetic flux tubes, carrying charges under $\pi_1[G]$, will be generated; they will have tension [radius] of order v'^2 $[1/v']$. Their $\pi_1[G]$ quantum numbers are precisely the ones they need to confine the $\pi_1[G]$-charged Dirac monopole sources of the high-energy theory. Thus, when G is completely broken, the Dirac monopoles disappear because they are confined by flux tubes.

Magnetic sources and fluxes in pure gauge theories carry a conserved $\pi_1[G]$ quantum number. If the gauge group is completely broken, then the confining magnetic flux tubes will also carry this quantum number.

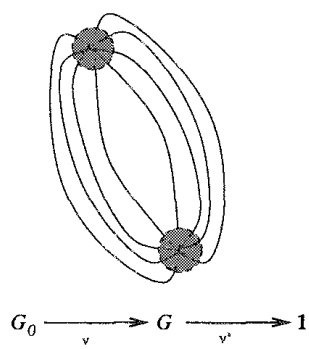


Figure 11: Confined monopole solitons in a theory with flux tubes.

3.3 Conclusion

There are two questions of interest, one of which is answered by this classification.

The first is: "*Can* a particular theory have confinement of flux into stable [or at least metastable] flux tubes?" The answer is "Only if there exist charged sources which cannot be screened: namely, if there is part of C_G under which all matter [or at least, the light matter] fields are neutral." This answer is group theoretic, and thus structural, or "kinematical."

The second is: "*Does* a particular theory have confinement?" This question cannot so easily be answered. It is a "dynamical" question, and it depends on the theory in much more detail. In general we cannot answer it in nonsupersymmetric theories. Fortunately,

we can sometimes answer it in supersymmetric theories, and we will spend the next lecture exploring one of them.

4 Duality and Confinement

4.1 Duality and QCD

What is duality? Let me propose an example of a duality which you may not normally think of as one, and I will gradually convince you that it is. Namely, think of what happens in real QCD. At high energy it is well described as weakly interacting quarks and gluons. At low energy, where the quarks and gluons are strongly interacting, it is well described as a weakly coupled theory of pions, kaons, and other hadrons. This latter theory is called the "chiral Lagrangian". I claim QCD and the chiral Lagrangian are dual.

The immediate objection is that the chiral Lagrangian is just an effective theory for the long-distance degrees of freedom of QCD — a description of quasiparticles, much as one would use in condensed matter systems with composite objects. But I want to argue strongly that this is too limited a view: that effective theories of this type are just special cases of a much larger story.

How is the chiral Lagrangian constructed? We want to write an effective theory of hadrons π, K, η, *etc.* We write the most general Lagrangian allowed by symmetries; we do not impose renormalizability, since the theory has a natural cutoff Λ of about 1 GeV, and we do not plan to take $\Lambda \to \infty$. There are many unknown constants, but in the low energy limit, when the scale μ of the process is much less than Λ, the theory becomes uniquely fixed. The statement of duality in this case is that the

$$\text{Physics of } \mathcal{L}_{chiral} = \text{Physics of } \mathcal{L}_{QCD} + \text{ corrections of order } \frac{\mu}{\Lambda}$$

Thus the theories describe the same physics in the infrared, though they differ enormously in the ultraviolet.

The power of duality is that we can alter the two theories in ways such that they remain dual. For instance, suppose we consider QCD with three flavors, and set all quark masses equal to zero. The dual chiral Lagrangian is tightly constrained, since the $SU(3) \times SU(3) \times U(1)$ flavor symmetries allow us to constrain $\mathcal{L}_{chiral}$ more completely. Now let us add finite m_s. If $m_s \gg \Lambda$, then we know what to do. (1) The QCD theory with $N_f = 3$ must be matched, in perturbation theory, to the theory with $N_f = 2$. (2) The latter becomes strongly coupled and makes hadrons. We can invent a chiral Lagrangian to which it becomes identical well below the scale Λ. This is not so exciting. But now suppose $m_s \ll \Lambda$. Now it is reversed. (1) The theory first becomes strongly coupled and makes hadrons, and we can invent a chiral Lagrangian which becomes identical to QCD with $N_f = 3$ at scales well below Λ. (2) The effect of m_s on QCD is then correctly described by the chiral Lagrangian, because the two theories become essentially identical *before* the mass of the strange quark becomes important. Since the theories are identical, and since the chiral Lagrangian is weakly coupled in this regime, we are better off using it for describing the physics of the theory. Note that in this regime we have *one* theory with two descriptions, not two theories.

What lessons do we learn? We must distinguish two levels of duality. The first is "Weak Duality," where we have two theories, each with its own set of variables, which become approximately equal below some scale $\hat{\mu}$; for $\mu \ll \hat{\mu}$ the theories become equal. The second is "Strong Duality", where we have one theory, with two sets of variables: two descriptions of a single theory. QCD and the chiral Lagrangian provide an example of the first, for $\hat{\mu} = \Lambda \sim$ 1 GeV, in that the chiral Lagrangian provides a description of hadrons accurate up to effects of order μ/Λ. And the effect of the strange quark mass is closer to an example of the second, since the physics of the transition from three flavors to two occurs at $\mu \ll \Lambda$ where the two theories are essentially identical. A precise example of the latter is given in Burgess and Quevedo 1994 paper [12]; you can also look at [13], or at the author's TASI 2002 lectures.

Let me also list some facts that we do not learn from this example. Generally, two dual descriptions of a single theory are mutually non-local and non-polynomial; the quasi-particles of QCD can be thought of as bound states of a small integer number of quarks and antiquarks, but this is not the general case. Another surprise is that the dual description of a gauge theory need not be a theory of massive bound states but may itself be a gauge theory, with its own gauge symmetry unrelated to that of the original description. Weak duality may hold for two asymptotically free gauge theories, neither of which has a UV cutoff; so $\hat{\mu}$ need not be a cutoff, but simply a scale at which the two theories start to look similar. Alternatively, *both* descriptions of a theory may be nonrenormalizable, and both then have an ultraviolet cutoff unrelated to $\hat{\mu}$.

4.2 Seiberg duality for $SO(N)$ gauge theories

Let us learn a bit about how this works in supersymmetric $SO(N)$ gauge theory. Gauge bosons $A_\mu = A_\mu^A T^A$, gauginos $\lambda_a = \lambda_a^A T^A$, and auxiliary fields $D = D^A T^A$ are all in the adjoint representation, and the kinetic terms are the minimal ones

$$S_{gauge} = \frac{1}{g^2} \int d^4x \ \mathrm{tr} \left[-\frac{1}{4} F^2 + i\bar{\lambda} \not{D} \lambda + \frac{1}{2} D^2 \right]$$

where

$$F_{\mu\nu} = F_{\mu\nu}^A T^A = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu] = \left(\partial_\mu A_\nu^A - \partial_\nu A_\mu^A + f^{ABC} A_\mu^B A_\nu^C \right) T^A,$$

$D_\mu \lambda_a = \partial_\mu \lambda_a + i[A_\mu, \lambda_a]$, and $D^2 \equiv \sum_A |D^A|^2$. Let's add chiral superfields Q_r and $r = 1, \ldots, N_f$, in the fundamental representation of $SO(N)$. Each Q^r contains a squark Q^r, a quark ψ^r, and an auxiliary field F^r. These fields have kinetic terms

$$S_{kin} = \sum_r \int d^dx \ \left[D_\mu Q_r^\dagger D^\mu Q_r + i\bar{\psi}_r \not{D} \psi_r + F_r^\dagger F_r + \lambda \bar{\psi}_r Q_r + \bar{\lambda} \psi_r Q_r^\dagger + Q_r^\dagger D Q_r \right]$$

where the contraction of gauge indices is in each case unique: for example in the term $Q_r^\dagger D Q_r$ the indices are contracted as

$$(Q_r^\dagger)_{\bar{j}} D^A (T^A)_i^{\bar{j}} (Q_r)^i$$

Finally there are interactions encoded in a "superpotential", a function of the chiral fields Q_r only (not their conjugates)

$$S_{int} = \int d^dx \ \left[-\frac{1}{2} \frac{\partial^2 W(Q)}{\partial Q^r \ \partial Q^s} \psi^r \psi^s + \frac{\partial W(Q)}{\partial \phi^r} F^r \right]$$

plus its complex conjugate.

Note the scalar potential $V(\phi)$, obtained by integrating out the D^A and F_i is a sum of squares! This means the vacua of the theory all have $V = 0$, which occurs in our theory only if

$$D^A = \sum_i Q^{i\dagger} T^a Q^i = 0$$

for each a and

$$F_i^\dagger = \frac{\partial W}{\partial Q^i} = 0$$

for each i.

Let us now consider a special case of Seiberg duality. Seiberg claims [14] that if we take a gauge theory with group $SO(N)$ and N_f squarks Q^i (and quarks too) in the $\mathbf{N}$ representation, and $W(Q^i)$=0, then this theory has a dual description as an $SO(N_f - N + 4)$ gauge theory. This dual theory has N_f squarks q_i (and quarks) in the $\mathbf{N_f} - \mathbf{N} + \mathbf{4}$ representation, along with $N_f(N_f + 1)/2$ scalars M^{ij} (and their fermionic superpartners) which are neutral under the gauge group. These neutral particles interact with the charged matter because the dual superpotential $\tilde{W}(q_i, M^{ij}) = hM^{ij}q_iq_j$ (we will set the coupling $h = 1$ since its value will play no role below.)

If the two theories are dual, then it must be true that their gauge-invariant operators are the same. (For example, in QCD $\psi_i\gamma^\mu\psi^j$ is a gauge-invariant operator, corresponding to the operator which creates the ρ meson in the chiral Lagrangian.) Here the operator Q^iQ^j is the same operator as M^{ij}. The operator q_iq_j (by equations of motion involving $\partial\tilde{W}/\partial M$) equals a derivative of $M^{ij\dagger}$; this is dual to a derivative of $(Q^iQ^j)^\dagger$.

Seiberg's claim is that these two theories are weakly dual; they are different in the ultraviolet but the same in the infrared. This conjecture has been tested in an immense number of ways; few people still doubt its validity. But note that it still has not been proven!

Now, how did Seiberg come to make this claim? Let's review how things were done in QCD. (1) We identified the correct low-energy degrees of freedom using experiment. (2) We wrote the most general Lagrangian consistent with the symmetries. (3) We checked the result by comparing with experiment.

Here, the situation is different. (1) Instead of experiment, Seiberg used raw brilliance to guess the low-energy degrees of freedom. (2) The next step is the same as before, except that supersymmetry adds more constraints, and allows the Lagrangian to be specified in much greater detail. (3) The ansatz is checked not by experiment but by a host of algebraic and physical consistency conditions; these include 't Hooft anomaly matching and a check that the duality remains consistent under all relevant perturbations of the theory.

For instance, let us change the superpotential to $W = mQ^1Q^1$, which reduces N_f to $N_f - 1$. In the dual theory this means adding mM^{11} to the superpotential. The condition

$$0 = \frac{\partial\tilde{W}}{\partial M^{11}} = \frac{\partial(M^{ij}q_iq_j + mM^{11})}{\partial M^{11}} = q_1q_1 + m$$

which means one of the dual squarks q_1 must have an expectation value, breaking the gauge group by one color. In the process the squark q_1 (and its quark) are absorbed into the

massive gauge bosons (and their fermionic partners) reducing the number of flavors in the dual theory to $N_f - 1$. Thus we end up with $SO(N_f - N_f + 3) = SO([N_f - 1] - N_c + 4)$ with $N_f - 1$ flavors, which is indeed the dual of $SO(N)$ with $N_f - 1$ flavors. So we have changed the orginal theory, and the change in the dual theory is such that the two theories remain consistent with Seiberg's conjecture.

This duality is consistent with something else we know. The case $N = 2$, $N_f = 0$ is pure electromagnetism (since $SO(2) = U(1)$) and Seiberg tells us this should be dual to $SO(0 - 2 + 4) = SO(2)$ with $N_f = 0$. This is just the electric-magnetic duality of Maxwell's equations! So we are looking at a nonabelian generalization of electric-magnetic duality!

Now, I want to concentrate our attention on $SO(8)$, just to have a specific example to study. Let's explore a few cases. Suppose $N_f = 20$, so that the dual theory has $SO(16)$ gauge group. In this case the $SO(8)$ theory has a positive beta function, and a Landau pole appears at some high scale Λ. The dual theory is asymptotically free. The two theories become similar, according to Seiberg, at low energy, where the $SO(8)$ theory becomes free. Thus the $SO(16)$ gauge theory, which is weakly coupled at high energy, and becomes strongly coupled at low energy, is well described in the infrared by the $SO(8)$ theory! And this $SO(8)$ is not a subgroup of $SO(16)$, but is an independent gauge group. The $SO(8)$ gluons are not a subset of the $SO(16)$ gauge bosons, but are quasiparticles of the $SO(16)$ gauge theory. This is a strange effective theory indeed!

Consider instead $N_f = 15$. In this case both the $SO(8)$ theory and its $SO(11)$ dual are asymptotically free, and both become strongly coupled in the infrared. The interpretation in this case is that both theories flow to the same nontrivial infrared fixed point — a conformally invariant theory, with nontrivial critical exponents. (This is the kind of theory which describes phase transitions in condensed matter.) The low energy fixed point does not in general have a simple Lagrangian description.

How about $N_f = 7$? In this case the dual $SO(3)$, with seven fields in the **3**, has a positive beta function; thus it becomes weakly coupled in the infrared. It serves as a good effective theory for the $SO(8)$ theory, which becomes strongly coupled in the infrared.

For $N_f = 6$ the dual theory is $SO(2)$ — electromagnetism — with 6 selectrons and electrons. Again the $SO(8)$ theory becomes strongly coupled in the infrared, and the abelian gauge theory is a good effective theory for it at long distances.

Finally, consider $N_f = 5$. In this case there is no dual gauge group. There are just fields M^{ij} and q_j. It turns out the q_j are, in this case, just baryons of $SO(8)$! The dual theory is thus an ordinary sigma model, free in the infrared, with bound states of the $SO(8)$ matter fields as its basic degrees of freedom. This is very much like the chiral Lagrangian of QCD.

Thus, we see that Seiberg duality draws together electric-magnetic duality, and the relationship between QCD and the chiral Lagrangian, into a unified duality structure.

4.3 Confinement or not?

Our goal in this section is to convince ourselves that $SO(8)$ with $N_f = 5$ is a confining theory, while the theories with $N_f = 6$ and 7 are not confining. To do so, we must introduce a source which is not screened either by the gauge fields or by the matter fields Q^j in the **8** representation. Fortunately $SO(8)$ (or rather, its double cover, $Spin(8)$) has *spinor*

representations. These happen to be 8 dimensional also, and are usually referred to as $\mathbf{8_s}$ and $\mathbf{8_c}$ (c for "conjugate spinor.") And $SO(8)$ has a $\mathbf{Z}_2 \times \mathbf{Z}_2$ center, under which the gauge fields are neutral — they transform as $(++)$ — the Q^i transform as $(+-)$, and the $\mathbf{8_s}$ and $\mathbf{8_c}$ transform as $(-+)$ and $(--)$. Since the spinors are odd under the first $\mathbf{Z}_2$ (whose nontrivial element is what we earlier called h_0) while the gauge and matter fields are even, no combination of gauge and matter fields can ever completely hide a spinor source. The theory will always know the source is there; even though it can't distinguish an $\mathbf{8_s}$ from an $\mathbf{8_c}$, it can distinguish between a spinor source (odd under h_0) and no source (even under h_0)). Therefore, it is possible that the theory has a stable confining flux tube. To look for it, all we have to do is introduce a spinor source and an antispinor source at a distance R, and compute $V(R)$.

Unfortunately the $SO(8)$ theory is very strongly coupled at long distance, so we can't calculate $V(R)$ directly. But fortunately the dual $SO(N_f - N + 4)$ theory *is* weakly coupled in the infrared for $N_f = 5, 6, 7$, so we should be able to do the computation using dual variables! The only problem is that we haven't talked yet about how to introduce the spinor source into the dual theory. What is this spinor source from the dual point of view?

Is it a magnetically charged source?

4.4 Spinor sources from the dual point of view

Yes! To see this, we need an extension of Seiberg duality [15]. As our original theory, let us take $Spin(8)$ with N_f fields Q^i in the $\mathbf{8}$ and one field P in the $\mathbf{8_s}$. We take $W(Q^i, P) = 0$. The dual theory is then not $SO(N_f - 4)$ but $SU(N_f - 4)$, with a field S in the symmetric tensor representation of the gauge group, N_f fields q_i in the antifundamental, singlets M^{ij} as before and one more gauge singlet T. As before Q^iQ^j is dual to M^{ij}, while PP is a gauge singlet dual to T. The superpotential

$$W = M^{ij} q_i S q_j + T \det S$$

is nonrenormalizable; but so is $\mathcal{L}_{chiral}$, so this should not disturb us.

We can use this theory to find a spinor source. If we make P massive, and take the mass to infinity, then a P particle will become exactly what we want: a nondynamical source in the $\mathbf{8_s}$ representation. Again, the theory will not be able to distinguish an $\mathbf{8_s}$ from an $\mathbf{8_c}$. Gauge bosons and Q^i particles, passing nearby, will constantly change the representation in which it appears to transform. However, they can never change the fact that it is odd under the first $\mathbf{Z}_2$ in the $\mathbf{Z}_2 \times \mathbf{Z}_2$ center. And so this source will be characterized by a $\mathbf{Z}_2$ quantum number.

Thus we should look for a monopole in the dual theory which is odd under some $\mathbf{Z}_2$ transformation. And we can find it! When we make the original superpotential $W = mPP$, we make the dual superpotential

$$\tilde{W} = MqSq + T \det S + mT$$

from which $\partial\tilde{Q}/\partial T$ implies $\det S = m$. This means S has a nonzero expectation value. Combining this with the other terms in the scalar potential we find

$$\langle S \rangle = m^{1/[N_f - 4]} \mathbf{1}$$

where **1** is the unit matrix. This expectation value breaks $SU(N_f-4)$ to $SO(N_f-4)$. This looks quite familiar. If $N_f=6$, this is the same as $SU(2)$ broken to $U(1)$ — the classic case of the 't Hooft-Polyakov monopole. The topology behind this was

$$\pi_2[SU(2)/SO(2)] = \pi_1[SO(2)] = \mathbf{Z}\ .$$

For $N_f > 6$ we have

$$\pi_2[SU(N_f-4)/SO(N_f-4)] = \pi_1[SO(N_f-4)] = \mathbf{Z}_2$$

and so for $N_f > 6$ we have a monopole carrying a $\mathbf{Z}_2$ charge. Clearly, this monopole is the dual of our spinor! And its mass goes to infinity as $m \to \infty$, making it into a nondynamical *magnetic* source.[1]

In the $N_f = 6$ case, we can now easily see that there is no confinement of spinors. We do so by considering what happens to magnetic monopoles in the dual theory, which is an abelian theory with massless charged matter. In such a theory, the coupling constant goes to zero logarithmically: $\alpha(r) = c/\ln r$, where c is a constant. This means that the potential energy between monopoles in a weakly coupled abelian theory

$$V(R) = \frac{1}{\alpha(R)}\frac{1}{R} \sim \frac{\ln R}{R}$$

goes to zero at infinity, as does the force. Thus these monopoles are not confined, so we can conclude that in the original description, spinor sources are not confined, and there are no flux tubes.

For $N_f = 7$, the situation is the same. We now have a nonabelian $SO(3)$ dual theory, but because its beta function is positive, its coupling again goes to zero as $\alpha(r) = c/\ln(r)$. As before, the potential between monopoles goes to zero as $\ln R/R$.

But what about $N_f = 5$? In this case there is no dual gauge group, either before or after $m \neq 0$. There therefore can be no 't Hooft-Polyakov monopole! How do we study this theory? Well, we know how the $N_f = 7$ case works. What happens to the unconfined spinors if we give masses to Q^7 and Q^6?

In other words, let's take $Spin(8)$ with 7 Q^i and P, and make $W = (Q^i, P) = mPP + m'(Q^6Q^6 + Q^7Q^7)$. In the dual $SU(3)$ gauge theory, this causes $\langle S\rangle$ to be nonzero *and* $\langle q_6q_6\rangle = \langle q_7q_7\rangle = -m'$. This breaks the gauge group completely! When $SU(3)$ breaks to $SO(3)$, monopoles are formed. But when $SO(3)$ is broken via the Higgs mechanism, magnetic monopoles are confined! And they are confined by magnetic flux tubes, which form because $\pi_1[SO(3)] = \pi_1[SU(2)/\mathbf{Z}_2] = \mathbf{Z}_2$.

Thus, we have found a nonabelian example of the dual Meissner effect. The confinement of $\mathbf{Z}_2$ monopoles by $\mathbf{Z}_2$ magnetic flux tubes in the dual theory is interpreted as confinement of spinors by $\mathbf{Z}_2$ electric flux tubes in the original $SO(8)$ theory. And so we have achieved our goal: we have shown $SO(8)$ with $N_f = 5$ is confining, and seen that it occurs through the dual of the Meissner effect, by condensation of magnetically charged scalar fields.

[1]The question of why we get $\mathbf{Z}$ and not $\mathbf{Z}_2$ for $N_f = 6$ has to do with physics which is not essential here.

4.5 Conclusion

However, we must not celebrate overly much. In addition to the scales m and m' in our theory, there is a scale $\hat{\mu}$ where the original and the dual theory, which are weakly dual, become equivalent. The dual theory of $SU(3)$ with which we began our analysis is asymptotically free; in the far ultraviolet it is weakly coupled and is not the same as the weakly coupled $SO(8)$ theory. For our analysis to be valid, it must be that $\hat{\mu} \gg m \gg m'$, because otherwise our dual description really wasn't describing quite the same theory. But in fact, what we want to know is whether $SO(8)$ with $N_f = 5$ with weak coupling in the ultraviolet confines in the infrared. In other words, we want m, m' much bigger than the strong-coupling scale Λ_8 of $SO(8)$. Unfortunately, it can't be that $\hat{\mu}$ is bigger than Λ_8. The dual description of the $SO(8)$ theory can't be weakly coupled unless the $SO(8)$ description is strongly coupled. So we want to study the region $m, m' \gg \Lambda_8 \geq \hat{\mu}$, but our analysis is not valid there. Have we lost everything?

Not everything. Supersymmetry is powerful; it ensures that as we vary m and m' there can be no phase transitions. (See Seiberg's work on holomorphy.) Consequently, the *fact* of confinement survives from our analysis. However, the mechanism — a semiclassical dual Meissner effect — does not survive. Our analysis must be taken into a strong-coupling regime of the dual theory, where we cannot, unfortunately, check its validity. So, we have shown $SO(8)$ with $N_f = 5$ confines, but we have not shown there is a simple description of the *dynamics* of confinement.

Now, earlier I said that the question "*Can* a theory confine?" is structural, or kinematical; whereas the question "*Does* a theory confine?" is dynamical. But we have seen here that duality ties together dynamics, topology and group theory together in an interesting way: the fact that a mass for P leads the dual SU group to break to SO is what leads to monopoles, and the masses for Qs break the dual SO and give us confinement. So perhaps we should refine this thinking. Perhaps "*Can* a theory confine?" is a question of semiclassical structure, "*Does* a theory confine" is a question of *nonperturbative* structure, and only "*How* does a theory confine" is a truly dynamical question. If this is true, then perhaps we should, even in nonsupersymmetric theories, be able to predict which theories confine, break their chiral symmetries, *etc.*, and which do not, without too much understanding of the underlying dynamics. This would vastly simplify our task! Let us hope that it is true.

References

[1] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).

[2] G. 't Hooft, Nucl. Phys. B **190**, 455 (1981), Nucl. Phys. B **153**, 141 (1979), Nucl. Phys. B **138**, 1 (1978).

[3] R. Donagi and E. Witten, Nucl. Phys. B **460**, 299 (1996) [hep-th/9510101].

[4] M. J. Strassler, Nucl. Phys. Proc. Suppl. **73**, 120 (1999) [hep-lat/9810059].

[5] M. Wingate and S. Ohta, Phys. Rev. D **63**, 094502 (2001) [hep-lat/0006016].

[6] B. Lucini and M. Teper, hep-lat/0103027.

[7] B. Lucini and M. Teper, Phys. Lett. B **501**, 128 (2001) [hep-lat/0012025].

[8] C. Vafa and E. Witten, Nucl. Phys. B **431**, 3 (1994) [hep-th/9408074].

[9] A. A. Abrikosov, Sov. Phys. JETP **5**, 1174 (1957) [Zh. Eksp. Teor. Fiz. **32**, 1442 (1957)].

[10] H. B. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973).

[11] S. Coleman *Aspects of Symmetry, Selected Erice lectures* (Cambridge University Press, Cambridge, 1985)

[12] C. P. Burgess and F. Quevedo, Nucl. Phys. B **421**, 373 (1994) [arXiv:hep-th/9401105].

[13] A. Kapustin and M. J. Strassler, JHEP **9904**, 021 (1999) [hep-th/9902033].

[14] N. Seiberg, Nucl. Phys. B **435**, 129 (1995) [hep-th/9411149].

[15] P. Pouliot and M. J. Strassler, Phys. Lett. B **370**, 76 (1996) [arXiv:hep-th/9510228].

CHAIRMAN: M. STRASSLER

Scientific Secretaries: V. Markov, M. Matveev

DISCUSSION I

- *Strassler:*

Let me put up the main results. Just to review and try to put everything together: we started off with the statement that Yang-Mills has confinement of flux, and that the various properties of QCD with heavy or light or massless quarks has everything to do with this fact, that QCD with heavy quarks has confinement because Yang-Mills has confinement of flux.

I then pointed out there are two separate questions one might ask about a theory: "Does a particular theory have confinement" – does it have flux tubes, is there a linear potential between charged sources – is a question of dynamics. But before that, one should ask "Can a theory have confinement" – is it even possible that a theory could have flux tubes – which is a question that goes to the structure of the theory. In particular, in QCD with light quarks, the presence of light matter in the triplet representation of SU(3) already ensures that the theory cannot possibly have stable flux tubes. And that is a structural question, not a dynamical question, so you should ask it first.

Then I spent a little time looking at the structural issue with electric confinement. You have a gauge group G; there are various electric sources you could imagine putting into the theory, in various representations of the group. But these sources are classified by their representation under the centre of the gauge group. And you can only have flux tubes classified by a subgroup S of the centre, and S is defined as the subgroup which leaves invariant all of the matter representations in the theory. (Part of the matter is already gluons, which are already only invariant under the centre.) Note that if you only want approximately stable flux tubes, then you can replace "matter" with "light matter".

The dynamics in the case of magnetic flux confinement we investigated in the context of an abelian Higgs model, or a nonabelian Higgs model. We had a group G, and a charged scalar in some representation of the gauge group. When the scalar gets an expectation value, it breaks the group down to a subgroup H, generating massive gauge bosons, screened electric fields, and confined magnetic flux via the Meissner effect. But there is also a structural issue in the magnetic case, which is that the magnetic sources that you can use to investigate the theory are classified by the possible Dirac monopoles in the theory, which are in turn classified by the possible Dirac strings. These are given by the first homotopy group of G. The flux tubes themselves come in the first homotopy group of G/H.

Finally, what is electric confinement really about? Well, having understood structure and dynamics in the magnetic case, and having understood structure in the electric case, we might hope that electric confinement could be understood by taking the dynamical magnetic picture of confinement and acting on it with electric-magnetic duality. A magnetically charged scalar, whatever that means, might condense, generating electric flux tubes via a dual Meissner effect.

- *Alessindro:*

If you could calculate exactly a Wilson loop in Yang Mills theory and if you found pure exponentiation of the area law, to what extent would you be sure that the theory is strictly confining?

- *Strassler:*

What we will have shown in that case is that non-dynamical heavy quarks have a potential energy which grows linearly with the distance between them. That is also a demonstration that if I physically produce a pair of quarks and try to pull them apart, it would take infinite energy. That is one of form of confinement. The other thing you could ask is "can I have isolated quarks." Put an isolated quark here; that is like having half a Wilson loop; the other half is off at infinity, so the loop has infinite area, which is equivalent to saying the state with the isolated quark has infinite energy. In theories that have screening, such as QCD, where you can actually produce quark pairs, you can never actually have an area law. It is much harder to say what is going on in such a theory. So in Yang Mills theory you have the possibility – and again, this is a purely structural issue – that there might be strict confinement, or a Wilson loop area law, but in QCD, before you even start, you know this cannot happen in QCD with light quarks, because any source you use can just be screened by taking a light quark and sticking it next to it.

- *Markov:*

The linear rising potential is an effective potential. What is its range of application? Why would you not add a constant term, or some other term, to it?

- *Strassler:*

The question really asked is about asymptotic behaviour. And the reason you know it will be asymptotically linear, if the theory has flux confinement, is the Gauss's law picture that I showed. If there are flux tubes in the theory then the energy will be extensive in the length, up to edge effects which you will not see if the flux tube is very long. What you are really interested in, anyway, is forces, and what you really learned from a linear potential is that the force is constant.

- *Markov:*

Does the constant flux always mean the field strength is constant?

- *Strassler:*

There are certainly possibilities beyond that which I discussed. As you pull the quark and antiquark pair apart, the radius of the flux tube could change, or something. It might not completely deconfine, but its asymptotic behaviour might not be linear. But I do not know of any context in 4d gauge theory where any other behaviour has been seen. It might be an interesting exercise to see if you could write down a theory which had some other behaviour than linear.

- *Maas:*

Does the reasoning about structural issues if a theory can or cannot confine hold on charge if the physical spectrum and the degrees of freedom in your Lagrangian are not the same?

- *Strassler:*

The physical spectrum and the spectrum you write down in the Lagrangian never are the same. When you write the "electron" in the Lagrangian it does not mean that the physical electron is the thing you wrote – the electron gets dressed in some way. How the electron gets dressed is really a question of charge conservation. It does not matter how many photons, electron and positron pairs surround the electron; there is one thing that cannot go away, and that is the charge. So, what is really going on in the structural issue is the question: does my source carry some charge which, no matter how it is dressed by the charged matter and gluons, cannot be removed? And that can only be true if the gluons and light matter are neutral under that particular charge.

- *Englert:*

Let me imagine that I have a theory with massless fermions in a nontrivial representation in the centre. Is it not a necessary effect of a "confining" phase that, no matter what, these massless fermions will break chiral symmetry, and therefore get a mass? And why can't that be a test for confinement?

- *Strassler:*

That definitely does not happen in general in supersymmetric theories. There are confining theories with massless fermions, which are non-trivially charged, that do not break chiral symmetries.

- *Englert:*

Because you compensated somehow...?

- *Strassler:*

The arguments that confinement automatically generates chiral symmetry breaking have never been transparent to me, so it is hard for me to say why they are

they wrong. But there are certain assumptions which go in. All I can tell you is that in the supersymmetric context one has very clear examples where it does not occur.

- *Englert:*

Does there exist a proof that N=1 supersymmetry is not broken by non-perturbative effects?

- *Strassler:*

Some supersymmetric theories do spontaneously break supersymmetry. There is no statement that it is impossible in all theories. But there are statements in individual theories that in that theory it is impossible. In most theories, in fact, one can prove supersymmetry breaking is impossible, including all nonperturbative effects. (Seiberg's work of 1993-1994.)

- *Abramovich:*

Would it not be enough to say that confinement exists because we do not observe light quarks? This is an experimental statement.

- *Strassler:*

But there is absolutely no guarantee that isolated quarks with large masses could not exist.

- *Abramovich:*

But the searches for isolated quarks were based on the search for fractional charges and were independent of the masses of the quarks.

- *Strassler:*

Yes, but this becomes a dynamical question. The likelihood that you will find one of these things is connected with the question of how many of them are left over from the early universe, after the QCD phase transition. I do not think anyone actually knows how to calculate this quantity. If the density is exponentially suppressed, then you will not see them.

- *Abramovich:*

But you actually claimed that in QCD quarks are not really confined.

- *Strassler:*

What I said was that there is no demonstration that. First of all, isolated quarks cannot exist, and second, there is no definite notion (as there would be for Yang-Mills with extremely heavy quarks) that it takes an infinite amount of energy to pull a pair of quarks apart – in fact, the latter is not true! If I want to pull an up quark out of a proton, it takes finite energy in QCD; all I have to do is make one additional quark-

antiquark pair, leaving myself with proton and a pion which I can pull as far apart as I want.

- *Abramovich:*
  But are you not happy with QCD? Are you saying it is the wrong theory?

- *Strassler:*
  No, QCD is great! QCD does everything we want it to do. The theory and experiment are consistent, and neither of them has what you would call strict confinement. My motivation for emphasizing this is that our usual language is extremely loose. We are not very clear in textbooks, or in our general conversation, what we are talking about. It is very important if you want to study this subject theoretically that you become much more precise. The strict confinement which one uses Wilson loops for and which one can hope to study in some topological manner is not of the form found in QCD. QCD has a much more limited sense in which quarks are confined.

- *Nobbenhuis :*
  Although confinement is generally assumed to be a non–perturbative phenomenon, it seems that, since confinement rests heavily on the masses of the quarks, that there must be a correspondence or constraints between confinement and the mechanism of chiral symmetry breaking. What are your ideas about that?

- *Strassler:*
  Even this is not obvious. The instanton liquid model has a very interesting property: it does a great job predicting the size of the chiral condensate and, as Shuryak himself has pointed out, a rather bad job of predicting the string tension in an Yang-Mills theory. This really is not clear either from the lattice data or from experimental data or from any other theoretical arguments. This remains very much an open question. And the fact that we know from supersymmetric theories that chiral symmetry breaking and confinement do not have to be related to one other – sometimes you find both of them, and sometimes you don't – certainly raises the question as to whether there are non-supersymmetric theories that have confinement but no chiral symmetry breaking. We simply do not know the answer to that. There is no proof one way or the other. So these are open questions which are very interesting, and I am glad you raised them, but I do not know the answer.

- *Markov:*
  Why consider groups other than SU(3), etc?

- *Strassler:*
  One of the things we learned from supersymmetric theories in the last decade, by studying a large number of them with very different matter content, was that these

theories are very different from one another. It used to be that people thought that QCD was basically all there was: basically, the gluons generate confined flux, and the quarks just go along for the ride. All the theories would basically be the same. But the supersymmetric case shows this is absolutely not the case. So, we have, from that point of view, strong reasons to think that nonsupersymmetric theories form a very rich system, with a wide variety of behaviours. Perhaps there are some with confinement but without chiral symmetry breaking. We already know there are some conformal field theories (at large N_f and N_c) and there may be many of these. Four-fermion interactions may also play an interesting role in the low-energy dynamics of a theory. And we will not really understand QCD as a theory until we understand the context in which it sits. In fact, people had guessed confinement occurs in certain supersymmetric gauge theories as far back as the 1980s, but they did not really understand what they were looking at because they did not know about all the other theories, and about duality, and about all the other phases these theories can be in. Only through the study of large classes of theories did the whole picture start to make sense. So, from my point of view, I would very much like to have a similar picture for all of nonsupersymmetric gauge theory, and understand QCD as a special case, which I will appreciate better because I understand where it sits. And this is not merely an academic exercise. We should remember that there are all sorts of unsolved problems in particles physics, which might have solutions through unexpected dynamics of nonsupersymmetric theories which as yet we have not guessed. For example, electroweak symmetry breaking... technicolor, in particular, is only ruled out in QCD-like theories, but we do not have any reason to think that most nonsupersymmetric theories are QCD-like. Since we are also now learning that questions in nonsupersymmetric string theory can be related to questions in nonsupersymmetric field theory, these issues are becoming all the more pressing. We live in a nonsupersymmetric world; we really have to understand much better how it works.

- *Maas:*

Would you also perhaps be able say that showing that all particles in a QCD-like theory are not part of the physical Hilbert space, by showing that they do not have a Källen-Lehman representation or something like that, would also qualify as showing confinement?

- *Strassler:*

What you really need to show in QCD, to show that isolated quarks cannot exist, is that all states in the theory with non-integer electric charge have infinite energy. I do not think the criterion you propose is sufficient, although it might well be necessary. In Yang-Mills theory the Wilson loop provides a simple test of confinement of flux. But there is no simple test of the absence of isolated quarks.

CHAIRMAN: M. STRASSLER

Scientific Secretaries: C. Pica, A. Nesterenko

DISCUSSION II

- *Strassler:*

One thing I want to emphasize is that what Seiberg duality has done is link together electric-magnetic duality, the relationship between QCD and the sigma model, which you might not have even thought was a duality, confinement-superconductor duality, and also a couple of others, those of extended supersymmetry: that of the famous Seiberg-Witten N=2 supersymmetric gauge theories, and Montonen-Olive duality of N=4 supersymmetric gauge theory. Thus it provides a unified framework where all these different pieces of physics, formerly thought to be distinct, are now known to be the same.

What we did today is that we found a confining gauge theory, or more precisely a set of theories. If you look carefully at what we did, you will see that we wanted to find out about SO(8) with five massless fields in the 8 (vector) representation, but what we actually did was study SO(8) with five massless fields *and two fairly light fields* in the 8 representation. The light fields do not change the long-distance dynamics, but their presence means that the dual description ends up being a weakly coupled non-abelian Higgs model, with an explicit dual Meissner effect. On the other hand, if we try to make the light fields very massive, the dual description becomes strongly coupled — and that is just the way things work. We are pretty convinced that confinement continues to hold, but the mechanism of confinement is no longer easy to describe in any set of variables at our disposal.

I think one of the most important things we saw was that many strongly coupled gauge theories do not confine at all. And that could very well be true beyond the supersymmetric case. I think this is a tremendously important result of Seiberg's work. We also saw that duality has a structure to it, and a lot of what we thought was dynamics is actually structure. It is really nice to see this in the supersymmetric case, but what is especially important is that, although the extra symmetries of supersymmetry made this discovery possible, nothing suggests that duality is a property of supersymmetric theories *only*. In fact, there are many examples of duality which are well known, including exactly soluble ones in continuum field theory in 1+1 dimensions, which are totally non-supersymmetric. So again, I presented all this physics in a supersymmetric context, because that is where we have been able to solve the problem — though only in part, since, as you should remember, Seiberg duality is still an unproven conjecture. But in the non-supersymmetric case, although we have absolutely no idea what is going on, we have every reason to think that similar physics holds: that the physics of confinement and of chiral symmetry breaking is

similar, and that the physics of duality is probably there. Maybe one of you will find it.

Now, we also saw nothing which suggests that confinement in Yang-Mills or in QCD really has a simple description. But maybe *showing* that confinement occurs is possible, and easier, if we understand a little bit more about how non-supersymmetric theories work in general.

- *Maas:*

What happens to duality if supersymmetry is either explicitly or dynamically broken?

- *Strassler:*

Generally you lose control, because you lose the symmetries that you need in order to do the calculations which allow you to check whether the duality applies or not. So what usually happens is you have a weakly coupled theory and a strongly coupled theory, or maybe two strongly coupled theories, and you add supersymmetry breaking on both sides. If the supersymmetry breaking is sufficiently small, small compared to the confinement scale, for example, or small compared to some strong coupling scale, you are still OK, and you can do the computation. So small supersymmetry breaking can be studied. However if the supersymmetry scale is large compared to the confinement scale, you do not have control, and so you cannot start with supersymmetric QCD, break supersymmetry a lot, and study ordinary QCD; that does not work. But you can study a small amount of supersymmetry breaking, both dynamical and explicit and that has been useful.

- *Catà:*

This morning you mentioned that weak duality can be adjusted so as to become strong duality. Could you comment a little more on this?

- *Strassler:*

Imagine you take the chiral lagrangian and you start adding operators to it, in some very clever way, so as to get every amplitude a little bit closer to QCD than it would be otherwise, and you keep changing the operators, and somehow you extremely cleverly manage to even match QCD up to, say, 5 or 6 GeV, 10, 40 GeV. With the proper sets of adjustments — an infinite series of terms, and some non-locality, probably — you somehow manage to reproduce all of QCD, including weakly propagating quarks and gluons, which from the point of view of hadrons are some sort of fractional solitons. It is hard to believe you could actually do this, but in principle it should be possible. I know a class of examples where we can actually do that kind of thing (supersymmetric QED in 2+1 dimensions, described in the work that I have done with Kapustin.) Now, why do we care? You would like to believe that the duality transformation you do at low energies when you go from QCD to hadrons, which is some sort of approximate thing, is itself approximating some

change of variables which is really exact, and that you could in principle perform in a path integral. Of course, such a dual description would be a mess in the ultraviolet; since QCD perturbation theory works well there, the corresponding hadronic description of the theory must be strongly coupled. In some sense, it is an accident of the physics of QCD that we start with weakly coupled QCD, where perturbation theory works — we are happy, so we do not try to change variables — and we go down to low energies until it becomes strongly coupled — we are not happy, so we change variables there. We only try to do it approximately, in part because it is all we can do in practice, but in part because we are only trying to describe physics at low energies; we do not need a change of variables that is exact all the way to high energy. But there may be some contexts, on a more theoretical level, where you would want to know this. For example, if you try to understand how the duality of QCD fits into the dualities of many other theories, finding the exact transformation, or at least knowing that it exists in principle, might be very important.

- *Zichichi:*

You said "Nothing suggests there is a simple description of the confinement of flux in YM (pure gauge) or of the absence of isolated quarks in QCD". Since a long time I am trying to convince my friends that confinement should not be taken for granted. This is interesting, because it prevents good experiments from being supported in the field of free quark searches.

- *Strassler:*

But with this comment, I am only making a statement about theory, not about experiment. The question is: can one start with the lagrangian or with the path integral of QCD and prove that these facts are true through a simple description of the dynamics?

- *Zichichi:*

But do we have a theorem that confinement occurs in QCD theory?

- *Strassler:*

Absolutely not. We do not have such a theorem. We have evidence, but we do not have a theorem. And I think it is a hard problem for many reasons. But in particular, there was a great hope in the 1970's — and I think Professor 't Hooft had an enormous amount to do with this — there was a real hope there should be some way to describe the mechanism of confinement using some sort of an effective description which one could ideally derive directly from QCD. In other words, throw away the irrelevant part of QCD, and focus on some piece which really is the heart of confinement — perhaps in analogy to the way that Shuryak tells us to throw away most of QCD, and to focus on instantons and think of them as a liquid, in order to capture much of the physics of chiral symmetry breaking. There is no evidence that confinement has such a simple description. In fact, every time I find a simple

description of confinement in some theory, it is always a theory which is close to QCD but not quite. And every time I try to go over to a theory which is very much like QCD, the simple description always goes away, in every single case. Furthermore, I am certainly not guaranteed that a theory such as QCD has a simple description of its dynamics waiting for us to find; some functions just do not have a nice perturbative expansion, so what can you do? So, maybe this is the wrong question. As I have suggested, we would also be happy if there is a different way to prove that confinement occurs, one that does not go into the dynamics, but somehow ties this theory together with a deep mathematical structure, from which we could derive many other properties as well as confinement. And maybe that is a better approach. But we are still far from knowing what form that could take in the non-supersymmetric case, so this is a pure speculation.

- *Zichichi:*

Since this is a School, it is interesting to make this point clear, because when I proposed the experiment at CERN for search in DIS neutrino–proton scattering for quarks, I had the Theory Division against me. And I could defend myself, because I knew something of what was going on. My answer was that "Why is the asymptotic freedom in QCD clearly resolved and confinement is not understood?" The experiment was approved against the advice of the Theory group at CERN. So, the quark should go on being searched for in high energy collisions.

- *'t Hooft:*

I want to add one remark. What experimentalists can tell us is that QCD does not produce light glueballs, say 10 MeV, or 100 MeV. The pion with 135 MeV is the lightest hadron. This means something for the vortex lines. If they would spread out a lot in the transverse direction, they would show physical fluctuations over large distance scales, i.e., light glueball excitations. So, there is a limit to the transverse size of the vortexes, and that should imply exact and not approximate confinement. It is hard however to pin down the numbers for this argument, but I think that the lightest glueball state gives a lower limit for the confining string constant.

- *Abramowic:*

Is confinement seen in lattice QCD calculations?

- *Kenway:*

In pure YM theories we measure an area law for Wilson loops at all values of the gauge coupling. We cannot prove this holds in the continuum limit, but the absence of a phase transition suggests that it does. For simulations with light dynamical quarks, Wilson loops do not provide a good signal for confinement, because quarks pop out of the vacuum to screen the colour charges, so the Wilson loop follows a perimeter law. We do see hadronic bound states with a large mass gap, just as in experiment, so we

are in a similar position to experimentalists in answering this question. There is no proof coming out of the lattice, just a set of consistency checks.

- *Strassler:*

It is important to keep in mind that I am not suggesting that anyone really *believes* that YM does not confine; we have a lot of evidence that it does, from many different places. But one would like to have some understanding of this which goes beyond simply either experimental observation or a numerical effort, where the answer comes out but you cannot clearly see why. So, from that point of view, the lattice does indeed confirm that the theory predicts confinement — it confirms the theory is correct — but it does not explain *why* it confines. We really do not understand gauge theory that well, if we do not understand this phenomenon. There is a still long way to go.

- *Markov:*

In strong interactions we have only two stable particles — pions and nucleons — and they form one Hilbert space. In effective theories we include for each resonance its own field. Are these two spaces connected by a field transformation? What does this have to do with confinement?

- *Strassler:*

I do not believe that in general it is exactly clear when you change variables, from quarks and gluons into a theory of bound states, exactly how you introduce the right degrees of freedom without double counting, for example. And I do not know of a toy model where you can actually see what happens to hadron interactions and how they might actually work in QCD. I think that part of the question is: given that you have so many resonances in QCD — and let me be more explicit, supposing one is looking at large N QCD where you have many stable resonances — how many fields are you supposed to introduce to describe the hadronic physics? We actually have a little bit of insight into that question now because we know that at least in some circumstances there is evidence that certain QCD-like confining gauge theories are actually equivalent to string theory. So at some level, you have to introduce a whole stringy spectrum. So the answer might be that one has to introduce a string's worth of fields in order to describe all the hadrons of large N gauge theories.

- *Bechtle:*

Is it possible to derive a realistic dual model to QCD, which has real phenomenological predictions in the non-perturbative regime?

- *Strassler:*

The chiral lagrangian should not be underestimated and people did a great deal with it. Again, the trick there, as I said this morning, is to combine certain experimental knowledge with the constraints of symmetries, and write the most

general lagrangian we can. Enormous progress was made in the 1960's even before we knew about QCD, and it continues to be made. There is wonderful literature on this stuff. But if you are asking the first principles question: can we start from QCD and derive the dual theory? the answer is absolutely not. In no non-supersymmetric theory in 4 dimensions can this be done. And even in the supersymmetric case, you learn some things, but you do not learn everything. For example, in the various dual descriptions of supersymmetric YM theory, we do not know the glueball spectrum. Even in the supersymmetric case, a great deal of work is left to be done, and clearly the non-supersymmetric case, which is, at least at the moment, more important, is just that much harder. These are really tough theoretical questions, and in fact, it is amazing to me that any progress has been made at all!

- *Gromov:*
  What is the strictest definition of weak duality?

- *Strassler:*
  The classic example is that of QCD and the theory of pions. And again, think about the case where the quarks are really massless and the pions are really light, so if you look at pion-pion scattering, it really has low-energy theorems which govern it. Those low-energy theorems we really believe are a property of QCD. When you get to higher energies, the closer you get to 1 GeV, the less you believe the pion theory is describing QCD correctly, because there are more and more ambiguities as to what the theory actually is. There are all sorts of non-renormalizable terms, many coefficients to fix, a cutoff you do not really understand — the loops are divergent, so you have to cut them off somehow — but it is still a very good theory down at low energies. This is an example which I called weak duality because you have two different theories, but they have the property that if you ask about their Green's functions at long distances, the Green's functions will agree, up to corrections that are small and get smaller as you consider larger distances.

  Now, there is a related concept that shows up especially in condensed matter physics, but is meaningful also in our context, which is the notion of "universality class". If I have two theories in the same universality class, what I mean is that their long distance physics is the same. Now, a trivial example of two theories which are in the same universality class would be (1) QCD and (2) QCD with some additional flavour-conserving four-quark operators, suppressed by powers of the Z-boson mass. These extra interactions are all short range, so the two theories are obviously the same at long distances. More interesting would be if you could find a theory with the same long distance physics as QCD, but which is qualitatively very different from it in appearance, and is related to it in some non-local way --- note that pions look fluffy from the QCD point of view, but are treated as point-like in the chiral Lagrangian. Then what you have done is found a theory which is in the same universality class as QCD, but also has degrees of freedom with a very non-trivial relationship to those of

QCD. In this case you would say this theory is weakly dual to QCD, as I have done with the chiral Lagrangian. This is distinct from a strong duality where you do not merely have two theories in the same universality class; rather, you actually have only one theory, with two descriptions related by a non-trivial change of variables in the path integral. Again, there are explicit examples of strong duality, and Burgess and Quevedo give you a great example, one that anyone with a one-year course in field theory can read.

- *Nobbenhuis:*

Could you say a few words about a possible relation between QCD defined in different spacetime geometries and confinement?

- *Strassler:*

This is not a problem that I have personally thought about a great deal, though it has been thought about by some people. A lot of work was done on QCD in two dimensions.

- *Wendland:*

QED has proven extremely successful in predicting experimental results. Yet, it is still poorly understood from a mathematical point of view, especially since it has infinities. Are mathematicians working on this? If so, what is their approach?

- *Strassler:*

As you said, QED has some difficulties from its infinities, but that is not the real problem. The real problem is of course the Landau pole, so the theory has a cut-off. The main mathematical work has been done with the case of φ^4 field theory, which of course is trivial in 4 dimensions, as it has a Landau pole like QED and has to be defined with a cut-off. If you take the cut-off away, you find to have a free theory. I do not know what mathematicians are doing with QED. But I do know this: the mathematicians are generally trying to start with a lattice definition of the theory, because then there is some reasonable cut-off of the theory that one can understand. All the supersymmetric field theories I wrote down have a very severe problem, which is that nobody knows how to define them, because you cannot put supersymmetry on a lattice in any controlled way. This is a huge outstanding problem which should also not be underestimated. I certainly believe these theories exist, but it is hard for mathematicians to prove things like duality in theories that they are not sure how to define.

- *Maas:*

In recent lattice calculations, it seems in some classes of partial gauge fixing, like maximal abelian gauge or centre gauge, that only a subgroup of the gauge group is responsible for confinement. Is something similar known in supersymmetric theories?

- *Strassler:*

The answer is yes and it is similar to that what we saw earlier. If you choose the right theory which is not QCD-like, you can find there is something like a maximal Abelian projection going on dynamically. The case where this is most clear is in the Seiberg-Witten theory itself, generalized to SU(N). That is an N=2 gauge theory, so N=1 supersymmetric YM plus an extra gluino plus a scalar in the adjoint representation. This theory, because of that scalar, effectively has a dynamical abelian projection. But again there is always a problem with abelian projections, in particular dynamical ones, which is that because SU(N) has a centre which is Z_N you expect that you have flux tubes carrying Z_N charges. If you really project the theory onto the abelian subgroup, then there is a tendency to end up with flux tubes that carry extra charges, approximate extra charges, but ones that dynamically they should not have. The maximal abelian projection has the problem that it again works for some particular variants of YM theory, but if you really want to go back to YM theory, you would lose control, and again you do not have any understanding of what really is going on. In addition, you have extra charges in those models that are approximate, which you know that flux tubes of the real YM theory do not carry.

- *Stamen:*

Duality seems to be a powerful tool. What else can you learn from duality in addition to the statement on confinement you made during the lecture? How can you judge which kind of physics can be transferred from one description to the other?

- *Strassler:*

Weak duality (let us stick with that case) helps you to figure out what is the essential long distance behaviour of a gauge theory. Now, that is a limited question. On the other hand, when you think about non-supersymmetric gauge field theories — suppose I handed you a non-supersymmetric gauge theory — you cannot answer even that basic question. So, what duality, at least in the N=1 supersymmetric context, allows you to do is to figure it out. Then, to go beyond that to some extent, you use a little bit of physics intuition. For example, even in understanding some properties of this conformal field theory that appeared in the weak duality of two asymptotically free gauge theories, there are some more things you can do. In some regimes, for example, the low-energy conformal field theory might almost be perturbative in one description, so you can try to calculate a few things, to see what you can learn about it. And you would not know that theory A flows to a conformal field theory which is weakly coupled in the B variables. That is something you cannot learn without duality. And then, of course, you can learn a great deal about the basic structure of field theory, because in understanding how relations like this can come about you are also motivated to look at field theory in new ways. But if you ask what else can you calculate, then you start finding that you really start using the supersymmetry. For example, I can calculate the dimensions of some operators in a low-energy conformal

field theory; that is not something I expect to be able to do in the non-supersymmetric case. Also, sometimes you learn some unexpected dynamical phenomena (such as accidental global symmetries (see my work with Leigh, and also Distler and Karch) that you cannot find out in another way.

At some point, we found that we had a monopole which carries a Z_2 charge which emerged from topology. Nobody asked me: why does it come out with a Z_2 charge? Remember this monopole is dual to P, the spinor of SO(8). I claim that P also carries a Z_2 charge, and that this charge is the only thing you can observe about the spinor when it is a massive field in an SO(8) gauge theory with vectors. It is therefore completely consistent that the dual monopoles should carry Z_2 charges as well. See if you can convince yourselves that this is true.

- *Englert:*

Are there basic relations between the duality in field theories and the duality in string theory?

- *Strassler:*

Absolutely. The dualities of field theory, in the different contexts where they show up, and the dualities of string theory, in the different context where they show up, fit together in an extraordinarily beautiful way. One could give an entire set of lectures on how one connects them all together. In addition, of course, I mentioned that field theory and string theory are dual, and you can see within that structure that the dualities of string theory and the dualities of field theory are in fact dual — that is, different descriptions of the same thing. It is all very complicated and very beautiful; I cannot give you a one-line answer as to exactly how it works. It is a huge subject. The mathematical structure goes all the way from the highest ethereal planes of M-theory down to 2 dimensional field theories, and somehow fills the whole space between them.

- *Englert:*

Is your feeling that the root of field theory dualities lies in string theory?

- *Strassler:*

No, I believe that duality comes from within the theory you are studying. You do not derive field theory duality from string theory; you should be able to write down the field theory path integral, do a duality transformation and get an answer. Again, Burgess and Quevedo give a lovely example of this. I believe that duality transformations should be self-contained, and thus you should be able to show field theory duality within field theory. And I would even go further as to say that I do not know of any quantum mechanical theory which does not potentially have duality. So, for all I know, duality in principle underlies all of quantum mechanics, all of quantum theory. And if that is true, the mathematics of duality is probably all the more profound.

Probing Grand Unification Through Neutrino Oscillations, Leptogenesis, and Proton Decay

Jogesh C. Pati
Department of Physics, University of Maryland,
College Park MD 20742

June 11, 2003

Abstract

Evidence in favor of supersymmetric grand unification including that based on the observed family multiplet-structure, gauge coupling unification, neutrino oscillations, baryogenesis, and certain intriguing features of quark-lepton masses and mixings is noted. It is argued that attempts to understand (a) the tiny neutrino masses (especially $\Delta m^2(\nu_2 - \nu_3)$), (b) the baryon asymmetry of the universe (which seems to need leptogenesis), and (c) the observed features of fermion masses such as the ratio m_b/m_τ, the smallness of V_{cb} and the maximality of $\Theta^{osc}_{\nu_\mu\nu_\tau}$, seem to select out the route to higher unification based on an effective string-unified $G(224) = SU(2)_L \times SU(2)_R \times SU(4)^c$ or $SO(10)$-symmetry, operative in 4D, as opposed to other alternatives.

A predictive framework based on an effective $SO(10)$ or $G(224)$ symmetry possessing supersymmetry is presented that successfully describes the masses and mixings of all fermions including neutrinos. It also accounts for the observed baryon asymmetry of the universe by utilizing the process of leptogenesis, which is natural to this framework. It is argued that a conservative upper limit on the proton lifetime within this $SO(10)/G(224)$-framework, which is so far most successful, is given by $(\frac{1}{3} - 2) \times 10^{34}$ years. This in turn strongly suggests that an improvement in the current sensitivity by a factor of five to ten (compared to SuperK) ought to reveal proton decay. Implications of this prediction for the next-generation nucleon decay and neutrino-detector are noted.

1 Introduction And An Overview

Since the discoveries (confirmations) of the atmospheric [1] and solar neutrino oscillations [2, 3], the neutrinos have emerged as being among the most effective probes into the nature of higher unification. Although almost the feeblest of all the entities of nature, simply by virtue of their tiny masses, they seem to possess a subtle clue to some of the deepest laws of nature pertaining to the unification-scale as well as the nature of the unification-symmetry. In this sense, the neutrinos provide us with a rare window to view physics at truly short distances. As we will see, these turn out to be as short as about 10^{-30} cm. Furthermore, it appears most likely that the origin of their tiny masses may be at the root of the origin of matter-antimatter asymmetry in the early universe. In short, the neutrinos may well be crucial to our own origin!

The main purpose of my talk here today will be to present the intimate links that exist in the context of supersymmetric grand unification between the following phenomena: (i) neutrino oscillations, (ii) the masses and mixing of quarks and charged leptons, (iii) gauge coupling unification, (iv) baryogenesis via leptogenesis, and last but not least (v) proton decay.

To set the background for a discussion along these lines, let us first recall that with only left-handed neutrinos, the standard model based on the gauge symmetry $SU(2)_L \times U(1)_Y \times SU(3)^c$, despite its numerous successes, fails to account for the magnitude of the mass-difference square $\Delta m^2(\nu_2 - \nu_3) \sim (1/20\ \text{eV})^2$ observed at Superkamiokande [1]. Incorporating effects of quantum gravity,[1] the standard model can lead to a neutrino-mass $\sim 10^{-5}$ eV, which is, however, too small to account for the SuperK effect. One can in fact argue that, to understand the magnitude of the SuperK effect in any natural way, one would need *new physics beyond the standard model* at an effective mass-scale $\sim 10^{15}$ GeV, rather than at the Planck scale $\sim 10^{19}$ GeV [4]. Interestingly enough, one can link this effective mass-scale to the scale of meeting of the three gauge couplings (to be discussed below) which is around 2×10^{16}GeV. That, in turn, hints at a link between the physics of neutrino-oscillations and grand unification!

The idea of "grand unification" was introduced in the early 1970's [5, 6, 7], purely on aesthetic grounds, in order to remove certain conceptual shortcomings of the standard model. Over the years, a set of key observations — some old and some new — have come to light, which together provide strong evidence in favor of this idea. Some of the observations in fact support the idea of both grand unification and low-energy supersymmetry [8, 9]. The evidence includes:

1. The observed family multiplet-structure — in particular the fact that the five (apparently disconnected) multiplets of the SM belonging to a family neatly become parts of a whole — *a single multiplet* — under grand unification, with *all* their quantum numbers predicted precisely as observed.

[1]See, e.g., S. Weinberg, *Phys. Rev. Lett.* **43**, 1566 (1979); *Proc. XXVI Int'l Conf. on High Energy Physics*, Dallas, TX, 1992; E. Akhmedov, Z. Berezhiani and G. Senjanovic, *Phys. Rev.* **D47**, 3245 (1993). Assuming that quantum gravity could induce violation of lepton number, one may allow for an effective non-renormalizable operator of the form $\lambda_L LLHH/M_{pl}$ + h.c., scaled by $M_{pl} = 1.2 \times 10^{19}$ GeV with $\langle H \rangle \approx$ 250 GeV. Such an operator would, however, yield a rather small Majorana mass $m(\nu_L) \sim 10^{-5}$ eV for the left-handed neutrinos, even for a maximal $\lambda_L \sim 1$, as mentioned in the text.

2. The observed quantization of electric charge and the fact that the electron and the proton have exactly equal but opposite charges.

3. The dramatic meeting of the three gauge couplings that is found to occur at a scale $M_X \approx 2 \times 10^{16}$GeV, when they are extrapolated from their values measured at LEP to higher energies, in the context of supersymmetry [10].

4. The tiny neutrino masses of the sort suggested by the discoveries/confirmations of atmospheric and solar neutrino oscillations. These, as we will see, not only go well with the scale of unification M_X mentioned above but also help select out a class of unification-symmetries which provide the right-handed neutrinos ($\nu_R's$) as a compelling feature and B-L as a local symmetry.

5. Certain intriguing features of the masses and mixings of the quarks and leptons, including the relation $m_b(M_X) \approx m_\tau$ and the largeness of the $\nu_\mu - \nu_\tau$ oscillation angle ($\sin^2 2\theta^{osc}_{\nu_\mu\nu_\tau} \geq 0.92$) together with the smallness of $V_{cb}(\approx 0.04)$ [11].

6. And last but not least, the likely need for leptogenesis [12, 13] to account for the observed baryon-asymmetry of the universe, which seems to require once again the existence of superheavy right-handed neutrinos ($\nu_R's$) and B-L as a local symmetry.

All of these features including the tiny neutrino masses and the observed baryon-asymmetry can be understood simply, and even quantitatively, within the concept of supersymmetric grand unification based on an *effective symmetry in four dimensions*, that is either

$$G(224) = SU(2)_L \times SU(2)_R \times SU(4)^C \quad [5]$$

$$\text{or} \qquad SO(10) \quad [14].$$

Believing in a unified theory of all forces including gravity, it is of course attractive to presume that such an effective symmetry in 4D ($G(224)$ or $SO(10)$) has its origin from a string theory [15] or the M-theory [16]. I will comment in Sec. 2 that, in the context of a string theory with the string-scale being close to the GUT-scale, the observed coupling unification may be understood even if the effective symmetry in 4D, below the string scale, is non-simple like $G(224)$. A string-derived $G(224)$-solution may, however, have an advantage over an $SO(10)$-solution in that it can neatly avoid the so-called doublet-triplet splitting problem (generic to SUSY GUTs, see Sec. 2). Motivated by the desire to avoid this problem, there have in fact been several attempts in the literature (many rather recent) which successfully obtain semi-realistic $G(224)$-solutions in 4D from compactification of a string theory [17], or of an effective five or six dimensional GUT-theory [18]. For most purposes, in particular for considerations of fermion masses, neutrino-oscillations, and leptogenesis, the symmetries $G(224)$ and $SO(10)$ provide essentially the same advantages. Differences between them in considerations of proton decay will be noted in Sec. 5.

Let us first recall the new features (relative to the SM) which are introduced through the symmetry $G(224)$ [5]. Subject to left-right discrete symmetry ($L \leftrightarrow R$), which is natural to $G(224)$, all members of the electron family become parts of a single left-right self-conjugate multiplet, consisting of:

$$F^e_{L,R} = \begin{bmatrix} u_r & u_y & u_b & \nu_e \\ d_r & d_y & d_b & e^- \end{bmatrix}_{L,R}. \tag{1}$$

The multiplets F_L^e and F_R^e are left-right conjugates of each other and transform respectively as (2,1,4) and (1,2,4) of $G(224)$; likewise for the muon and the tau families. The symmetry $SU(2)_L$ treats each column of F_L^e as a doublet; likewise $SU(2)_R$ for F_R^e. The symmetry $SU(4)$-color unifies quarks and lepotons by treating each row of F_L^e *and* F_R^e as a quartet; *thus lepton number is treated as the fourth color.* As mentioned above, because of the parallelism between $SU(2)_L$ and $SU(2)_R$, the symmetry $G(224)$ naturally permits the notion that the fundamental laws of nature possess a left $\leftrightarrow$ right discrete symmetry (i.e. parity invariance) that interchanges $F_L^e \leftrightarrow F_R^e$ and $W_L \leftrightarrow W_R$. With suitable requirements on the Higgs sector, observed parity violation can be attributed, in this case, entirely to a spontaneous breaking of the $L \leftrightarrow R$ discrete symmetry [19].

Furthermore, the symmetry $G(224)$ introduces an elegant charge formula: $Q_{em} = I_{3L} + I_{3R} + (B-L)/2$, that applies to all forms of matter (including quarks and leptons of all six flavors, Higgs and gauge bosons). Note that the weak hypercharge of the standard model, given by $Y_W = I_{3R} + (B-L)/2$, is now completely determined for all members of a family. Quite clearly, the charges I_{3L}, I_{3R}, and B-L, being generators respectively of $SU(2)_L$ $SU(2)_R$, and $SU(4)^c$, are quantized; so also then is the electric charge Q_{em}. Using the expression for Q_{em}, one can now explain why the electron and the proton have exactly equal but opposite charges.

Note also that postulating either $SU(4)$-color or $SU(2)_R$ forces one to introduce a right-handed neutrino (ν_R) for each family as a singlet of the SM symmetry. *This requires that there be sixteen two-component fermions in each family, as opposed to fifteen for the SM.* Furthermore, $SU(4)$-color possesses B-L as one of its generators. This in turn helps to protect the Majorana masses of the right-handed neutrinos from being of the order string or Planck-scale.[2] In addition, $SU(4)$-color provides the Dirac mass of the tau-neutrino by relating it to the top-quark mass at the unification-scale, and simultaneously the mass of the bottom quark in terms of that of the tau-lepton. In short, $SU(4)$-color introduces *three characteristic features* — i.e.

1. the right-handed neutrinos as a compelling feature,
2. B-L as a local symmetry, and
3. the two GUT-scale mass relations:

$$m_b(M_X) \approx m_\tau \quad \text{and} \quad m(\nu^\tau_{\text{Dirac}}) \approx m_{\text{top}}(M_X) \tag{2}$$

These two relations arise from the SU(4)-color preserving leading entries in the fermion mass-matrices (see Sec. 3) which contribute to the masses of the third family. The sub-leading corrections to the fermion mass matrices that arise from $SU(4)$-color-breaking in the (B-L)-direction turn out to be important for the masses and mixings of the fermions belonging to the first two families [11]. As we will see, these three ingredients, as well as the SUSY unification-scale M_X, play *crucial roles* in providing us with an understanding of the

[2]Without a protection by a local symmetry, $\nu_R's$ (being singlets of the SM) are likely to acquire Majorana masses of the order string or Planck scale through effects of quantum gravity. Such ultraheavy ν_R-masses would, however, lead via the seesaw mechanism (see later), to too small masses for the light neutrinos ($\leq 10^{-5}$eV) and thereby to too small a value for Δm^2_{23} compared to observation. Hence the need for B-L as an effective local symmetry in 4D near the string scale.

tiny masses of the neutrinos as well as of the baryon-asymmetry of the universe, by utilizing respectively the seesaw mechanism [20] and the idea of leptogenesis [12]. *The success of the predictions in this regard (see below), speaks in favor of the seesaw mechanism and suggests that the effective symmetry in 4D, below the string-scale, should contain $SU(4)$-color.*

Now the minimal symmetry containing $SU(4)$-color on the one hand and also possessing a rationale for the quantization of electric charge on the other hand is provided by the group $G(224)$. The group $G(224)$ being isomorphic to $SO(4)\times SO(6)$ embeds nicely into the simple group $SO(10)$. The group $SO(10)$, which historically was proposed after the suggestion of G(224), of course retains all the advantages of $G(224)$, in particular the features (a)-(c) listed above. The interesting point is that $SO(10)$ even preserves the 16-plet multiplet structure of $G(224)$ by putting $\{F_L + (F_R)^c\}$ as its spinorial 16-dimensional representation, thereby avoiding the need for any new matter fermions. By contrast, if one extends $G(224)$ to the still higher symmetry E_6 [21], one must extend the family-structure from a 16 to a 27-plet, by postulating additional fermions.

Now utilizing the three ingredients (1), (2), and (3) listed above (thus assuming that $SU(4)$-color holds in 4D near the GUT-scale), together with the SUSY unification-scale (M_X) and the seesaw mechanism, one arrives at a set of predictions (see Sec. 3) which include [11]:

$$\begin{aligned} m_b(m_b) &\approx 4.7-4.9 \text{ GeV} \\ m(\nu_3) &\approx (\frac{1}{24}\text{ eV})(\frac{1}{2}-2) \\ \sin^2 2\theta^{\text{osc}}_{\nu_\mu\nu_\tau} &\approx 0.99 \\ V_{cb} &\approx 0.044 \end{aligned} \tag{3}$$

Each of these predictions agrees remarkably well with observations.The most intriguing feature is that this framework provides a compelling reason for why V_{cb} is so small (≈ 0.04), and simultaneously why $\sin^2 2\theta_{\nu_\mu\nu_\tau}$ is so large (≈ 1), both in accord with observations. It is worth noting that the last two results, showing a sharp difference between V_{cb} and $\theta_{\nu_\mu\nu_\tau}$, go against the often expressed (naive) view that the quark-lepton unification should lead to similar mixing angles in the quark and lepton sectors. Quite to the contrary, as we will see in Sec. 3, the minimal Higgs system provides a natural breaking of $SU(4)$-color along the (B–L)-direction which particularly contributes to a mixing between the second and the third families [11]. That in turn provides a compelling *group-theoretical reason* for a distinction between the masses and mixings of the quarks and leptons as in fact observed empirically.

One important consequence of having an effective $G(224)$ or $SO(10)$-symmetry in 4D is that spontaneous breaking of such a symmetry (thereby of B-L) into the SM symmetry naturally generates Majorana masses of the RH neutrinos that are of order GUT-scale or smaller. In correlation with the flavor symmetries which provide the hierarchical masses of the quarks and the leptons, the Majorana masses of the three RH neutrinos are found to be [22]: $(M_{N_1}, M_{N_2}, M_{N_3}) \approx (10^{15}, 2\times 10^{12}, (1/3-2)\times 10^{10})$ GeV. Given lepton number (thus of B-L) violation associated with these Majorana masses, and C and CP violating phases that generically arise in the Dirac and/or Majorana mass-matrices, the out of equilibrium decays of the lightest of these heavy RH neutrinos (produced after inflation[3]) into $l+H$ and $\bar{l}+\bar{H}$, and

[3]The lightest N_1 may be produced either thermally after reheating or non-thermally through inflaton-

the corresponding SUSY modes, generates a lepton-asymmetry. The latter is then converted into a baryon-asymmetry by the electroweak sphaleron process [12, 13]. In conjunction with an understanding of the fermion-masses and neutrino-oscillations (atmospheric and solar), the baryon excess thus generated is found to be (see Sec. 4 and Ref. [22]:

$$Y_B = \left(\frac{n_B - n_{\bar{B}}}{n_s}\right) \approx (\sin 2\phi_{21})(7 - 100) \times 10^{-11} \tag{4}$$

While the relevant phase angle ϕ_{21} arising from C and CP-violating phases in the Dirac and Majorana mass-matrices of the neutrinos is not predictable within the framework, it is rather impressive that for plausible and natural values of the phase angle $\phi_{21} \approx \frac{1}{2} - \frac{1}{20}$ (say), the calculated baryon excess Y_B agrees with the observed value based on big bang nucleosynthesis [23] and CMB data [24]. This may be contrasted from many alternative mechanisms, such as GUT and electroweak baryogenesis, which are either completely ineffective (owing to inflation and gravitino contraint) or yield too small a baryon excess even for a maximal phase. For a recent review and other relevant references on the topic of baryogenesis, see Ref. [25].

It should be stressed that the five predictions shown in Eqs(3) and (4), together make a crucial use of the three features (a)-(c) listed in Eq. (2), as well as of the SUSY unification-scale M_X and the seesaw mechanism. Now the properties (a)-(c) are the distinguishing features of the symmetry $G(224)$. They are of course available within any symmetry that contains $G(224)$ as a subgroup. Thus they are present in $SO(10)$ and E_6, though not in $SU(5)$. Effective symmetries like $[SU(3)]^3$ [26] or $SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times SU(3)^c$ [27] possess the first two features (a) and (b) but not (c). Flipped $SU(5) \times U(1)$ [28] on the other hand offers (a) and (b) but not the relation $m_b(M_X) \approx m_\tau$, which, however, is favored empirically.

The empirical success of the features (1)-(6), including specifically the predictions listed in Eqs. (3) and (4), seems to be non-trivial. Together they make a strong case for both the *conventional ideas of supersymmetric grand unification*[4] *and simultaneously for the symmetry* $G(224)$ *or* $SO(10)$ *being relevant to nature in four dimensions, just below the string scale.*

As mentioned before, the main purpose of my talk here will be to present the intimate links that exist, in the context of supersymmetric grand unification based on an effective $G(224)$ or $SO(10)$ symmetry, between (i) neutrino oscillations, (ii) the masses and mixings of quarks and charged leptons, (iii) gauge coupling unification, (iv) baryogenesis via leptogenesis, and last but not least (v) proton decay.

Perhaps the most dramatic prediction of grand unification is proton decay. This important process which would provide the window to view physics at truly short distance ($< 10^{-30}$ cm) and would greatly complement studies of neutrino oscillations in this regard is yet to be seen. One can, however, argue that the evidence listed above in favor of supersymmetric grand unification, based on an effective $G(224)$ or $SO(10)$ symmetry in 4D, strongly suggests

decay during reheating. Both possibilities are considered in Sec. 3.

[4]By "conventional" I mean gauge coupling unification occurring in 4D at a scale of few $\times 10^{16}$ Gev (for the case of MSSM), with the string-scale being somewhat larger. This is to be contrasted from the case of large extra dimensions with unification occurring at the TeV-scale, on the one hand, or from 5D GUTs possessing unification only in higher dimensions leading to the SM symmetry $G(213)$ in 4D, on the other hand.

that an upper limit on proton lifetime is given by

$$\tau_{\text{proton}} \lesssim (\frac{1}{3} - 2) \times 10^{34} \text{ yrs},$$

with $\bar{\nu}K^+$ being the dominant mode, and quite possibly μ^+K^o and $e^+\pi^o$ being prominent. This in turn suggests that an improvement in the current sensitivity by a factor of five to ten (relative to SuperK) ought to reveal proton decay. A next-generation megaton-size detector of the kind being contemplated by the UNO [29] and the Hyperkamiokande [30] proposals would thus be needed to probe efficiently into the prediction of the supersymmetric $G(224)/SO(10)$-framework as regards proton decay.

I have discussed in a recent review [31] in some detail the updated results for proton decay in the context of supersymmetric $SU(5)$, $SO(10)$ and $G(224)$-symmetries by taking into account (a) the recently improved (and enhanced) matrix elements as well as short and long-distance renormalization effects, (b) the dependence of the "standard" d=5 proton-decay operator on GUT-scale threshold corrections that are restricted by the requirement of natural coupling unification, and (c) its link with the masses and the mixings of all fermions including neutrinos [11]. The latter give rise to a new set of $d = 5$ operators, related to the Majorana masses of the RH neutrinos [32], which are found to be important. I will present a summary of the main results in this regard in Sec. 5 and also comment on the recent works which tend to avoid the standard $d = 5$ proton decay operators which generically arise in the context of supersymmetric grand unification.

In Sec. 2, I discuss the implications of the meeting of the three gauge couplings in the context of string-unification. In Sec. 3, I discuss fermion masses and neutrino oscillations within a predictive framework based on the $G(224)$ or $SO(10)$ symmetry in 4D which lead to predictions of the type shown in Eq.(3), and in Sec. 4, I discuss leptogenesis within the same framework. Results on proton decay which arise within this framework and within related approaches are summarized in Sec. 5. Concluding remarks are presented in Sec. 6, where the case for building a major underground detector with improved sensitivity to detecting proton decay and neutrino oscillations is made.

2 MSSM Versus String Unifications: $G(224)$ Versus $SO(10)$ as Effective Symmetries

As mentioned in the introduction, the three gauge couplings are found to meet when they are extrapolated from their values measured at LEP to higher energies by assuming that the SM is replaced by the minimal supersymmetric standard model (MSSM) above a threshold of about 1 TeV [10]. The meeting occurs to a very good approximation, barring a few percent discrepancy which can legitimately be attributed to GUT-scale threshold corrections. Their scale of meeting is given by:

$$M_X \approx 2 \times 10^{16} \text{ GeV (MSSM or SUSY } SU(5)) \tag{5}$$

This dramatic meeting of the three gauge couplings provides a strong support for the ideas of both grand unification and supersymmetry, as being relevant to physics at short distances $\lesssim (10^{16} \text{ GeV})^{-1}$.

In addition to being needed for achieving coupling unification there is of course an independent motivation for low-energy supersymmetry—i.e. for the existence of SUSY partners of the standard model particles with masses of order 1 TeV. This is because it protects the Higgs boson mass from getting large quantum corrections, which would (otherwise) arise from grand unification and Planck scale physics. It thereby provides at least a technical resolution of the so-called gauge-hierarchy problem. *In this sense low-energy supersymmetry seems to be needed for the consistency of the hypothesis of grand unification.* Supersymmetry is of course also needed for the consistency of string theory. It is fortunate that low-energy supersymmetry can be tested at the LHC, and possibly at the Tevatron, and the proposed NLC.

The most straightforward interpretation of the observed meeting of the three gauge couplings and of the scale M_X, is that a supersymmetric grand unification symmetry (often called GUT symmetry), like $SU(5)$ or $SO(10)$, breaks spontaneously at M_X into the standard model symmetry $G(213)$, and that supersymmetry-breaking induces soft masses of order one TeV.

Even if supersymmetric grand unification may well be a good effective theory below a certain scale $M \gtrsim M_X$, it ought to have its origin within an underlying theory like the string/M theory. Such a theory is needed to unify all the forces of nature including gravity, and to provide a good quantum theory of gravity. It is also needed to provide a rationale for the existence of flavor symmetries (not available within grand unification), which distinguish between the three families and can resolve certain naturalness problems including those associated with inter-family mass hierarchy. As alluded to in the introduction, in the context of string or M-theory, an alternative interpretation of the observed meeting of the gauge couplings is however possible. This is because, even if the effective symmetry in four dimensions emerging from a higher dimensional string theory is non-simple, like $G(224)$ or even $G(213)$, string theory can still ensure familiar unification of the gauge couplings at the string scale. In this case, however, one needs to account for the small mismatch between the MSSM unification scale M_X (given above), and the string unification scale, given by $M_{st} \approx g_{st} \times 5.2 \times 10^{17}$ GeV $\approx 3.6 \times 10^{17}$ GeV (Here we have put $\alpha_{st} = \alpha_{\rm GUT}(\rm MSSM) \approx 0.04$) [33]. Possible resolutions of this mismatch have been proposed. These include: (i) utilizing the idea of *string-duality* [34] which allows a lowering of M_{st} compared to the value shown above, or alternatively (ii) the idea of the so-called "Extended Supersymmetric Standard Model" (ESSM) that assumes the existence of two vector-like families, transforming as $(\mathbf{16} + \overline{\mathbf{16}})$ of $SO(10)$, with masses of order one TeV [35], in addition to the three chiral families. The latter leads to a semi-perturbative unification by raising $\alpha_{\rm GUT}$ to about 0.25-0.3. Simultaneously, it raises M_X, in two loop, to about $(1/2 - 2) \times 10^{17}$ GeV. (Other mechanisms resolving the mismatch are reviewed in Ref. [36]). In practice, a combination of the two mechanisms mentioned above may well be relevant.[5]

[5] I have in mind the possibility of string-duality [34] lowering M_{st} for the case of semi-perturbative unification in ESSM (for which $\alpha_{st} \approx 0.25$, and thus, without the use of string-duality, M_{st} would have been about 10^{18} GeV) to a value of about (1-2)$\times 10^{17}$ GeV (say), and semi-perturbative unification [35] raising the MSSM value of M_X to about 5×10^{16} GeV $\approx M_{st}$ (1/2 to 1/4) (say). In this case, an intermediate symmetry like $G(224)$ emerging at M_{st} would be effective only within the short gap between M_{st} and M_X, where it would break into $G(213)$. Despite this short gap, one would still have the benefits of $SU(4)$-color that are needed to understand neutrino masses (see Section 3), and to implement baryogenesis via leptogenesis. At the same time, since the gap is so small, the couplings of $G(224)$, unified at M_{st} would remain essentially so

While the mismatch can thus quite plausibly be removed for a non-GUT string-derived symmetry like $G(224)$ or $G(213)$, a GUT symmetry like $SU(5)$ or $SO(10)$ would have an advantage in this regard because it would keep the gauge couplings together between M_{st} and M_X (even if $M_X \sim M_{st}/20$), and thus not even encounter the problem of a mismatch between the two scales. A supersymmetric four dimensional GUT-solution [like $SU(5)$ or $SO(10)$], however, has a possible disadvantage as well, because it needs certain color triplets to become superheavy by the so-called doublet-triplet splitting mechanism in order to avoid the problem of rapid proton decay. However, no such mechanism has emerged yet, in string theory, for the four-dimensional GUT-like solutions [37].[6]

Non-GUT string solutions, based on symmetries like $G(224)$ or $G(2113)$ for example, have a distinct advantage in this regard, in that the dangerous color triplets, which would induce rapid proton decay, are often naturally projected out for such solutions [15, 17, 38]. Furthermore, the non-GUT solutions invariably possess new "flavor" gauge symmetries, which distinguish between families and also among members within a family. These symmetries are immensely helpful in explaining qualitatively the observed fermion mass-hierarchy (see e.g. Ref. [38]) and resolving the so-called naturalness problems of supersymmetry such as those pertaining to the issues of squark-degeneracy [39], CP violation [40] and quantum gravity-induced rapid proton decay [41].

Weighing the advantages and possible disadvantages of both, it seems hard at present to make a priori a clear choice between a GUT versus a non-GUT string-solution. As expressed elsewhere [42], it therefore seems prudent to keep both options open and pursue their phenomenological consequences. Given the advantages of $G(224)$ or $SO(10)$ in understanding the neutrino masses and implementing leptogenesis (see Sections 3 and 4), I will thus proceed by assuming that either a suitable four dimensional $G(224)$-solution with the scale M_X being close to M_{st} (see footnote 5), or a realistic four-dimensional $SO(10)$-solution with the desired mechanism for doublet-triplet splitting, emerges effectively from an underlying string theory, at the "conventional" string-scale $M_{st} \sim 10^{17} - 10^{18}$ GeV, and that the $G(224)/SO(10)$ symmetry in turn breaks spontaneously at the conventional GUT-scale of $M_X \sim 2 \times 10^{16}$ GeV (or at $M_X \sim 5 \times 10^{16}$ GeV for the case of ESSM, as discussed in footnote 4) to the standard model symmetry $G(213)$. The extra dimensions of string/M-theory are assumed to be tiny with sizes $\leq M_X^{-1} \sim 10^{-30}$ cm, so as not to disturb the successes of GUT. In short, I assume that essentially *the conventional (good old) picture of grand unification, proposed and developed sometime ago [5, 6, 7, 10], holds as a good effective theory above the unification scale M_X and up to some high scale $M \lesssim M_{st}$, with the added presumption that it may have its origin from the string/M-theory.*[7]

We will see that with the broad assumption mentioned above, an economical and predictive framework emerges, which successfully accounts for a host of observed phenomena pertaining to the masses and the mixings of all fermions, including neutrinos, and the baryon

at M_X, so as to match with the "observed" coupling unification, of the type suggested in Ref. [35].

[6]Some alternative mechanisms for doublet-triplet splitting, and for suppression of the $d = 5$ proton decay operators have been proposed in the context of higher dimensional theories. These will be mentioned briefly in Section 5.

[7]Alternative scenarios such as those based on TeV-scale large extra dimensions [43] or string-scale being at a few TeV [44], or submillimeter-size even larger extra dimensions with the fundamental scale of quantum gravity being a few TeV [45], though intriguing, do not seem to provide simple explanations of these features: (a), (b), and (c). They will be mentioned briefly in Section 5.2.5.

asymmetry of the universe. It also makes some crucial testable predictions for proton decay.

3 Link Between Fermion Masses and Neutrino Oscillations within a $G(224)/SO(10)$ Framework

Following Ref. [11], I now present a simple and predictive fermion mass-matrix based on $SO(10)$ or the $G(224)$-symmetry.[8] One can obtain such a mass mass-matrix for the fermions by utilizing only the minimal Higgs system that is needed to break the gauge symmetry $SO(10)$ to $SU(3)^c \times U(1)_{em}$. It consists of the set:

$$H_{\text{minimal}} = \left\{\mathbf{45_H}, \mathbf{16_H}, \overline{\mathbf{16}}_\mathbf{H}, \mathbf{10_H}\right\} \tag{6}$$

Of these, the VEV of $\langle \mathbf{45_H} \rangle \sim M_X$ breaks $SO(10)$ in the B-L direction to $G(2213) = SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times SU(3)^c$, and those of $\langle \mathbf{16_H} \rangle = \langle \overline{\mathbf{16}}_\mathbf{H} \rangle$ along $\langle \tilde{\bar{\nu}}_{RH} \rangle$ and $\langle \tilde{\nu}_{RH} \rangle$ break $G(2213)$ into the SM symmetry $G(213)$ at the unification-scale M_X. Now $G(213)$ breaks at the electroweak scale by the VEV of $\langle \mathbf{10_H} \rangle$ to $SU(3)^c \times U(1)_{em}$.[9]

The question is: can the minimal Higgs system provide a realistic pattern for fermion masses and mixings? Now $\mathbf{10_H}$ (even several $\mathbf{10}$'s) can not provide certain desirable features — i.e. family-antisymmetry and (B-L)-dependence in the mass matrices — which are, however, needed respectively to suppress V_{cb} while enhancing $\theta_{\nu_\mu \nu_\tau}$ on the one hand, and accounting for features such as $m_\mu^0 \neq m_s^0$ on the other hand (see e.g. Ref. [31], Sec. 5). Furthermore, a single $\mathbf{10_H}$ cannot generate CKM mixings. At the same time, $\mathbf{10_H}$ is the only multiplet among the ones in the minimal Higgs system (Eq. (6)) which can have cubic couplings with the matter fermions which are in the $\mathbf{16}$'s. This impasse disappears as soon as one allows for not only cubic but also effective non-renormalizable quartic couplings of the minimal set of Higgs fields with the fermions. Such effective couplings can of course arise quite naturally through exchanges of superheavy states (e.g. those in the string-tower or those having GUT-scale masses) involving renormalizable couplings, and/or through quantum gravity.

The 3×3 Dirac masses matrices for the four sectors (u, d, l, ν) proposed in Ref. [11] are motivated in part by the group theory of $SO(10)/G(224)$, which severely restricts the effective cubic and quartic couplings (and thus the associated mass-patterns), for the minimal Higgs system. They are also motivated in part by the notion that flavor symmetries [47] distinguishing between the three families lead to a hierarchical pattern for the mass matrices (i.e. with the element "33" $\gg$ "23" $\gg$ "22" $\gg$ "12" $\gg$ "11" etc.), *so that the lighter family gets its mass primarily through its mixing with the heavier ones.* It turns out that the allowed forms of effective couplings and the corresponding pattern of mass-matrices, satisfying the

[8] I will present the Higgs system for $SO(10)$. The discussion would remain essentially unaltered if one uses the corresponding $G(224)$-submultiplets instead.

[9] Large dimensional tensorial multiplets of $SO(10)$ like$\mathbf{126_H}$, $\overline{\mathbf{126}}_H$, $\mathbf{120_H}$, and $\mathbf{54_H}$ are not used for the purpose in part because they do not seem to arise at least in weakly interacting heterotic string solutions [46], and in part because they tend to give too large threshold corrections to $\alpha_3(m_Z)$ (typically exceeding 20%), which would render observed coupling unification fortuitous [see e.g. discussions in Appendix D of Ref. [11]].

constraints of group-theory and flavor-hierarchy (as above), are rather unique, barring a few discrete variants. The mass matrices proposed in Ref. [11] are as follows:[10] [11]

$$M_u = \begin{bmatrix} 0 & \epsilon' & 0 \\ -\epsilon' & 0 & \sigma+\epsilon \\ 0 & \sigma-\epsilon & 1 \end{bmatrix} \mathcal{M}_u^0; \qquad M_d = \begin{bmatrix} 0 & \eta'+\epsilon' & 0 \\ \eta'-\epsilon' & 0 & \eta+\epsilon \\ 0 & \eta-\epsilon & 1 \end{bmatrix} \mathcal{M}_d^0$$
$$M_\nu^D = \begin{bmatrix} 0 & -3\epsilon' & 0 \\ -3\epsilon' & 0 & \sigma-3\epsilon \\ 0 & \sigma+3\epsilon & 1 \end{bmatrix} \mathcal{M}_u^0; \quad M_l = \begin{bmatrix} 0 & \eta'-3\epsilon' & 0 \\ \eta'+3\epsilon' & 0 & \eta-3\epsilon \\ 0 & \eta+3\epsilon & 1 \end{bmatrix} \mathcal{M}_d^0 \tag{7}$$

These matrices are defined in the gauge basis and are multiplied by $\bar{\Psi}_L$ on left and Ψ_R on right. For instance, the row and column indices of M_u are given by $(\bar{u}_L, \bar{c}_L, \bar{t}_L)$ and (u_R, c_R, t_R) respectively. Note the group-theoretic up-down and quark-lepton correlations: the same σ occurs in M_u and M_ν^D, and the same η occurs in M_d and M_l. It will become clear that the ϵ and ϵ' entries are proportional to B-L and are antisymmetric in the family space (as shown above). Thus, the same ϵ and ϵ' occur in both (M_u and M_d) and also in (M_ν^D and M_l), but $\epsilon \to -3\epsilon$ and $\epsilon' \to -3\epsilon'$ as $q \to l$. Such correlations result in an enormous reduction of parameters and thus in increased predictivity. Although the entries σ, η, ϵ, η', and ϵ' will be treated as parameters, consistent with assignment of flavor-symmetry charges (see below), we would expect them to be hierarchical with $(\sigma, \eta, \epsilon) \sim 1/10$ and $(\eta', \epsilon') \sim 10^{-3} - 10^{-4}$ (say). Such a hierarchical pattern for the mass-matrices can be obtained, using a minimal Higgs system $\mathbf{45}_H, \mathbf{16}_H, \overline{\mathbf{16}}_H$ and $\mathbf{10}_H$ and a singlet S of $SO(10)$, through effective couplings as follows [49]:

$$\begin{aligned}
\mathcal{L}_{\rm Yuk} &= h_{33}\mathbf{16}_3\mathbf{16}_3\mathbf{10}_H \\
&+ \left[h_{23}\mathbf{16}_2\mathbf{16}_3\mathbf{10}_H(S/M) + a_{23}\mathbf{16}_2\mathbf{16}_3\mathbf{10}_H(\mathbf{45}_H/M')(S/M)^p + g_{23}\mathbf{16}_2\mathbf{16}_3\mathbf{16}_H^d(\mathbf{16}_H/M'')(S/M)^q\right] \\
&+ \left[h_{22}\mathbf{16}_2\mathbf{16}_2\mathbf{10}_H(S/M)^2 + g_{22}\mathbf{16}_2\mathbf{16}_2\mathbf{16}_H^d(\mathbf{16}_H/M'')(S/M)^{q+1}\right] \\
&+ \left[g_{12}\mathbf{16}_1\mathbf{16}_2\mathbf{16}_H^d(\mathbf{16}_H/M'')(S/M)^{q+2} + a_{12}\mathbf{16}_1\mathbf{16}_2\mathbf{10}_H(\mathbf{45}_H/M')(S/M)^{p+2}\right]
\end{aligned} \tag{8}$$

Typically we expect M', M'' and M to be of order $M_{\rm string}$ [50]. The VEV's of $\langle \mathbf{45}_H \rangle$ (along B-L), $\langle \mathbf{16}_H \rangle = \langle \overline{\mathbf{16}}_H \rangle$ (along standard model singlet sneutrino-like component) and of the SO(10)-singlet $\langle S \rangle$ are of the GUT-scale, while those of $\mathbf{10}_H$ and of the down type $\mathrm{SU}(2)_L$-doublet component in $\mathbf{16}_H$ (denoted by $\mathbf{16}_H^d$) are of the electroweak scale [11, 51]. Depending upon whether $M'(M'') \sim M_{\rm GUT}$ or $M_{\rm string}$ (see comment in [50]), the exponent $p(q)$ is either one or zero [52].

The entries 1 and σ arise respectively from h_{33} and h_{23} couplings, while $\hat{\eta} \equiv \eta - \sigma$ and η' arise respectively from g_{23} and g_{12}-couplings. The (B-L)-dependent antisymmetric entries ϵ

[10] The zeros in "11", "13", "31", and "22" elements signify that they are relatively small. For instance, the "22"-elements are set to zero because (restricted by flavor symmetries, see below), they are meant to be less than ("23")("32")/"33" $\sim 10^{-2}$, and thus unimportant for our purposes [11]. Likewise, for the other "zeros."

[11] A somewhat analogous pattern, also based on $SO(10)$, has been proposed by C. Albright and S. Barr [AB] [48]. One major difference between the work of AB and that of BPW [11] is that the former introduces the so-called "lop-sided" pattern in which some of the "23" elements are even greater than the "33" element; in BPW on the otherhand, the pattern is consistently hierarchical with individual "23" elements (like η, ϵ and σ) being much smaller in magnitude than the "33" element of 1.

and ϵ' arise respectively from the a_{23} and a_{12} couplings. This is because, with $\langle \mathbf{45}_H \rangle \propto$ B-L, the product $\mathbf{10}_H \times \mathbf{45}_H$ contributes as a $\mathbf{120}$, whose coupling is family-antisymmetric. Thus, for the minimal Higgs system (see Eq. 6), (B-L)-dependence can enter only through family off-diagonal couplings of $\mathbf{10_H} \cdot \mathbf{45_H}$ as in a_{23} and a_{12}-terms. *Thus, for such a system, the diagonal "33" entries are necessarily (B-L)-independent (as shown in Eq. (7)). This in turn makes the relations like $m_b(M_X) \approx m_\tau$ (barring corrections of order ϵ^2 [11]) robust.* This feature would, however, be absent if one had used $\overline{\mathbf{126}}_\mathbf{H}$, whose coupling is family-symmetric and can give (B-L) dependent contributions to the "33"-elements.

As alluded to above, such a hierarchical form of the mass-matrices, with h_{33}-term being dominant, is attributed in part to flavor gauge symmetry(ies) that distinguishes between the three families [53], and in part to higher dimensional operators involving for example $\langle \mathbf{45}_H \rangle / M'$ or $\langle \mathbf{16}_H \rangle / M''$, which are supressed by $M_{\rm GUT}/M_{\rm string} \sim 1/10$, if M' and/or $M'' \sim M_{\rm string}$.

To discuss the neutrino sector one must specify the Majorana mass-matrix of the RH neutrinos as well. These arise from the effective couplings of the form [54]:

$$\mathcal{L}_{\rm Maj} = f_{ij} \mathbf{16}_i \mathbf{16}_j \overline{\mathbf{16}}_H \overline{\mathbf{16}}_H / M \tag{9}$$

where the f_{ij}'s include appropriate powers of $\langle S \rangle / M$, in accord with flavor charge assignments of $\mathbf{16}_i$ (see [53]). For the f_{33}-term to be leading, we must assign the charge $-a$ to $\overline{\mathbf{16}}_H$. This leads to a hierarchical form for the Majorana mass-matrix [11]:

$$M_R^\nu = \begin{bmatrix} x & 0 & z \\ 0 & 0 & y \\ z & y & 1 \end{bmatrix} M_R \tag{10}$$

Following the flavor-charge assignments given in [53], we expect $|y| \sim \langle S/M \rangle \sim 1/10$, $|z| \sim (\langle S/M \rangle)^2 \sim 10^{-2}$(1 to 1/2), $|x| \sim (\langle S/M \rangle)^4 \sim (10^{-4}\text{-}10^{-5})$ (say). The "22" element (not shown) is $\sim (\langle S/M \rangle)^2$ and its magnitude is taken to be $< |y^2/3|$, while the "12" element (not shown) is $\sim (\langle S/M \rangle)^3$. We expect

$$M_R = f_{33} \langle \overline{\mathbf{16}}_\mathbf{H} \rangle^2 / M_{\rm string} \approx 10^{15}\ {\rm GeV}(1/2 - 2) \tag{11}$$

where we have put $\langle \overline{\mathbf{16}}_\mathbf{H} \rangle \approx M_X \approx 2 \times 10^{16}$ GeV, $M_{\rm string} \approx 4 \times 10^{17}$ GeV [33], and $f_{33} \approx 1$, and have allowed for an uncertainty by a factor of 2 in the estimate around a centrally expected value of about 10^{15}GeV. Allowing for 2-3 family-mixing in the Dirac and the Majorana sectors as in Eqs. 7 and 10, the seesaw mechanism leads to [11]:

$$m(\nu_3) \approx B \frac{m(\nu^\tau_{\rm Dirac})^2}{M_R} \tag{12}$$

The quantity B represents the effect of 2-3 mixing and is given by $B = (\sigma + 3\epsilon)(\sigma + 3\epsilon - 2y)/y^2$ (see Eq. (24) of Ref. [11]). Thus B is fully calculable within the model once the parameters σ, η, ϵ, and y are determined in terms of inputs involving some quark and lepton masses (as noted below). In this way, one obtains $B \approx (2.9 \pm 0.5)$. The Dirac mass of the tau-neutrino is obtained by using the $SU(4)$-color relation (see Eq. (2)): $m(\nu^\tau_{\rm Dirac}) \approx m_{\rm top}(M_X) \approx 120$ GeV. One thus obtains from Eq. (12):

$$m(\nu_3) \approx \frac{(2.9)(120\ {\rm GeV})^2}{10^{15}\ {\rm GeV}} (1/2 - 2) \approx (1/24\ {\rm eV})(1/2 - 2) \tag{13}$$

Noting that for hierarchical entries — i.e. for (σ, ϵ, and y) $\sim 1/10$ — one naturally obtains a hierarchical spectrum of neutrino-masses: $m(\nu_1) \lesssim m(\nu_2) \sim (1/10)m(\nu_3)$, we thus get:

$$\left[\sqrt{\Delta m^2_{23}}\right]_{\text{Theory}} \approx m(\nu_3) \approx (1/24 \text{ eV})(1/2 - 2) \tag{14}$$

This agrees remarkably well with the SuperK value of $(\sqrt{\Delta m^2_{23}})_{\text{SK}}(\approx 1/20 \text{ eV})$, which lies in the range of nearly (1/15 to 1/30) eV. As mentioned in the introduction, the success of this prediction provides clear support for (i) the existence of ν_R, (ii) the notion of $SU(4)$-color symmetry that gives $m(\nu^\tau_{\text{Dirac}})$, (iii) the SUSY unification-scale that gives M_R, and (iv) the seesaw mechanism.

We note that alternative symmetries such as $SU(5)$ would have no compelling reason to introduce the ν_R's. Even if one did introduce ν^i_R by hand, there would be no symmetry to relate the Dirac mass of ν_τ to the top quark mass. Thus $m(\nu^\tau_{\text{Dirac}})$ would be an arbitrary parameter in $SU(5)$. Furthermore, without B-L as a local symmetry, the Majorana masses of the RH neutrinos, which are singlets of $SU(5)$, can naturally be of order string scale $\sim 4 \times 10^{17}$ GeV (say). That would, however, give too small a mass for $m(\nu_3)$ ($< 10^{-4}$ eV) compared to the SuperK value.

Other effective symmetries such as $[SU(3)]^3]$ [26] and $SU(2)_L \times SU(2)_R \times U(1)_{B-L} \times SU(3)^C$ [27] would give ν_R and B-L as a local symmetry, but not the desired $SU(4)$-color mass-relations: $m(\nu^\tau_{\text{Dirac}}) \approx m_t(M_X)$ and $m_b(M_X) \approx m_\tau$. Flip $SU(5) \times U(1)$ [28] on the other hand would yield the desired features for the neutrino-system, but not the b-τ mass relation. Thus, combined with the observed b/τ mass-ratio, the SuperK data on atmospheric neutrino oscillation seems to clearly select out the effective symmetry in 4D being either $G(224)$ or $SO(10)$, as opposed to the other alternatives mentioned above. *It is in this sense that the neutrinos, by virtue of their tiny masses, provide crucial information on the unification-scale as well as on the nature of the unification-symmetry in 4D, as alluded to in the introduction.*

Ignoring possible phases in the parameters and thus the source of CP violation for a moment, as was done in Ref. [11], the parameters ($\sigma, \eta, \epsilon, \epsilon', \eta', \mathcal{M}^0_u, \mathcal{M}^0_D$, and y) can be determined by using, for example, $m_t^{\text{phys}} = 174$ GeV, $m_c(m_c) = 1.37$ GeV, $m_S(1 \text{ GeV}) = 110$-116 MeV, $m_u(1 \text{ GeV}) = 6$ MeV, the observed masses of e, μ, and τ and $m(\nu_2)/m(\nu_3) \approx 1/(7 \pm 1)$ (as suggested by a combination of atmospheric [1] and solar neutrino data [2], the latter corresponding to the LMA MSW solution, see below) as inputs. One is thus led, *for this CP conserving case*, to the following fit for the parameters, and the associated predictions [11]. [In this fit, we leave the small quantities x and z in M^ν_R undetermined and proceed by assuming that they have the magnitudes suggested by flavor symmetries (i.e., $x \sim (10^{-4}\text{-}10^{-5})$ and $z \sim 10^{-2}(1 \text{ to } 1/2)$ (see remarks below Eq. (10))]:

$$\begin{aligned} &\sigma \approx 0.110, \quad \eta \approx 0.151, \quad \epsilon \approx -0.095, \quad |\eta'| \approx 4.4 \times 10^{-3}, \\ &\qquad \epsilon' \approx 2 \times 10^{-4}, \qquad \mathcal{M}^0_u \approx m_t(M_X) \approx 120 \text{ GeV}, \\ &\mathcal{M}^0_D \approx m_b(M_X) \approx 1.5 \text{ GeV}, \qquad y \approx -(1/17). \end{aligned} \tag{15}$$

These in turn lead to the following predictions for the quarks and light neutrinos [11]:[12]

$$
\begin{array}{l}
m_b(m_b) \approx (4.7\text{-}4.9)\ \text{GeV}, \\
m(\nu_3) \approx (1/24\ \text{eV})(1/2-2), \\
V_{cb} \approx \left| \sqrt{\frac{m_s}{m_b}} \left|\frac{\eta+\epsilon}{\eta-\epsilon}\right|^{1/2} - \sqrt{\frac{m_c}{m_t}} \left|\frac{\sigma+\epsilon}{\sigma-\epsilon}\right|^{1/2} \right| \approx 0.044, \\
\left\{ \begin{array}{l} \theta^{\text{osc}}_{\nu_\mu\nu_\tau} \approx \left| \sqrt{\frac{m_\mu}{m_\tau}} \left|\frac{\eta-3\epsilon}{\eta+3\epsilon}\right|^{1/2} + \sqrt{\frac{m_{\nu_2}}{m_{\nu_3}}} \right| \approx |0.437 + (0.378 \pm 0.03)|, \\ \text{Thus, } \sin^2 2\theta^{\text{osc}}_{\nu_\mu\nu_\tau} \approx 0.99, \quad (\text{for } m(\nu_2)/m(\nu_3) \approx 1/7), \end{array} \right. \\
V_{us} \approx \left| \sqrt{\frac{m_d}{m_s}} - \sqrt{\frac{m_u}{m_c}} \right| \approx 0.20, \\
\left|\frac{V_{ub}}{V_{cb}}\right| \approx \sqrt{\frac{m_u}{m_c}} \approx 0.07, \\
m_d(1\ \text{GeV}) \approx 8\ \text{MeV}, \\
\theta^{\text{osc}}_{\nu_e\nu_\mu} \approx 0.06 \text{ (ignoring non-seesaw contributions; see, however, remarks below)}
\end{array}
\tag{16}
$$

The Majorana masses of the RH neutrinos (N_i) are given by:[13]

$$
\begin{array}{rcl}
M_3 & \approx & M_R \approx 10^{15}\ \text{GeV (1/2-1)}, \\
M_2 & \approx & |y^2| M_3 \approx (2.5\times 10^{12}\ \text{GeV})(1/2\text{-}1), \\
M_1 & \approx & |x - z^2| M_3 \sim (1/2\text{-}2) 10^{-5} M_3 \sim 10^{10}\ \text{GeV}(1/4\text{-}2).
\end{array}
\tag{17}
$$

Note that we necessarily have a hierarchical spectrum for the light as well as the heavy neutrinos (see discussions below on m_{ν_1}). Leaving out the ν_e-ν_μ oscillation angle for a moment, it seems remarkable that the first seven predictions in Eq. (16) agree with observations, to within 10%. Particularly intriguing is the (B-L)-dependent *group-theoretic correlation* between the contribution from the first term in V_{cb} and that in $\theta^{\text{osc}}_{\nu_\mu\nu_\tau}$, which explains simultaneously why one is small (V_{cb}) and the other is large ($\theta^{\text{osc}}_{\nu_\mu\nu_\tau}$) [55]. That in turn provides some degree of confidence in the gross structure of the mass-matrices.

As regards ν_e-ν_μ and ν_e-ν_τ oscillations, the standard seesaw mechanism would typically lead to rather small angles as in Eq. (16), within the framework presented above [11]. It has, however, been noted recently [31] that small intrinsic (non-seesaw) masses $\sim 10^{-3}$ eV of the LH neutrinos can arise quite plausibly through higher dimensional operators of the form [56]: $W_{12} \supset \kappa_{12} \mathbf{16}_1 \mathbf{16}_2 \mathbf{16}_H \mathbf{16}_H \mathbf{10}_H \mathbf{10}_H / M^3_{\text{eff}}$, without involving the standard seesaw mechanism [20]. One can verify that such a term would lead to an intrinsic Majorana mixing mass term of the form $m^0_{12} \nu^e_L \nu^\mu_L$, with a strength given by $m^0_{12} \approx \kappa_{12} \langle \mathbf{16}_H \rangle^2 (175\ \text{GeV})^2 / M^3_{\text{eff}} \sim (1.5\text{-}6) \times 10^{-3}$ eV, for $\langle \mathbf{16}_H \rangle \approx (1\text{-}2) M_{\text{GUT}}$ and $\kappa_{12} \sim 1$, if $M_{\text{eff}} \sim M_{\text{GUT}} \approx 2 \times 10^{16}$ GeV [57]. Such an intrinsic Majorana $\nu_e \nu_\mu$ mixing mass $\sim$ few$\times 10^{-3}$ eV, though small compared to $m(\nu_3)$, is still much larger than what one would generically get for the corresponding term from the standard seesaw mechanism [as in Ref. [11]]. Now, the diagonal ($\nu^\mu_L \nu^\mu_L$) mass-term, arising from the standard seesaw mechanism is expected to be $\sim (3-8) \times 10^{-3}$ eV for a natural value of $|y| \approx 1/20\text{-}1/15$, say [11]. Thus, taking the net values of $m^0_{22} \approx 7 \times 10^{-3}$

[12]These predictions are based on the fact that the pattern given in Eq. 7 leads to $m_b(M_X) \approx m_\tau(1 - 8\epsilon^2)$. They also reflect the recent trend in the atmospheric and solar neutrino data which suggests $m(\nu_2)/m(\nu_3) \approx 1/7$.

[13]The range in M_3 and M_2 is constrained by the values of $m(\nu_3)$ and $m(\nu_2)$ suggested by the atmospheric and solar neutrino data.

eV, $m_{12}^0 \approx 3 \times 10^{-3}$ eV as above and $m_{11}^0 \ll 10^{-3}$ eV, which are all plausible, we obtain $m_{\nu_2} \approx 7 \times 10^{-3}$ eV, $m_{\nu_1} \sim$ (1 to few) $\times 10^{-3}$ eV, so that $\Delta m_{12}^2 \approx 5 \times 10^{-5}$ eV^2 and $\sin^2 2\theta_{12}^{\rm osc} \approx 0.6 - 0.7$. These go well with the LMA MSW solution of the solar neutrino problem.

In summary, *the intrinsic non-seesaw contribution* to the Majorana masses of the LH neutrinos (neglected in making the predictions of Eq. (16)) can plausibly have the right magnitude for ν_e-ν_μ mixing so as to lead to the LMA solution within the $G(224)/SO(10)$-framework, without upsetting the successes of the first seven predictions in Eq. (16). [In contrast to the near maximality of the ν_μ-ν_τ oscillation angle, however, which emerges as a compelling prediction of the framework [11], the LMA solution, as obtained above, should be regarded only as a consistent possibility, rather than as a compelling prediction, within this framework.]

It is worth noting at this point that in a theory leading to Majorana masses of the LH neutrinos as above, *one would of course expect the neutrinoless double beta decay process (like $n + n \to ppe^-e^-$), satisfying $|\Delta L| = 2$ and $|\Delta B| = 0$, to occur at some level.* Search for this process is most important because it directly tests a fundamental conservation law and can shed light on the Majorana nature of the neutrinos, as well as on certain CP violating phases in the neutrino-system (assuming that the process is dominated by neutrino-exchange). The crucial parameter which controls the strength of this process is given by $m_{ee} \equiv |\sum_i m_{\nu_i} U_{ei}^2|$. With a non-seesaw contribution leading to $m_{\nu_1} \sim$ few $\times 10^{-3}$ eV, $m_{\nu_2} \approx 7 \times 10^{-3}$ eV, $\sin^2 2\theta_{12} \approx 0.6 - 0.7$, and an expected value for $\sin\theta_{13} \sim m_{13}^0/m_{33}^0 \sim (1-5) \times 10^{-3}$ eV $/(5 \times 10^{-2}$ eV $) \sim (0.02 - 0.1)$, one would expect $m_{ee} \approx (1-5) \times 10^{-3}$ eV. Such a strength, though compatible with current limits [58], would be accessible if the current sensitivity is improved by about a factor of 50–100. Improving the sensitivity to this level would certainly be most desirable.

I would now like to turn to a discussion of leptogenesis within the $G(224)/SO(10)$-framework for fermion masses and mixings presented above. Before discussing leptogenesis, we need to discuss, however, the origin of CP violation within the same framework. The discussion so far has ignored, for the sake of simplicity, possible CP violating phases in the parameters (σ, η, ϵ, η', ϵ', $\zeta_{22}^{u,d}$, y, z, and x) of the Dirac and Majorana mass matrices [Eqs. (7), and (10)]. In general, however, these parameters can and generically will have phases [59]. Some combinations of these phases enter into the CKM matrix and define the Wolfenstein parameters ρ_W and η_W [60], which in turn induce CP violation by utilizing the standard model interactions. As observed in Ref. [61], an additional and potentially important source of CP and flavor violations (as in $K^0 \leftrightarrow \bar{K}^0$, $B_{d,s} \leftrightarrow \bar{B}_{d,s}$, $b \to s\bar{s}s$, etc. transitions) arise in the model through supersymmetry [62], involving squark and gluino loops (box and penguin), simply because of the embedding of MSSM within a string-unified $G(224)$ or $SO(10)$-theory near the GUT-scale, and the assumption that primordial SUSY-breaking occurs near the string scale ($M_{\rm string} > M_{\rm GUT}$) [63]. It is shown in [61] that complexification of the parameters (σ, η, ϵ, η', ϵ', etc.), through introduction of phases $\sim$ 1/20-1/2 (say) in them, can still preserve the successes of the predictions as regards fermion masses and neutrino oscillations shown in Eq. (16), as long as one maintains nearly the magnitudes of the real parts of the parameters and especially their relative signs as obtained in Ref. [11] and shown in Eq. (15) [64]. Such a picture is also in accord with the observed features of CP and flavor violations in ϵ_K, Δm_{Bd}, and asymmetry parameter in $B_d \to J/\Psi + K_s$, while

predicting observable new effects in processes such as $B_s \to \bar{B}_s$ and $B_d \to \Phi + K_s$ [61].

We therefore proceed to discuss leptogenesis concretely within the framework presented above by adopting the Dirac and Majorana fermion mass matrices as shown in Eqs. (7) and (10) and assuming that the parameters appearing in these matrices can have natural phases $\sim$ 1/20-1/2 (say) with either sign up to addition of $\pm\pi$, *while their real parts have the relative signs and nearly the magnitudes given in Eq. (16).*

4 Leptogenesis

Finally, the observed matter-antimatter asymmetry of the universe provides an additional important clue to physics at truly short distances. This issue has taken a new turn since the discovery of the non-perturbative electroweak sphaleron effects [13], which violate B+L but conserve B-L. These remain in thermal equilibrium in the temperature range of 200 GeV to about 10^{12} GeV. As a result, they efficiently erase any pre-existing baryon/lepton asymmetry that satisfies $\Delta(\mathrm{B}+\mathrm{L}) \neq 0$, but $\Delta(\mathrm{B}-\mathrm{L}) = 0$. This is one reason why standard GUT-baryogenesis satisfying $\Delta(\mathrm{B}-\mathrm{L}) = 0$ (as in minimal SU(5)) becomes irrelevant to the observed baryon asymmetry of the universe[14]. On the other hand, purely electroweak baryogenesis based on the sphaleron effects - although a priori an interesting possibility - appears to be excluded for the case of the standard model without supersymmetry, and highly constrained as regards the available parameter space for the case of the supersymmetric standard model, owing to LEP lower limit on Higgs mass $\geq$ 114 GeV. As a result, in the presence of electroweak sphalerons, baryogenesis via leptogenesis [12] appears to be an attractive and promising mechanism to generate the observed baryon asymmetry of the universe.

To discuss leptogenesis concretely within the G(224)/SO(10) - framework presented above, I follow the discussion of Ref.[22] and first consider the case of thermal leptogenesis. In the context of an inflationary scenario [65], with a plausible reheat temperature $T_{RH} \sim$ (1 to few) $\times\, 10^9$ GeV (say), one can avoid the well known gravitino problem if $m_{3/2} \sim$ (1 to 2) TeV [66] and yet produce the lightest heavy neutrino N_1 efficiently from the thermal bath if $M_1 \sim$ (3 to 5) $\times\, 10^9$ GeV (say), in accord with Eq. (17) [N_2 and N_3 are of course too heavy to be produced at $T \sim T_{RH}$]. Given lepton number (and B-L) violation occurring through the Majorana mass of N_1, and C and CP violating phases in the Dirac and/or Majorana fermion mass-matrices as mentioned above, the out-of-equilibrium decays of N_1 (produced from the thermal bath) into $l + H$ and $\bar{l} + \bar{H}$ and into the corresponding SUSY modes $\tilde{l}+\tilde{H}$ and $\bar{\tilde{l}}+\bar{\tilde{H}}$ would produce a B-L violating lepton asymmetry; so also would the decays of $\tilde{N}_1$ and $\bar{\tilde{N}}_1$. Part of this asymmetry would of course be washed out due to inverse decays and lepton number violating 2$\leftrightarrow$2-scatterings. I will assume this commonly adopted mechanism for the so-called thermal leptogenesis (At the end, I will consider an interesting alternative that would involve non-thermal leptogenesis). This mechanism has been extended to incorporate supersymmetry by several authors (see e.g., [67, 68, 69]). The net lepton asymmetry of the universe [$Y_L \equiv (n_L - n_{\bar{L}})/s$] arising from decays of N_1 into $l+H$

[14]Standard GUT-baryogenesis involving decays of X and Y gauge bosons (with $M_X \sim 10^{16}$ GeV) and/or of superheavy Higgs bosons is hard to realize anyway within a plausible inflationary scenario satisfying the gravitino-constraint [see e.g. E. W. Kolb and M. S. Turner, "The Early Universe", Addison-Wesley, 1990].

and $\bar{l} + \bar{H}$ as well as into the corresponding SUSY modes ($\tilde{l} + \tilde{H}$ and $\bar{\tilde{l}} + \bar{\tilde{H}}$) and likewise from ($\tilde{N}_1, \bar{\tilde{N}}_1$)-decays [67, 68, 69] is given by:

$$Y_L = \kappa\epsilon_1 \left(\frac{n_{N_1} + n_{\tilde{N}_1} + n_{\bar{\tilde{N}}_1}}{s} \right) \approx \kappa\epsilon_1/g^* \tag{18}$$

where ϵ_1 is the lepton-asymmetry produced per N_1 (or ($\tilde{N}_1 + \bar{\tilde{N}}_1$)-pair) decay (see below), and κ is the efficiency or damping factor that represents the washout effects mentioned above (thus κ incorporates the extent of departure from thermal equilibrium in N_1-decays; such a departure is needed to realize lepton asymmetry).[15] The parameter $g^* \approx 228$ is the number of light degrees of freedom in MSSM.

The lepton asymmetry Y_L is converted to baryon asymmetry, by the sphaleron effects, which is given by:

$$Y_B = \frac{n_B - n_{\bar{B}}}{s} = C\,Y_L, \tag{19}$$

where, for MSSM, $C \approx -1/3$. Taking into account the interference between the tree and loop-diagrams for the decays of $N_1 \to lH$ and $\bar{l}\bar{H}$ (and likewise for $N_1 \to \tilde{l}\tilde{H}$ and $\bar{\tilde{l}}\bar{\tilde{H}}$ modes and also for $\tilde{N}_1$ and $\bar{\tilde{N}}_1$-decays), the CP violating lepton asymmetry parameter in each of the four channels (see e.g., [68] and [69]) is given by

$$\epsilon_1 = \frac{1}{8\pi v^2 (M_D^\dagger M_D)_{11}} \sum_{j=2,3} \mathrm{Im}\left[(M_D^\dagger M_D)_{j1}\right]^2 f(M_j^2/M_1^2) \tag{20}$$

where M_D is the Dirac neutrino mass matrix evaluated in a basis in which the Majorana mass matrix of the RH neutrinos M_R^ν [see Eq. (10)] is diagonal, $v = (174\ \mathrm{GeV}) \sin\beta$ and the function $f \approx -3(M_1/M_j)$, for the case of SUSY with $M_j \gg M_1$.

Including inverse decays as well as $\Delta L \neq 0$-scatterings in the Boltzmann equations, a recent analysis [70] shows that in the relevant parameter-range of interest to us (see below), the efficiency factor (for the SUSY case) is given by [71]:

$$\kappa \approx (0.7 \times 10^{-4})(\mathrm{eV}/\tilde{m}_1) \tag{21}$$

where $\tilde{m}_1$ is an effective mass parameter (related to K [72]), and is given by [73]:

$$\tilde{m}_1 \equiv (m_D^\dagger m_D)_{11}/M_1. \tag{22}$$

Eq. (21) should hold to better than 20% (say), when $\tilde{m}_1 \gg 5 \times 10^{-4}$ eV [70] (This applies well to our case, see below).

I proceed to make numerical estimates of the lepton-asymmetry by taking the magnitudes and the relative signs of the real parts of the parameters (σ, η, ϵ, η', ϵ', and y) approximately the same as in Eq. (15), but allowing in general for natural phases in them (as in [61]).

[15]The efficiency factor mentioned above, is often expressed in terms of the parameter $K \equiv [\Gamma(N_1)/2H]_{T=M_1}$ [65]. Assuming initial thermal abundance for N_1, κ is normalized so that it is 1 if N_1's decay fully out of equilibrium corresponding to $K \ll 1$ (in practise, this actually requires $K < 0.1$).

	$\lvert\zeta_{31} - z\rvert$		
	$(1/200)(1/3)$	$(1/200)(1/4)$	$(1/200)(1/5)$
$\tilde{m}_1$(eV)	0.83×10^{-2}	0.47×10^{-2}	0.30×10^{-2}
κ	$1/73$	$1/39$	$1/24$
$Y_L/\sin(2\phi_{21})$	-11.8×10^{-11}	-22.4×10^{-11}	-36×10^{-11}
$Y_B/\sin(2\phi_{21})$	4×10^{-11}	7.5×10^{-11}	12×10^{-11}
ϕ_{21}	$\sim \pi/4$	$\sim \pi/12 - \pi/4$	$\sim \pi/18 - \pi/4$

Table 1: Baryon Asymmetry for the Case of Thermal Leptogenesis

In the evaluation of the lepton asymmetry, I allow for small "31" and "13" entries in M_ν^D, denoted by ζ_{31} and ζ_{13} respectively, which are not exhibited in Eq. (7). Following assignment of flavor-charges [53], these are expected to be of order (1/200)(1 to 1/2) (say). As such, they have no noticeable effects on fermion masses and mixings discussed above, but they can be relevant to lepton asymmetry.

Using the values of the parameters (σ, ϵ, ϵ', y, and z) determined from our consideration of fermion masses (see Eq. (15)) and the expected magnitudes of ζ_{31} and z, one obtains the following estimates (see Ref. [22] for details):

$$\begin{aligned}
\frac{\left(M_D^\dagger M_D\right)_{11}}{\left(\mathcal{M}_u^0\right)^2} &= |3\epsilon' - z(\sigma - 3\epsilon)|^2 + |\zeta_{31} - z|^2 \\
&\approx 2.5 \times 10^{-5}(1/4 \text{ to } 1/6) \qquad (23) \\
\epsilon_1 &\approx \frac{1}{8\pi}\left(\frac{\mathcal{M}_u^0}{v}\right)^2 |(\sigma + 3\epsilon) - y|^2 \sin(2\phi_{21}) \left(\frac{-3M_1}{M_2}\right) \approx -2.0 \times 10^{-6} \sin(2\phi_{21}) \qquad (24)
\end{aligned}$$

where, $\phi_{21} = \arg[(\zeta_{31} - z)(\sigma^* + 3\epsilon^* - y^*)] + (\phi_1 - \phi_2)$. Here $(\phi_1 - \phi_2)$ is a phase angle that arises from diagonalization of the Majorana mass matrix M_R^ν (see [22]). *The effective phase ϕ_{21} thus depends upon phases in both the Dirac and the Majorana mass matrices.* In writing Eq. (24), we have put $(\mathcal{M}_u^0/v)^2 \approx 1/2$, $|\sigma + 3\epsilon - y| \approx 0.13$ (see Eq. (15) and Ref. [64]), and for concreteness (for the present case of thermal leptogenesis) $M_1 \approx 4 \times 10^9$ GeV and $M_2 \approx 2 \times 10^{12}$ GeV [see Eq. (17)]. Since $|\zeta_{31}|$ and $|z|$ are each expected to be of $(1/200)(1 to 1/2)$ (say) by flavor symmetry, we have allowed for a possible mild cancellation between their contributions to $|\zeta_{31} - z|$ by putting $|\zeta_{31} - z| \approx (1/200)(1/2 - 1/5)$. The parameter $\tilde{m}_1$, given by Eq. (22), turns out to be (approximately) proportional to $|\zeta_{31} - z|^2$ [see Eq. (23)]. It is given by:

$$\tilde{m}_1 \approx |\zeta_{31} - z|^2 (\mathcal{M}_u^0)^2 / M_1 \approx (1.9 \times 10^{-2} \text{ eV})(1 \text{ to } 1/6)\left(\frac{4 \times 10^9 \text{GeV}}{M_1}\right) \qquad (25)$$

where, as before, we have put $|\zeta_{31} - z| \approx (1/200)(1/2 \text{ to } 1/5)$. The corresponding efficiency factor κ [given by Eq. (21)], lepton and baryon-asymmetries Y_L and Y_B [given by Eqs. (18) and (19)] and the requirement on the phase-parameter ϕ_{21} are listed in Table 1.

The constraint on ϕ_{21} is obtained from considerations of Big-Bang nucleosynthesis, which requires $3.7 \times 10^{-11} \lesssim (Y_B)_{BBN} \lesssim 9 \times 10^{-11}$ [23]; this is consistent with the CMB data [24], which suggests somewhat higher values of $(Y_B)_{CMB} \approx (7-10) \times 10^{-11}$ (say). We see that the first case $|\zeta_{31} - z| \approx 1/200(1/3)$ leads to a baryon asymmetry Y_B that is on the low side of the BBN-data, even for a maximal $\sin(2\phi_{21}) \approx 1$. The other cases with $|\zeta_{31} - z| \approx (1/200)(1/4 \text{ to } 1/5)$, which are of course perfectly plausible, lead to the desired magnitude of the baryon asymmetry for natural values of the phase parameter $\phi_{21} \sim (\pi/18 \text{ to } \pi/4)$. We see that, for the thermal case, the CMB data would suggest somewhat smaller values of $|\zeta_{31} - z| \sim 10^{-3}$. This constraint would be eliminated for the case of non-thermal leptogenesis.

We next consider briefly the scenario of non-thermal leptogenesis [74, 75]. In this case the inflaton is assumed to decay, following the inflationary epoch, directly into a pair of heavy RH neutrinos (or sneutrinos). These in turn decay into $l + H$ and $\bar{l} + \bar{H}$ as well as into the corresponding SUSY modes, and thereby produce lepton asymmetry, during the process of reheating. It turns out that this scenario goes well with the fermion mass-pattern of Sec. 2 [in particular see Eq. (17)] and the observed baryon asymmetry, provided $2M_2 > m_{\text{infl}} > 2M_1$, so that the inflaton decays into $2N_1$ rather than into $2N_2$ (contrast this from the case proposed in Ref. [74]). In this case, the reheating temperature (T_{RH}) is found to be much less than $M_1 \sim 10^{10}$ GeV (see below); thereby (a) the gravitino constraint is satisfied quite easily, even for a rather low gravitino-mass ~ 200 GeV (unlike in the thermal case); at the same time (b) while N_1's are produced non-thermally (and copiously) through inflaton decay, they remain out of equilibrium and the wash out process involving inverse decays and $\Delta L \neq 0$-scatterings are ineffective, so that the efficiency factor κ is 1.

To see how the non-thermal case can arise naturally, we recall that the VEV's of the Higgs fields $\Phi = (1,2,4)_H$ and $\bar{\Phi} = (1,2,\bar{4})_H$ have been utilized to (i) break $SU(2)_R$ and B-L so that G(224) breaks to the SM symmetry [5], and simultaneously (ii) to give Majorana masses to the RH neutrinos via the coupling in Eq. (9) (see e.g., Ref. [11]; for SO(10), $\bar{\Phi}$ and Φ would be in $\mathbf{16}_H$ and $\bar{\mathbf{16}}_H$ respectively). It is attractive to assume that the same Φ and $\bar{\Phi}$ (in fact their $\nu_{\text{RH}}^{\sim}$ and $\bar{\tilde{\nu}}_{\text{RH}}$-components), which acquire GUT-scale VEV's, also drive inflation [74]. In this case the inflaton would naturally couple to a pair of RH neutrinos by the coupling of Eq. (9). To implement hybrid inflation in this context, let us assume following Ref. [74], an effective superpotential $W_{\text{eff}}^{\text{infl}} = \lambda S(\bar{\Phi}\Phi - M^2) + (\text{non-ren. terms})$, where S is a singlet field [76]. It has been shown in Ref. [74] that in this case a flat potential with a radiatively generated slope can arise so as to implement inflation, with $G(224)$ broken during the inflationary epoch to the SM symmetry. The inflaton is made of two complex scalar fields (i.e., $\theta = (\delta\tilde{\nu}_H^C + \delta\bar{\tilde{\nu}}_H^C)/\sqrt{2}$ that represents the fluctuations of the Higgs fields around the SUSY minimum, and the singlet S). Each of these have a mass $m_{\text{infl}} = \sqrt{2}\lambda M$, where $M = \langle (1,2,4)_H \rangle \approx 2 \times 10^{16}$ GeV and a width $\Gamma_{\text{infl}} = \Gamma(\theta \to \Psi_{\nu_H}\Psi_{\nu_H}) = \Gamma(S \to \tilde{\nu}_H\tilde{\nu}_H) \approx [1/(8\pi)](M_1/M)^2 m_{\text{infl}}$ so that

$$T_{\text{RH}} \approx (1/7)(\Gamma_{\text{infl}} M_{\text{Pl}})^{1/2} \approx (1/7)(M_1/M)[m_{\text{infl}} M_{\text{Pl}}/(8\pi)]^{1/2} \tag{26}$$

For concreteness, take [77] $M_2 \approx 2 \times 10^{12}$ GeV, $M_1 \approx 2 \times 10^{10}$ GeV (1 to 2) [in accord with Eq. (17)], and $\lambda \approx 10^{-4}$, so that $m_{\text{infl}} \approx 3 \times 10^{12}$ GeV. We then get: $T_{\text{RH}} \approx (1.7 \times$

10^8 GeV)(1 to 2), and thus (see e.g., Sec. 8 of Ref. [65]):

$$\begin{aligned}(Y_B)_{Non-Thermal} &\approx -(Y_L/3) \\ &\approx (-1/3)[(n_{N_1} + n_{\tilde{N}_1} + n_{\tilde{\tilde{N}}_1})/s]\epsilon_1 \\ &\approx (-1/3)[(3/2)(T_{\rm RH}/m_{\rm infl})\epsilon_1] \\ &\approx (30 \times 10^{-11})(\sin 2\phi_{21})(1 \text{ to } 2)^2 \end{aligned} \tag{27}$$

Here we have used Eq. (24) for ϵ_1 with appropriate (M_1/M_2), as above. Setting $M_1 \approx 2\times 10^{10}$ for concreteness, we see that Y_B obtained above agrees with the (nearly central) observed value of $\langle Y_B\rangle^{\rm central}_{\rm BBN(CMB)} \approx (6(8.6)) \times 10^{-11}$, again for a natural value of the phase parameter $\phi_{21} \approx \pi/30(\pi/20)$. As mentioned above, one possible advantage of the non-thermal over the thermal case is that the gravitino-constraint can be met rather easily, in the case of the former (because T_{RH} is rather low $\sim 10^8$ GeV), whereas for the thermal case there is a significant constraint on the lowering of the T_{RH} (so as to satisfy the gravitino-constraint) vis a vis a raising of $M_1 \sim T_{RH}$ so as to have sufficient baryon asymmetry (note that $\epsilon_1 \propto M_1$, see Eq. (24)). Furthermore, for the non-thermal case, the dependence of Y_B on the parameter $|\zeta_{31} - z|^2$ (which arises through κ and $\tilde{m}_1$ in the thermal case, see Eqs. (21), (22), and (23)) is largely eliminated. Thus the expected magnitude of Y_B (Eq. (27) holds without a significant constraint on $|\zeta_{31} - z|$ (in contrast to the thermal case).

To conclude this part, I have considered two alternative scenarios (thermal as well as non-thermal) for inflation and leptogenesis. *We see that the $G(224)/SO(10)$ framework provides a simple and unified description of not only fermion masses and neutrino oscillations (consistent with maximal atmospheric and large solar neutrino oscillation angles) but also of baryogenesis via leptogenesis, for the thermal as well as non-thermal cases, in accord with the gravitino constraint.* Each of the features — (a) the existence of the right-handed neutrinos, (b) B-L local symmetry, (c) quark-lepton unification through SU(4)-color, (d) the seesaw mechanism, and (e) the magnitude of the supersymmetric unification-scale — plays a crucial role in realizing this unified and successful description. These features in turn point to the relevance of either the G(224) or the SO(10) symmetry being effective between the string and the GUT scales, in four dimensions. While the observed magnitude of the baryon asymmetry seems to emerge naturally from within the framework, understanding its observed sign (and thus the relevant CP violating phases) remains a challenging task.

5 Proton Decay: The Hallmark of Grand Unification

5.1 Preliminaries

Turning to proton decay, I present now the reason why the unification framework based on SUSY $SO(10)$ or SUSY $G(224)$, together with the understanding of fermion masses and mixings discussed above, strongly suggest that proton decay should be imminent.

In supersymmetric unified theories there are in general three distinct mechanisms for proton decay - two realized sometime ago and one rather recently. Briefly, they are:

1. The familiar $d = 6$ operators mediated by X and Y gauge bosons of $SU(5)$ or $SO(10)$ (Fig. 1). These lead to $e^+\pi^0$ as the dominant mode.

2. The "standard" $d = 5$ operators [78] (Fig. 2) which arise through exchanges of color triplet Higgsinos which are in the $\mathbf{5}_{\rm H} + \bar{\mathbf{5}}_{\rm H}$ of $SU(5)$ or $10_{\rm H}$ of $SO(10)$. In the presence of these operators, one crucially needs, for consistency with the empirical lower limit on proton lifetime, a suitable doublet-triplet splitting mechanism which assigns GUT-scale masses to the color triplets in the $\mathbf{10}_{\rm H}$ of $SO(10)$ while keeping the electroweak doublets light (see e.g. Ref. [11] for discussion of such mechanisms and relevant references). Following the constraints of Bose symmetry, color antisymmetry and hierarchical Yukawa couplings, these standard $d = 5$ operators lead to dominant $\bar{\nu}K^+$ and comparable $\bar{\nu}\pi^+$ modes, but in all cases to highly suppressed $e^+\pi^0$, e^+K^0 and even μ^+K^0 modes.

3. The "new" $d = 5$ operators [32] which arise (see Fig. 3) through exchanges of color triplet Higgsinos in the Higgs multiplets like $\mathbf{16}_{\rm H} + \overline{\mathbf{16}}_{\rm H}$ of $SO(10)$, which are used to give superheavy Majorana masses to the RH neutrinos. These operators generically arise through the joint effects of (a) the couplings as in Eq. (9) which assign superheavy Majorana masses to the RH neutrinos, and (b) the couplings of the form $g_{ij}\mathbf{16}_i\mathbf{16}_j\mathbf{16}_{\rm H}\mathbf{16}_{\rm H}/M$ as in Eq. (8) which are needed to generate CKM mixings (see Sec. 3). Thus these new $d = 5$ operators are directly linked not only to the masses and mixings of quarks and leptons, but also to the Majorana masses of the RH neutrinos.

The contributions of these three operators to proton decay has been considered in detail in Ref. [31], which provides an update in this regard of the results of Ref. [11]. Here, I will highlight only the main ingredients that enter into the calculations of proton decay rate, based on the three contributions listed above, and then present a summary of the main results. The reader is referred to these two references for a more detailed presentation and explanations.

Relative to other analyses, the study of proton decay carried out in Refs. [11] and [31] have the following distinctive features:

(i) Link with Fermion Masses: It systematically takes into account the link that exists between the $d = 5$ proton decay operators and the masses and mixings of all fermions including neutrinos, within a realistic $G(224)/SO(10)$-framework, as discussed in Sec. 3.

(ii) Inclusion of the standard and the new $d = 5$ operators: In particular, it includes contributions from both the standard and the new $d = 5$ operators (Fig. 3), related to the Majorana masses of the RH neutrinos. These latter, invariably omitted in the literature, are found to be generally as important as the standard ones.

(iii) Restricting GUT-scale Threshold Corrections: The study restricts GUT-scale threshold corrections to $\alpha_3(m_Z)$ so as to be in accord with the demand of "natural" coupling unification. This restriction is especially important for SUSY $SO(10)$, for which, following the mechanism of doublet-triplet splitting (see Appendix of Ref. [11]), the standard $d = 5$ operators become inversely proportional to an effective mass-scale given by $M_{\rm eff} \equiv [\lambda\langle\mathbf{45}_{\rm H}\rangle]^2/M_{10'} \sim M_X^2/M_{10'}$, rather than to the physical masses of the color-triplets in the $\mathbf{10}_{\rm H}$ of $SO(10)$. Here $M_{10'}$ represents the mass of $\mathbf{10}'_{\rm H}$, that enters into the D-T splitting mechanism through an effective coupling $\lambda\mathbf{10}_{\rm H}\mathbf{45}_{\rm H}\mathbf{10}'_{\rm H}$ in the superpotential. Now, $M_{10'}$ can be naturally suppressed compared to M_X owing to flavor symmetries, and thus $M_{\rm eff}$ can be

correspondingly larger than M_X by even two to three orders of magnitude.[16] [17]

Although $M_{\rm eff}$ can far exceed M_X, it still gets bounded from above by demanding that coupling unification, as observed,[18] should emerge as a natural prediction of the theory as opposed to being fortuitous. That in turn requires that there be no large (unpredicted) cancellation between GUT-scale threshold corrections to the gauge couplings that arise from splittings within different multiplets as well as from Planck scale physics. Following this point of view, we have argued (see Ref. [11]) that the net "other" threshold corrections to $\alpha_3(m_Z)$ arising from the Higgs and the gauge multiplets should be negative, but conservatively and quite plausibly no more than about 10%, at the electroweak scale. Such a requirement is in fact found to be well satisfied not only in magnitude but also in sign by the minimal Higgs system consisting of ($\mathbf{45}_{\rm H}$, $\mathbf{16}_{\rm H}$ $\overline{\mathbf{16}}_{\rm H}$, and $\mathbf{10}_{\rm H}$) [11]. This in turn restricts how big can be the threshold corrections to $\alpha_3(m_Z)$ that arise from (D-T) splitting (which is positive). Since the latter turns out to be proportional to $\ln(M_{\rm eff}\cos\gamma/M_X)$, we thus obtain an upper limit on $M_{\rm eff}\cos\gamma$, where $\cos\gamma \approx (\tan\beta)/(m_t/m_b)$. An upper limit on $M_{\rm eff}\cos\gamma$ thus provides an upper limit on $M_{\rm eff}$ which is inversely proportional to $\tan\beta$. In this way, our demand of natural coupling unification, together with the simplest model of D-T splitting, introduces an upper limit on $M_{\rm eff}$ given by $M_{\rm eff} \leq 2.7 \times 10^{18}$ GeV$(3/\tan\beta)$ for the case of MSSM embedded in $SO(10)$. This in turn introduces an implicit dependence on $\tan\beta$ into the lower limit of the SO(10)-amplitude—i.e. $\widehat{A}(SO(10)) \propto 1/M_{\rm eff} \geq$ [(a quantity) $\propto \tan\beta$]. These considerations are reflected in the results given below. [More details can be found in Ref.[11] and [31]].

(iv) Allowing for the ESSM Extension of MSSM: The case of the extended supersymmetric standard model (ESSM), briefly alluded to in Sec. 2, is an interesting variant of MSSM, which can be especially relevant to a host of observable phenomena, including (a) proton decay, (b) possible departure of muon $(g-2)$ from the SM prediction [79], and (c) a lowering of the LEP neutrino-counting from the SM value of 3 [80]. Briefly speaking, ESSM introduces an extra pair of vectorlike families transforming as $\mathbf{16} + \overline{\mathbf{16}}$ of $SO(10)$, having masses of order 1 TeV [35, 81]. Adding such complete $SO(10)$-multiplets would of course preserve coupling unification. From the point of view of adding extra families, ESSM seems to be the minimal and also the maximal extension of the MSSM, that is allowed in that it is compatible with (a) precision electroweak tests, as well as (b) a semi-perturbative

[16]It should be noted that $M_{\rm eff}$ does not represent the physical masses of the color-triplets or of the other particles in the theory. It is simply a parameter of order $M_X^2/M_{10'}$ that is relevant to proton decay. Thus values of $M_{\rm eff}$, close to or even exceeding the Planck scale, does not in any way imply large corrections from quantum gravity.

[17]Accompanying the suppression due to $M_{\rm eff}$, it turns out that the standard $d=5$ operators for $SO(10)$ possess an intrinsic enhancement as well, compared to those for $SU(5)$, primarily due to correlations between the Yukawa couplings in the up and down sectors in $SO(10)$. The standard $d=5$ amplitude for proton decay in $SO(10)$ is thus based on these two opposing effects — suppression through $M_{\rm eff}$ and enhancement through the Yukawa couplings [11].

[18]For instance, in the absence of GUT-scale threshold corrections, the MSSM value of $\alpha_3(m_Z)_{MSSM}$, assuming coupling unification, is given by $\alpha_3(m_Z)^{\circ}_{MSSM} = 0.125 \pm 0.13$ [10], which is about 5-8% higher than the observed value: $\alpha_3(m_Z)^{\circ}_{MSSM} = 0.118 \pm 0.003$. We demand that this discrepancy should be accounted for accurately by a net *negative* contribution from D-T splitting and from "other" GUT-scale threshold corrections, without involving large cancellations. That in fact does happen for the minimal Higgs system $(\mathbf{45}, \mathbf{16}, \overline{\mathbf{16}})$ (see Ref. [11]).

as opposed to non-perturbative gauge coupling unification [35, 82].[19] *The existence of two extra vector-like families of quarks and leptons can of course be tested at the LHC.*

Theoretical motivations for the case of ESSM arise on several grounds: (a) it provides a better chance for stabilizing the dilaton by having a semi-perturbative value for $\alpha_{\rm unif} \approx 0.35 - 0.3$ [35], in contrast to a very weak value of 0.04 for MSSM; (b) owing to increased two-loop effects [35, 82], it raises the unification scale M_X to $(1/2 - 2) \times 10^{17}$ GeV and thereby considerably reduces the problem of a mismatch [36] between the MSSM and the string unification scales (see Section 2); (c) it lowers the GUT-prediction for $\alpha_3(m_Z)$ to (0.112–0.118) (in absence of unification-scale threshold corrections), which is in better agreement with the data than the corresponding value of (0.125–0.113) for MSSM; and (d) it provides a simple reason for inter-family mass-hierarchy [35, 81]. In this sense, ESSM, though less economical than MSSM, offers some distinct advantages.

In the present context, because of raising of M_X and lowering of $\alpha_3(m_Z)$, ESSM naturally enhances the GUT-prediction for proton lifetime, in full accord with the data [83]. Specifically, for ESSM, one obtains: $M_{\rm eff} \leq (6 \times 10^{18}\ {\rm GeV})(30/\tan\beta)$ [11, 31].

As a result, in contrast to MSSM, ESSM can allow for larger values of $\tan\beta$ (like 10), or lighter squark masses ($\sim$ 1 TeV) without needing large threshold corrections, and simultaneously without conflicting with the limit on proton lifetime (see below).

5.2 Proton Decay Rate

Some of the original references on contributions of the standard $d = 5$ operators to proton decay may be found in [84, 85, 86, 87, 88, 11, 31, 89, 90]. I now specify some of the parameters involving the matrix element, renormalization effects and the spectrum of the SUSY partners of the SM particles that are relevant to calculations of proton decay rate.

The hadronic matrix element is defined by $\beta_H u_L(\vec{k}) \equiv \epsilon_{\alpha\beta\gamma}\langle 0 \left| (d_L^\alpha u_L^\beta) u_L^\gamma \right| p, \vec{k}\rangle$. A recent improved lattice calculation yields $\beta_H \approx 0.014$ GeV^3 [91] (whose systematic errors that may arise from scaling violations and quenching are hard to estimate). We will take as a conservative, but plausible, range for β_H to be given by (0.014 GeV^3)$(1/2 - 2)$. A_S denotes the short distance renormalization effect for the $d = 5$ operator which arises owing to extrapolation between GUT and SUSY-breaking scales [85, 87, 92]. The average value of $A_S = 0.67$, given in Ref. [87] for $m_t = 100$ GeV, has been used in most early estimates. For $m_t = 175$ GeV, a recent estimate yields: $A_S \approx 0.93$ to 1.2 [92]. Conservatively, I would use $A_S = 0.93$; this would enhance the rate by a factor of two compared with previous estimates. A_L denotes the long-distance renormalization effect of the $d = 6$ operator due to QCD interaction that arises due to extrapolation between the SUSY breaking scale and 1 GeV [85]. Using the two-loop expression for A_L [93], together with the two-loop value of α_3, Babu and I find: $A_L \approx 0.32$. This by itself would also increase the rate by a factor of $(0.32/0.22)^2 \approx 2$, compared to the previous estimates [85, 86, 87, 88, 11]. Including the enhancements in both A_S and A_L, we thus see that the net increase in the proton decay rate solely due to new evaluation of renormalization effects is nearly a factor of four, compared to the previous estimates (including that in Ref. [11]).

[19] For instance, addition of *two* pairs of vector-like families at the TeV-scale, to the three chiral families, would cause gauge couplings to become non-perturbative below the unification scale.

In Ref. [11], guided by the demand of naturalness (i.e. the absence of excessive fine tuning), in obtaining the Higgs boson mass, squark masses were assumed to lie in the range of 1 TeV$(1/\sqrt{2} - \sqrt{2})$, so that $m_{\tilde{q}} \lesssim 1.4$ TeV. Work, based on the notion of focus point supersymmetry however suggests that squarks may be quite a bit heavier without conflicting with the demands of naturalness [94]. In the interest of obtaining a conservative upper limit on proton lifetime, we will therefore allow squark masses to be as heavy as about 2.4 TeV.

Allowing for plausible and rather generous uncertainties in the matrix element and the spectrum I take:

$$\begin{aligned} \beta_H &= (0.014\ \mathrm{GeV}^3)(1/2-2), \\ m_{\tilde{q}} \approx m_{\tilde{l}} &\approx 1.2\ \mathrm{TeV}(1/2-2), \\ m_{\tilde{w}}/m_{\tilde{q}} &= 1/6(1/2-2), \\ M_{H_c}(\text{minimal SU(5)}) &\leq 10^{16}\ \mathrm{GeV}, \\ A_L &\approx 0.32, \\ A_S &\approx 0.93 \quad \text{and} \quad \tan\beta \geq 3\ . \end{aligned} \tag{28}$$

For evaluation of the strengths of the $d = 6$ operators, generated by exchanges of X and Y gauge bosons, for the cases of SUSY $SO(10)$ or $SU(5)$ with MSSM spectrum, we take:[20]

$$\left.\begin{aligned} M_X \approx M_Y &\approx 10^{16}\ \mathrm{GeV}(1 \pm 25\%) \\ \alpha_i(GUT) &\approx 0.04 \end{aligned}\right\} \text{(for MSSM)}$$

$$\alpha_H = 0.015\ \mathrm{GeV}^3(1/2-2) \tag{29}$$

Before presenting the theoretical predictions, I note the following experimental results on inverse proton decay rates provided by the SuperK studies [95, 83]:

$$\begin{aligned} \Gamma^{-1}(p \to e^+\pi^0)_{\text{expt}} &\gtrsim 6 \times 10^{33}\ \text{yrs} \\ \left[\sum_l \Gamma(p \to \bar{\nu}_l K^+)\right]^{-1}_{\text{expt}} &\gtrsim 1.9 \times 10^{33}\ \text{yrs} \end{aligned} \tag{30}$$

Before the theoretical predictions for proton decay can be given, a few comments are in order.

1. I present the results separately for the standard $d = 5$ and the new $d = 5$ operators, allowing for both the MSSM and the ESSM alternatives. (The contributions of the new $d = 5$ operators are in fact the same for these two alternatives.) Although the proton decay amplitude receives contributions from both the standard and the new operators, in practice, the standard d=5 operators dominate over the new ones for the case of MSSM in the parameter-range of interest that corresponds to predicted proton lifetimes in the upper end, while the new operators dominate over the standard ones for the case of ESSM, in the same range. (This may be inferred from the results listed below.) Thus, in practice, it suffices to consider the contributions of the standard and the new operators separately.

[20]For the central value of α_H, I take the value quoted in Ref. [91] and allow for an uncertainty by a factor of two either way around this central value.

2. In evaluating the contributions of the new $d = 5$ operators to proton decay, allowance is made for the fact that for the f_{ij} couplings (see Eq. (9)), there are two possible $SO(10)$-contractions (leading to a **45** or a **1**) of the pair $\mathbf{16}_i\overline{\mathbf{16}}_H$, both of which contribute to the Majorana masses of the ν_Rs, but only the contraction via the **45** contributes to proton decay. In the presence of non-perturbative quantum gravity one would in general expect both contractions to be present having comparable strengths. For example, the couplings of the **45**s lying in the string-tower or possibly below the string scale, and likewise of the singlets, to the $\mathbf{16}_i\overline{\mathbf{16}}_H$ pair would respectively generate the two contractions. Allowing for a difference between the relevant projection factors for ν_R-masses versus proton decay operator, we set $(f_{ij})_p \equiv (f_{ij})_\nu K$, where $(f_{ij})_\nu$ defined in Sec. 3 directly yields ν_R-masses and K is a relative factor of order unity.[21] As a plausible range, we take $K \approx 1/5 - 2$ (say), where $K = 1/5$ seems to be a conservative value on the low side that would correspond to proton lifetimes near the upper end.

The theoretical predictions for proton decay for the case of the minimal supersymmetric SU(5) model, and the supersymmetric SO(10) and G(224)-models developed in Secs. 3 and 4, are summarized below. They are based on (a) the items (i)–(iv) listed above, (b) the two comments mentioned above, and (c) the values of the relevant parameters listed in Eqns. (28) and (29).[22]

A Summary of Results on Proton Decay and Discussions

$$\left.\frac{\text{Min. SUSY } SU(5)}{\text{MSSM (std. } d=5)}\right\} \quad \Gamma^{-1}(p \to \bar{\nu}K^+) \quad \leq \quad 1.2 \times 10^{31} \text{ yrs} \quad \left(\begin{array}{c}\text{Excluded by}\\ \text{SuperK}\end{array}\right) \tag{31}$$

$$\left.\frac{\text{SUSY } SO(10)}{\text{MSSM (std. } d=5)}\right\} \quad \Gamma^{-1}(p \to \bar{\nu}K^+) \quad \leq \quad 1 \times 10^{33} \text{ yrs} \quad \left(\begin{array}{c}\text{Tightly constrained}\\ \text{by SuperK}\end{array}\right) \tag{32}$$

$$\left.\frac{\text{SUSY } SO(10)}{\text{ESSM (std. } d=5)}\right\} \quad \begin{array}{l}\Gamma^{-1}(p \to \bar{\nu}K^+)_{\text{Med.}} \\ \Gamma^{-1}(p \to \bar{\nu}K^+)\end{array} \quad \begin{array}{l}\approx \\ \lesssim\end{array} \quad \begin{array}{l}(1\text{–}10) \times 10^{33} \text{ yrs} \\ 10^{35} \text{ yrs}\end{array} \quad \left(\begin{array}{c}\text{Fully SuperK}\\ \text{Compatible}\end{array}\right) \tag{33}$$

$$\left.\frac{\text{SUSY } G(224)/SO(10)}{\text{MSSM or ESSM (new } d=5)}\right\} \quad \begin{array}{l}\Gamma^{-1}(p \to \bar{\nu}K^+) \\ B(p \to \mu^+K^0)\end{array} \quad \begin{array}{l}\lesssim \\ \approx\end{array} \quad \begin{array}{l}2 \times 10^{34} \text{ yrs} \\ (1-50)\%\end{array} \quad \left(\begin{array}{c}\text{Fully Compatible}\\ \text{with SuperK}\end{array}\right) \tag{34}$$

$$\left.\frac{\text{SUSY } SU(5) \text{ or } SO(10)}{\text{MSSM } (d=6)}\right\} \quad \Gamma^{-1}(p \to e^+\pi^0) \quad \approx \quad 10^{34.9\pm1} \text{ yrs} \quad \left(\begin{array}{c}\text{Fully Compatible}\\ \text{with SuperK}\end{array}\right) \tag{35}$$

It should be stressed that the upper limits on proton lifetimes given above are quite conservative in that they are obtained (especially for the top two cases) by stretching the uncertainties in the matrix element and the SUSY spectra as given in Eq. (28) to their extremes so as to prolong proton lifetimes. In reality, the lifetimes should be shorter than the upper limits quoted above. With this in mind, the following comments are in order:

[21] Thus the new set of proton decay operators become proportional to $(f_{ij})_\nu g_{kl} K(\overline{\mathbf{16}}_H)(\mathbf{16}_H)/(M^2)(M_{16})$ where $M \approx M_{\text{st}}$ and $M_{16}(\sim M_{\text{GUT}})$ is the mass of the $\mathbf{16}_H$ (see Ref. [31] for a discussion limiting the strength of this operator).

[22] In obtaining the rate for the $e^+\pi^o$-mode induced by the d=6 operator, we have used the net renormalization factor $A_R \approx 2.5$ representing long and short-distance effects[96] and the chiral lagrangian parameters – D and F as in Ref. [97]

1. By comparing the upper limit given in Eq. (31) with the experimental lower limit (Eq. (30)), we see that the *minimal* SUSY $SU(5)$ with the conventional MSSM spectrum is clearly excluded by a large margin by proton decay searches. This is in full agreement with the conclusion reached by other authors (see e.g. Ref. [90])[23]. We have of course noted in Sec. 3 that SUSY $SU(5)$ does not go well with neutrino oscillations observed at SuperK.

2. By comparing Eq. (32) with the empirical lower limit (Eq. (30)), we see that the case of MSSM embedded in $SO(10)$ is already tightly constrained to the point of being disfavored by the limit on proton lifetime. The constraint is of course augmented by our requirement of *natural coupling unification*, which prohibits accidental large cancellation between different threshold corrections.[24] On the positive side, improvement in the current limit by even a factor of 2–3 (say) ought to reveal proton decay. Otherwise the case of MSSM embedded in $SO(10)$ would be clearly excluded.

3. In contrast to the case of MSSM, that of ESSM embedded in $SO(10)$ (see Eq. (33)) is fully compatible with the SuperK limit. In this case, $\Gamma^{-1}_{\rm Med}(p \to \bar{\nu}K^+) \approx 10^{33} - 10^{34}$ yrs, given in Eq. (33), corresponds to the parameters involving the SUSY spectrum and the matrix element β_H being in the *median range*, close to their central values (see Eq. (28)). In short, confining to the standard operators only, if ESSM represents low energy physics and if $\tan\beta$ is rather small (3 to 5 say), *we do not have to stretch the uncertainties in the SUSY spectrum and the matrix elements to their extreme values in order to understand why proton decay has not been seen as yet, and still can be optimistic that it ought to be discovered in the near future with a lifetime* $\lesssim 10^{34}$ *yrs.*[25]

4. We see from Eq. (34) that the contribution of the new operators related to the Majorana masses of the RH neutrinos (Fig. 3) (which is the same for MSSM and ESSM and is independent of $\tan\beta$) is fully compatible with the SuperK limit. These operators can quite naturally lead to proton lifetimes in the range of $10^{33} - 10^{34}$ yrs with an upper limit of about 2×10^{34} yrs.

   It should be remarked that if in the unlikely event all the parameters (β_H, $m_{\tilde{W}}/m_{\tilde{q}}$, and $m_{\tilde{q}}$, etc.) happen to be closer to their extreme values (see Eq. (28)) so as to extend proton lifetime, the standard $d = 5$ operators for the case of ESSM embedded in $SO(10)$ would lead to lifetimes as long as about 10^{35} years (see Eq. (33)). But in this case the new $d = 5$ operators related to neutrino masses are likely to dominate and quite naturally lead to lifetimes bounded above in the range of $(1 - 20) \times 10^{33}$ years (as noted in Eq. (34)). *Thus in the presence of the new operators, the range of*

[23]See, however, Refs [98] and [99], where attempts are made to save minimal SUSY SU(5) by a set of scenarios, which seems (to me) contrived. These include a judicious choice of sfermion mixings, higher dimensional operators and squarks of first two families having masses of order 10 TeV.

[24]For instance, had we allowed the "other" GUT-scale threshold corrections (in our case, those arising from $\mathbf{45}_H$, $\mathbf{16}_H$, $\overline{\mathbf{16}}_H$ and the gauge multiplet, see Refs. [11, 31]) to $\alpha_3(m_Z)$ to be negative in sign and large as about 15% (rather than 10%), as some authors do [89], the upper limit on proton lifetime would have been higher by about a factor of 5, compared to Eq. (32).

[25]The results on proton lifetimes for a wide variation of the parameters for the case of MSSM and ESSM embedded in $SO(10)$ are listed in Tables 1 and 2 of Ref. [31].

$10^{33} - 10^{34}$ *years for proton lifetime is not only very plausible, but it also provides a reasonable upper limit for the same, for the case of ESSM embedded in* $SO(10)$.

5. We see that the gauge boson mediated $d = 6$ operators, for the case of MSSM embedded in $SU(5)$ or $SO(10)$, though typically suppressed compared to the $d = 5$ operators, can lead to inverse decay rates $\Gamma^{-1}(p \to e^+\pi^0)$ as short as about 10^{34} years (see Eq. (35)). It should be stressed that the $e^+\pi^0$-mode is the *common denominator* of all GUT models ($SU(5)$, $SO(10)$, etc.) which unify quarks and leptons and the three gauge forces. Its rate is determined essentially by α_H and the SUSY unification scale, without the uncertainty of the SUSY spectrum. I should also mention that the $e^+\pi^0$-mode is predicted to be the dominant mode in the flipped $SU(5) \times U(1)$-model [28], and also as it turns out in certain higher dimensional GUT-models [100], as well as in a model of compactification of M-theory on a manifold of G_2 holonomy [101]. For these reasons, intensifying the search for the $e^+\pi^0$-mode to the level of sensitivity of about 10^{35} years in a next-generation proton decay detector should be well worth the effort.

   It may be noted that for the case of ESSM embedded in $SO(10)$ or $SU(5)$, since $\alpha_{\rm unif}$ and the unification scale (thereby the masses of the X, Y gauge bosons) are raised by nearly a factor of (6 to 7) and (2.5 to 5) respectively, compared to those for MSSM (see Ref. [35] and discussions above), while the inverse decay rate is proportional to $(M_X^4/\alpha_{\rm unif}^2)$ we expect

$$\Gamma^{-1}(p \to e^+\pi^0)^{d=6}_{\rm ESSM} \approx (1 \text{ to } 17)\Gamma^{-1}(p \to e^+\pi^0)^{d=6}_{\rm MSSM}.$$

   This raises the interesting possibility that for ESSM embedded in $SO(10)$, both $\bar{\nu}K^+$ (arising from $d = 5$) and $e^+\pi^0$ (arising from $d = 6$) can have comparable rates, with proton having a lifetime $\sim (1/2 - 2) \times 10^{34}$ years.

Before concluding I should mention that there have been several old and new attempts in the literature based on compactification of string/M-theory [101, 102], as well as of a presumed 5D or 6D point-particle GUT-theory [103, 104], which project out the color triplets (anti-triplets) belonging to $\mathbf{5}_H$ ($\bar{\mathbf{5}}_H$) of $SU(5)$ or $\mathbf{10}_H$ of $SO(10)$ from the massless spectrum in 4D, through the process of compactification. As a result, they obtain a non-GUT SM-like gauge symmetry, and in some cases the $G(224)$ symmetry (see e.g. [17] and [18]) in 4D. In the process they eliminate (often using discrete symmetries) or strongly suppress the standard $d = 5$ proton decay operators, though not necessarily the $d = 6$.

These approaches are interesting in their own right. There are, however, some constraints which should be satisfied if one wishes to understand certain observed features of low energy physics. In particular, it seems to me that at the very least B-L should emerge as a gauge symmetry in 4D so as to protect the RH neutrinos from getting a string-scale mass (see footnote 2) *and* equally important to implement baryogenesis via leptogenesis, as discussed in Secs. 3 and 4. This feature is not available in models which start with $SU(5)$ in 5D, or in those that obtain only a standard model-like gauge symmetry without B-L in 4D. Furthermore, the full $SU(4)$ color symmetry, which of course contains B-L, plays a crucial role in yielding not only (a) the (empirically favored) relation $m_b(M_X) \approx m_\tau$, but also (b) the relation $m(\nu^\tau_{\rm Dirac}) = m_{\rm top}(M_X)$ which is needed to account for the observed value of

$\Delta m^2(\nu_2 - \nu_3)$ (see Sec. 3), and (c) the smallness of V_{cb} together with the near maximality of $\sin^2 2\theta^{\rm osc}_{\nu_\mu\nu_\tau}$, as observed. The symmetry $SU(2)_L \times SU(2)_R$ is also most useful in that it relates the masses and mixings in the up and the down sectors. Without these correlations, the successful predictions listed in Eq. (16) will not emerge.

In short, as noted in Secs. 3 and 4, an understanding of neutrino oscillations and leptogenesis as well as of certain intriguing features of the masses and mixings of all fermions including neutrinos seems to strongly suggest that minimally the symmetry $G(224)$, or maximally the symmetry $SO(10)$, should survive as an effective symmetry in 4D. If the symmetry $G(224)$ rather than $SO(10)$ survives in 4D near the string scale, the familiar color triplets would be projected out through compactification [see e.g. [17] and [18]].[26] In this case, there is no need for a doublet-triplet splitting mechanism and the standard $d = 5$ operators are either strongly suppressed or completely eliminated. *However, as long as the Majorana masses of the RH neutrinos and the CKM mixings are generated through the minimal Higgs system as in Sec. 3, the new $d = 5$ operators (Fig. 3) would still generically be present, and would be the dominant source of proton decay.* Like the standard $d = 5$ operators (Fig. 2), the new $d = 5$ operators also lead to $\bar{\nu}K^+$ and $\bar{\nu}\pi^+$ as dominant modes, but in contrast to the standard operators, the new ones can lead to prominent μ^+K^0-modes [32] (see Eq. (35)).

Given the empirical success of the supersymmetric $G(224)/SO(10)$-framework, derivation of this framework, at least that based on an effective $G(224)$-symmetry in 4D leading to the pattern of Yukawa couplings as in Sec. 3, from an underlying theory remains a challenge. At the same time, based on its empirical support so far, it makes sense to test this picture thoroughly. There are two notable pieces of this picture still missing. One is supersymmetry, which will be tested at the LHC. The other is proton decay.

5.3 Section Summary

Given the importance of proton decay, a systematic study of this process has been carried out within a supersymmetric $SO(10)/G(224)$ framework that successfully describes the fermion masses, neutrino oscillations and leptogenesis. Special attention is paid in this study to the dependence of the d=5 proton decay amplitude on the masses and mixings of the fermions and the GUT-scale threshold effects. Allowing for both the MSSM and the ESSM alternatives within this $SO(10)/G(224)$ framework and including the standard as well as the new $d = 5$ operators, one obtains (see Eqs. (31)–(35)) a conservative upper limit on proton lifetime given by:

$$\tau_{\rm proton} \lesssim (1/3 - 2) \times 10^{34} \text{ yrs} \begin{pmatrix} \text{SUSY} \\ SO(10)/G(224) \end{pmatrix} \tag{36}$$

with $\bar{\nu}K^+$ and $\bar{\nu}\pi^+$ being the dominant modes and quite possibly μ^+K^0 being prominent.

The $e^+\pi^o$-mode induced by gauge boson-exchanges should have an inverse decay rate in the range of $10^{34} - 10^{36}$ years (see Eq. (35)). The implication of these predictions for a next-generation detector is emphasized in the next section.

[26]The issue of gauge coupling unification for this case is discussed in Sec. 2.

6 Concluding Remarks

In this talk, I have argued that but for two missing pieces – supersymmetry and proton decay – the evidence in favor of supersymmetric grand unification is now strong. It includes: (i) the observed family multiplet-structure, (ii) quantization of electric charge, (iii) the meeting of the three gauge couplings, (iv) neutrino oscillations (atmospheric and solar), (v) the intricate pattern of the masses and mixings of all fermions, including neutrinos, and (vi) the likely need for leptogenesis to account for the observed baryon asymmetry of the universe. All of these features can be understood simply and even quantitatively (see e.g. Eqs. (3), (4), and (16)) within the concept of supersymmetric grand unification based on an effective string-unified $G(224)$ or $SO(10)$-symmetry in 4D. As discussed in Secs 3 and 4, attempts to understand especially (a) the tiny neutrino masses, (b) the baryon asymmetry of the universe (via leptogenesis), as well as (c) certain features of quark-lepton masses and mixings seem to select out the G(224)/SO(10) route to unification, as opposed to other alternatives.

A systematic study of proton decay has thus been carried out within this SO(10)/G(224) framework [11, 31], allowing for the possibilities of both MSSM and ESSM, and including the contributions for the gauge boson-mediated d=6, the standard d=5 as well as the new d=5 operators related to the Majorana masses of the RH neutrinos. Based on this study, I have argued that a conservative upper limit on the lifetime of the proton is about $(\frac{1}{3} - 2) \times 10^{34}$ years.

So, unless the fitting of all the pieces (i)-(vi) listed above is a mere coincidence, it is hard to believe that that is the case, discovery of proton decay should be around the corner. Allowing for the possibility that proton lifetime may well be near the upper limit stated above, a next generation detector, of the type proposed by UNO and Hyperkamiokande, providing a net gain in sensitivity by about a factor of five to ten, compared to SuperK, would thus be needed to produce real proton decay events and distinguish them from the background.

The reason for pleading for such improved searches is that proton decay would provide us with a wealth of knowledge about physics at truly short distances ($< 10^{-30}$ cm), which cannot be gained by any other means. Specifically, the observation of proton decay, at a rate suggested above, with $\overline{\nu}K^+$ mode being dominant, would not only reveal the underlying unity of quarks and leptons but also the relevance of supersymmetry. It would also confirm a unification of the fundamental forces at a scale of order 2×10^{16} GeV. Furthermore, prominence of the μ^+K^0 mode, if seen, would have even deeper significance, in that in addition to supporting the three features mentioned above, it would also reveal the link between neutrino masses and proton decay, as discussed in Sec. 5. *In this sense, the role of proton decay in probing into physics at the most fundamental level is unique.* In view of how valuable such a probe would be and the fact that the predicted upper limit on the proton lifetime is at most a factor of three to ten higher than the empirical lower limit, the argument in favor of building an improved detector seems compelling.

Such a detector should of course be designed to serve multiple goals including especially improved studies of neutrino oscillations and supernova signals. These ideas and others including that of a neutrino factory were discussed intensively at the NeSS meeting held recently in Washington [105].

To conclude, the discovery of proton decay would constitute a landmark in the history of

physics. That of supersymmetry would do the same. The discoveries of these two features – supersymmetry and proton decay – would fill the two missing pieces of gauge unification and would shed light on how such a unification may be extended to include gravity in the context of a deeper theory. The question thus poses: Will our generation give itself a chance to realize *both*?

Acknowledgments: I would like to thank Gerard 't Hooft and Antonino Zichichi for the kind invitation to lecture at the Erice School. I have benefitted from many collaborative discussions with Kaladi S. Babu and Frank Wilczek on topics covered in this lecture. I am also grateful to Pasquale Di Bari and Qaisar Shafi for discussions on the topic covered in Sec. 4. The research presented here is supported in part by DOE grant no. DE-FG02-96ER-41015.

References

[1] Y. Fukuda et al. (Super-Kamiokande), Phys. Rev. Lett. **81**, 1562 (1998), hep-ex/9807003; K. Nishikawa (K2K) Talk at Neutrino 2002, Munich, Germany.

[2] Q. R. Ahmad et al (SNO), Phys. Rev. Lett. **81**, 011301 (2002); B. T. Cleveland et al (Homestake), Astrophys. J. **496**, 505 (1998); W. Hampel et al. (GALLEX), Phys. Lett. **B447**, 127, (1999); J. N. Abdurashitov et al (SAGE) (2000), astro-ph/0204245; M. Altmann eta al. (GNO), Phys. Lett. **B 490**, 16 (2000); S. Fukuda et al. (SuperKamiokande), *Phys. Lett.* **B539**, 179 (2002). Disappearance of $\bar{\nu}_e$'s produced in earth-based reactors is established by the KamLAND data: K. Eguchi et al., hep-ex/0212021. For a recent review including analysis of several authors see, for example, S. Pakvasa and J. W. F. Valle, hep-ph/0301061.

[3] For a historical overview of theoretical calculations of expected solar neutrino flux, see J. Bahcall, astro-ph/0209080.

[4] J. C. Pati, "*Implications of the SuperKamiokande Result on the Nature of New Physics*", in Neutrino 98, Takayama, Japan, June 98, hep-ph/9807315; Nucl. Phys. B (Proc. Suppl.) **77**, 299 (1999).

[5] J. C. Pati and Abdus Salam; Phys. Rev. **D8**, 1240 (1973); J. C. Pati and Abdus Salam, Phys. Rev. Lett. **31**, 661 (1973); Phys. Rev. **D10**, 275 (1974).

[6] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[7] H. Georgi, H. Quinn and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).

[8] Y. A. Golfand and E. S. Likhtman, JETP Lett. **13**, 323 (1971). J. Wess and B. Zumino, Nucl. Phys. **B70**, 139 (1974); D. Volkov and V. P. Akulov, JETP Lett. **16**, 438 (1972).o.

[9] E. Witten, Nucl. Phys. **B 188**, 513 (1981); R. K. Kaul, Phys. Lett. **109 B**, 19 (1982).

[10] The essential features pertaining to coupling unification in SUSY GUTS were noted by S. Dimopoulos, S. Raby and F. Wilczek, Phys. Rev. **D 24**, 1681 (1981); W. Marciano

and G. Senjanovic, Phys. Rev. **D 25**, 3092 (1982) and M. Einhorn and D. R. T. Jones, Nucl. Phys. **B 196**, 475 (1982). For work in recent years, see P. Langacker and M. Luo, Phys. Rev. **D 44**, 817 (1991); U. Amaldi, W. de Boer and H. Furtenau, Phys. Rev. Lett. **B 260**, 131 (1991); F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, Nuov. Cim. **A 104** 1817 (1991).

[11] K. S. Babu, J. C. Pati and F. Wilczek, "*Fermion Masses, Neutrino Oscillations and Proton Decay in the Light of the SuperKamiokande*" hep-ph/981538V3; Nucl. Phys. B (to appear).

[12] M. Fukugita and T. Yanagida, Phys. Lett. **B 174**, 45 (1986); M. A. Luty, Phys. Rev. **D 45**, 455 (1992); W. Buchmuller and M. Plumacher, hep-ph/9608308.

[13] V. Kuzmin, V. Rubakov and M. Shaposhnikov, Phys. Lett **BM155**, 36 (1985).

[14] H. Georgi, in Particles and Fields, Ed. by C. Carlson (AIP, NY, 1975), p. 575; H. Fritzsch and P. Minkowski, Ann. Phys. **93**, 193 (1975).

[15] M. Green and J. H. Schwarz, Phys. Lett. **149B**, 117 (1984); D. J. Gross, J. A. Harvey, E. Martinec and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985); P. Candelas, G. T. Horowitz, A. Strominger and E. Witten, Nucl. Phys. **B 258**, 46 (1985). For introductions and reviews, see: M. B. Green, J. H. Schwarz and E. Witten, "Superstring Theory" Vols. 1 and 2 (Cambridge University Press); J. Polchinski, "String Theory" vols.1 and 2 (Cambridge University Press).

[16] For a few pioneering papers on string-duality and M-theory, relevant to gauge-coupling unification, see E. Witten, Nucl. Phys. **B 443**, 85 (1995) and P. Horava and E. Witten, Nucl. Phys. **B 460**, 506 (1996). For reviews, see e.g. J. Polchinski, hep-th/9511157; and A. Sen, hep-th/9802051, and references therein.

[17] Promising string-theory solutions yielding the G(224)-symmetry in 4D have been obtained using different approaches, by a number of authors. They include: I. Antoniadis, G. Leontaris, and J Rizos, Phys. Lett. **B245**, 161 (1990); G. K. Leontaris, *Phys. Lett.* **B372**, 212 (1996), hep-ph/9601337; A. Murayama and T. Toon, Phys. Lett. **B318**, 298 (1993); Z. Kakushadze, Phys. rev. **D58**, 101901 (1998); G. Aldazabal, L. I. Ibanez and F. Quevedo, hep-th/9909172; C. Kokorelis, hep-th/0203187, hep-th/0209202; M. Cvetic, G. Shiu, and A. M. Uranga, *Phys. Rev. Lett.* **87**, 201801 (2001), hep-th/0107143, and *Nucl. Phys.* **B615**, 3 (2001), hep-th/0107166; M. Cvetic and I. Papadimitriou, hep-th/0303197; R. Blumenhagen, L. Gorlich and T. Ott, hep-th/0211059. For a type I string-motivated scenario leading to the G(224) symmetry in 4D, see L. I. Everett, G. L. Kane, S. F. King, S. Rigolin and L. T. Wang, hep-th/0202100.

[18] Recently there have been several attempts based on compactifications of five and six-dimensional GUT-theories which lead to the G(224) symmetry in 4D with certain desirable features. See e.g. R. Dermisek and A. Mafi, Phys. Rev. **D65**, 055002 (2002) [hep-ph/0108139]; Q. Shafi and Z. Tavartkiladze, hep-ph/0108247, hep-ph/0303150; C. H. Albright and S. M. Barr, hep-ph/0209173; H. D. Kim and S. Raby, hep-ph/0212348;

I. Gogoladze, Y. Mimura and S. Nandi, hep-ph/0302176; B. Kyae and Q. Shafi, hep-ph/0211059; H. Baer et al., hep-ph/0204108; For a global phenomenological analysis of a realistic string-inspired supersymmetric model based on the $G(224)$ symmetry see T. Blazek, S. F. King, and J. K. Perry (hep-ph/0303192); and also references therein.

[19] J. C. Pati and Abdus Salam, *Phys. Rev.* **D10**, 275 (1974); R. N. Mohapatra and J. C. Pati, *Phys. Rev.* **D11**, 566 and 2558 (1975); G. Senjanovic and R. N. Mohapatra, *Phys. Rev.* **D12**, 1502 (1975).

[20] M. Gell-Mann, P. Ramond and R. Slansky, in: *Supergravity*, eds. F. van Nieuwenhuizen and D. Freedman (Amsterdam, North Holland, 1979) p. 315; T. Yanagida, in: *Workshop on the Unified Theory and Baryon Number in the Universe*, eds. O. Sawada and A. Sugamoto (KEK, Tsukuba) 95 (1979); R. N. Mohapatra and G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980).

[21] F. Gursey, P. Ramond and P. Sikivie, Phys. Lett. **B 60**, 177 (1976).

[22] J. C. Pati, "*Leptogenesis and Neutrino Oscillations Within a predictive G(224)/SO(10) Framework*", hep-ph/0209160.

[23] For a recent analysis based on Big Bang Nucleosynthesis, see B. D. Fields and S. Sarkar, Review of Particle Physics, Phys. Rev. **D66**, 010001 (2002), which yields: $Y_B^{BBN} \approx (3.7 - 9) \times 10^{-11}$.

[24] Most recently, the WMAP surveying the entire celestial sphere with high sensativity yields: $(Y_B)_{\rm WMAP} \approx (8.7 \pm 0.4) \times 10^{-11}$ (WMAP collaboration, astro-ph/0302207–09–13–15, 17, 18, 20, 22–25).

[25] M. Dine and A. Kusenko, hep-ph/0303065.

[26] F. Gursey, P. Ramond and R. Slansky, Phys. Lett, **B60**, 177 (1976); Y. Achiman and B. Stech, Phys. Lett. **B77**, 389 (1978); Q. Shafi, Phys. Lett. **B79**, 301 (1978); A. deRujula, H. Georgi and S. L. Glashow, 5th Workshop on Grand Unification, edited by K. Kang et al., World Scientific, 1984, p88.

[27] J. C. Pati and A. Salam, Phys. Rev. **D10**, 275 (1974); R. N. Mohapatra and J. C. Pati, Phys. Rev. **D11**, 566 and 2558 (1975).

[28] S. M. Barr, Phys. Lett. **B112**, 219 (1982); J. P. Derendinger, J. E. Kim and D. V. Nanopoulos, Phys. Lett. **B139**, 170 (1984); I. Antoniadis, J. Ellis, J. Hagelin and D. V. Nanopoulos, Phys. Lett. **B194**, 231 (1987).

[29] UNO Proposal: See the talk by C. K. Jung in "*Next Generation Nucleon Decay and Neutrino Detector*", AIP Conference Proceedings (Sept, 1999), Editors M. Diwan and C. K. Jung.

[30] Plans for Hyperkamiokande: Private Communications, Y. Totsuka. See also talks by M. Shiozawa and Y. Suzuki in the AIP Conf. Proceedings (Ref. [29]).

[31] J. C. Pati, "*Confronting The Conventional ideas on Grand Unification With Fermion Masses, Neutrino Oscillations and Proton Decay*", hep-ph/0204240; To appear in the Proceedings of the ICTP, Trieste School and Proceedings of the DAE Meeting, Allahabad, India (2002).

[32] K. S. Babu, J. C. Pati and F. Wilczek, "*Suggested New Modes in Supersymmetric Proton Decay*", Phys. Lett. **B 423**, 337 (1998).

[33] P. Ginsparg, Phys. Lett. **B 197**, 139 (1987); V. S. Kaplunovsky, Nucl. Phys. **B 307**, 145 (1988); Erratum: *ibid.* **B 382**, 436 (1992).

[34] E. Witten, hep-th/9602070.

[35] K. S. Babu and J. C. Pati, "*The Problems of Unification – Mismatch and Low α_3: A Solution with Light Vector-Like Matter*", hep-ph/9606215, Phys. Lett. **B 384**, 140 (1996).

[36] For a recent discussion, see K. Dienes, Phys. Rep. **287**, 447 (1997), hep-th/9602045 and references therein; J. C. Pati, "*With Neutrino Masses Revealed, Proton Decay is the Missing Link*", hep-ph/9811442; Proc. Salam Memorial Meeting (1998), World Scientific; Int'l Journal of Modern Physics A, vol. 14, 2949 (1999).

[37] See e.g. D. Lewellen, Nucl. Phys. **B 337**, 61 (1990); A. Font, L. Ibanez and F. Quevedo, Nucl. Phys. **B 345**, 389 (1990); S. Chaudhari, G. Hockney and J. Lykken, Nucl. Phys. **B 456**, 89 (1995) and hep-th/9510241; G. Aldazabal, A. Font, L. Ibanez and A. Uranga, Nucl. Phys. **B 452**, 3 (1995); *ibid.* **B 465**, 34 (1996); D. Finnell, Phys. Rev. **D 53**, 5781 (1996); A.A. Maslikov, I. Naumov and G.G. Volkov, Int. J. Mod. Phys. **A 11**, 1117 (1996); J. Erler, hep-th/9602032 and G. Cleaver, hep-th/9604183; and Z. Kakushadze and S.H. Tye, hep-th/9605221, and hep-th/9609027; Z. Kakushadze *et al.*, hep-ph/9705202.

[38] A. Faraggi, Phys. Lett. **B 278**, 131 (1992); Phys. Lett. **B 274**, 47 (1992); Nucl. Phys. **B 403**, 101 (1993); A. Faraggi and E. Halyo, Nucl. Phys. **B 416**, 63 (1994).

[39] A. Faraggi and J.C. Pati, "*A Family Universal Anomalous U(1) in String Models as the Origin of Supersymmetry Breaking and Squark-degeneracy*", hep-ph/9712516v3, Nucl. Phys. **B 256**, 526 (1998).

[40] K.S. Babu and J.C. Pati, "A Resolution of Supersymmetric Flavor-changing CP Problems Through String Flavor symmetries", UMD-PP0067, to appear.

[41] J.C. Pati, "*The Essential Role of String Derived Symmetries in Ensuring Proton Stability and Light Neutrino Masses*", hep-ph/9607446, Phys. Lett. **B 388**, 532 (1996).

[42] J.C. Pati, (Ref. [36]).

[43] I. Antoniadis, Phys. Lett **B246**, 377 (1990).

[44] See e.g., J. Lykken, Phys. Rev. **D54**, 3693 (1996).

[45] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B429**, 263 (1998); I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B436**, 357 (1998); K. Dienes, E. Dudas, T. Gherghetta, Phys. Lett. **B436**, 55 (1998). For a recent comprehensive review of this scenario and other references, see N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Physics Today, February 2002 (pages 35-40).

[46] See e.g., K.R. Dienes and J. March-Russell, hep-th/9604112; K.R. Dienes, hep-ph/9606467.

[47] These have been introduced in various forms in the literature. For a sample, see e.g., C. D. Frogatt and H. B. Nielsen, Nucl. Phys. **B147**, 277 (1979); L. Hall and H. Murayama, Phys. Rev. Lett. **75**, 3985 (1995); P. Binetruy, S. Lavignac and P. Ramond, Nucl. Phys. **B477**, 353 (1996). In the string theory context, see e.g., A. Faraggi, Phys. Lett. **B278**, 131 (1992).

[48] C. Albright and S. Barr, Phys. Rev. lett. **85**, 244 (2000).

[49] For G(224), one can choose the corresponding sub-multiplets – that is $(1, 1, 15)_H$, $(1, 2, \bar{4})_H$, $(1, 2, 4)_H$, $(2, 2, 1)_H$ – together with a singlet S, and write a superpotential analogous to Eq. (8).

[50] If the effective non-renormalizable operator like $\mathbf{16}_2\mathbf{16}_3\mathbf{10}_H\mathbf{45}_H/M'$ is induced through exchange of states with GUT-scale masses involving renormalizable couplings, rather than through quantum gravity, M' would, however, be of order GUT-scale. In this case $(\mathbf{45}_H\rangle/M' \sim 1$, rather than 1/10.

[51] While $\mathbf{16}_H$ has a GUT-scale VEV along the SM singlet, it turns out that it can also have a VEV of EW scale along the "$\tilde{\nu}_L$" direction due to its mixing with $\mathbf{10}_H^d$, so that the H_d of MSSM is a mixture of $\mathbf{10}_H^d$ and $\mathbf{16}_H^d$. This turns out to be the origin of non-trivial CKM mixings (See Ref. [11]).

[52] The flavor charge(s) of $\mathbf{45}_H(\mathbf{16}_H)$ would get determined depending upon whether $p(q)$ is one or zero (see below).

[53] The basic presumption here is that effective dimensionless couplings allowed by $SO(10)/G(224)$ and flavor symmetries are of order unity [i.e., $(h_{ij}, g_{ij}, a_{ij}) \approx$ 1/3-3 (say)]. The need for appropriate powers of (S/M) with $(S\rangle/M \sim M_{\rm GUT}/M_{\rm string} \sim (1/10 - 1/20)$ in the different couplings leads to a hierarchical structure. As an example, consider just one U(1)-flavor symmetry with one singlet S. The hierarchical form of the Yukawa couplings exibited in Eqs. (7) and (8) would be allowed, for the case of $p = 1$, $q = 0$, if ($\mathbf{16}_3$, $\mathbf{16}_2$, $\mathbf{16}_1$, $\mathbf{10}_H$, $\mathbf{16}_H$, $\mathbf{45}_H$ and S) are assigned U(1)-charges of (a, $a+1$, $a+2$, $-2a$, $-a-1/2$, 0, -1). It is assumed that other fields are present that would make the U(1) symmetry anomaly-free. With this assignment of charges, one would expect $|\zeta_{22}^{u,d}| \sim ((S\rangle/M)^2$; one may thus take $|\zeta_{22}^{u,d}| \sim (1/3) \times 10^{-2}$ without upsetting the success of Ref. [11]. In the same spirit, one would expect $|\zeta_{13}, \zeta_{31}| \sim ((S\rangle/M)^2 \sim 10^{-2}$ and $|\zeta_{11}| \sim ((S\rangle/M)^4 \sim 10^{-4}$ (say). The value of "a" would get fixed by the presence of other operators (see later).

[54] These effective non-renormalizable couplings can of course arise through exchange of (for example) **45** in the string tower, involving renormalizable $\mathbf{16}_i\bar{\mathbf{16}}_H\mathbf{45}$ couplings. In this case, one would expect $M \sim M_{\text{string}}$.

[55] Note that the magnitudes of η, ϵ and σ are fixed by the input quark masses. Furthermore, one can argue that the two contributions for $\theta^{\text{osc}}_{\nu_\mu\nu_\tau}$ [see Eq. (16)] necessarily add to each other as long as $|y|$ is hierarchical ($\sim 1/10$) [11]. As a result, once the sign of ϵ relative to η and σ is chosen to be negative, the actual magnitudes of $V_{cb} \approx (0.044)$ and $\sin^2 2\theta^{\text{osc}}_{\nu_\mu\nu_\tau} \approx 0.92 - 0.99$ emerge as predictions of the model [11, 31].

[56] Note that such an operator would be allowed by the flavor symmetry defined in Ref. [53] if one sets $a = 1/2$. In this case, operators such as W_{23} and W_{33} that would contribute to $\nu^\mu_L\nu^\tau_L$ and $\nu^\tau_L\nu^\tau_L$ masses would be suppressed relative to W_{12} by flavor symmetry.

[57] A term like W_{12} can be induced in the presence of, for example, a singlet $\hat{S}$ and a ten-plet $(\hat{\mathbf{10}})$, possessing effective renormalizable couplings of the form $a_i\mathbf{16}_i\mathbf{16}_H\hat{\mathbf{10}}$, $b\hat{\mathbf{10}}\mathbf{10}_H\hat{S}$ and mass terms $\hat{M}_S\hat{S}\hat{S}$ and $\hat{M}_{10}\hat{\mathbf{10}}\hat{\mathbf{10}}$. In this case $\kappa_{12}/M^3_{\text{eff}} \approx a_1a_2b^2/(\hat{M}^2_{10}\hat{M}_S)$. Setting the charge $a = 1/2$ (see Ref. [53] and [56]), and assigning charges (-3/2, 5/2) to $(\hat{\mathbf{10}}, \hat{S})$, the couplings a_1, and b would be flavor-symmetry allowed, while a_2 would be suppressed but so also would be the mass of $\hat{\mathbf{10}}$ compared to the GUT-scale. One can imagine that $\hat{S}$ on the other hand acquires a GUT-scale mass through for example the Dine-Seiberg-Witten mechanism, violating the U(1)-flavor symmetry. One can verify that in such a picture, one would obtain $\kappa_{12}/M^3_{\text{eff}} \sim 1/M^3_{\text{GUT}}$.

[58] See e.g. P. Vogel, Review of Particle Physics, *Phys. Rev.* **D66**, 010001 (2002) for a review of current status and a discussion of the Heidelberg-Moscow collaboration data which suggest a possible signal at the level of 2.2–3.15*sigma*.

[59] For instance, consider the superpotential for $\mathbf{45}_H$ only: $W(\mathbf{45}_H) = M_{45}\mathbf{45}^2_H + \lambda\mathbf{45}^4_H/M$, which yields (setting $F_{\mathbf{45}_H} = 0$), either $\langle\mathbf{45}_H\rangle = 0$, or $\langle\mathbf{45}_H\rangle^2 = -[2M_{45}M/\lambda]$. Assuming that "other physics" would favor $\langle\mathbf{45}_H\rangle \neq 0$, we see that $\langle\mathbf{45}_H\rangle$ would be pure imaginary, if the square bracket is positive, with all parameters being real. In a coupled system, it is conceivable that $\langle\mathbf{45}_H\rangle$ in turn would induce phases (other than "0" and π) in some of the other VEV's as well, and may itself become complex rather than pure imaginary.

[60] L. Wolfenstein. *Phys. Rev. Lett.*, 51, 1983.

[61] K. S. Babu and J. C. Pati. "*Link Between Neutrino Oscillations and CP Violation within Supersymmetric Unification*". In G. Branco, Q. Shafi, and J. I. Silva-Marcos, editors, *Recent Developments in Particle Physics and Cosmology.* NATO Advanced Study Institute, June 6–July 7, 2000. Abstract only (paper to appear).

[62] Within the framework developed in Ref. [61], the CP violating phases entering into the SUSY contributions (for example those entering into the squark-mixings), though distinct from the CKM phase, also arise entirely through phases in the fermion mass matrices, just as the CKM phase does.

[63] An intriguing feature is the prominence of the $\delta^{23}_{RR}(\tilde{b}_R \to \tilde{s}_R)$-parameter which gets enhanced in part because of the largeness of the ν_μ-ν_τ oscillation angle. This leads to large departures from the predictions of the standard model, especially in transitions such as $B_s \to \bar{B}_s$ and $B_d \to \Phi K_s\,(b \to s\bar{s}s)$ [61]. This feature has independently been noted recently by D. Chang, A. Massiero, and H. Murayama (hep-ph/0205111).

[64] As an example, one such fit with complex parameters assigns [61]: $\sigma = 0.10 - 0.012\,i$, $\eta = 0.12 - 0.05\,i$, $\epsilon = -0.095$, $\eta' = 4.0 \times 10^{-3}$, $\epsilon' = 1.54 \times 10^{-4} e^{i\pi/4}$, $\zeta^u_{22} = 1.25 \times 10^{-3} e^{i\pi/9}$ and $\zeta^d_{22} = 4 \times 10^{-3} e^{i\pi/2}$, $\mathcal{M}^0_u \approx 110$ GeV, $\mathcal{M}^0_D \approx 1.5$ GeV, $y \approx -1/17$ (compare with Eq. (15) for which $\zeta^u_{22} = \zeta^d_{22} = 0$). One obtains as outputs: $m_{b,s,d} \approx$ (5 GeV, 132 MeV, 8 MeV), $m_{c,u} \approx$ (1.2 GeV, 4.9 MeV), $m_{\mu,e} \approx$ (102 MeV, 0.4 MeV) with $m_{t,\tau} \approx$ (167 GeV, 1.777 GeV), $(V_{us}, V_{cb}, |V_{ub}|, |V_{td}|) \approx (0.217, 0.044, 0.0029, 0.011)$, while preserving the predictions for neutrino masses and oscillations as in Eq. (16). The above serves to demonstrate that complexification of parameters of the sort presented above can preserve the successes of Eq. (16) ([11]). This particular case leads to $\eta_W = 0.29$ and $\rho_W = -0.187$ [61], to be compared with the corresponding standard model values (obtained from ϵ_K, V_{ub} and Δm_{Bd}) of $(\eta_W)_{\rm SM} \approx 0.33$ and $(\rho_W)_{\rm SM} \approx +0.2$. The consistency of such values for η_W and ρ_W (especially reversal of the sign of ρ_W compared to the SM value), in the light of having both standard model and SUSY-contributions to CP and flavor-violations, and their distinguishing tests, are discussed in Ref. [61].

[65] For reviews, see chapters 6 and 8 in E. W. Kolb and M. S. Turner, *The Early Universe*, Addison-Wesley, 1990.

[66] J. Ellis and J. E. Kim and D. Nanopoulos, *Phys. Lett.* **145B** 181 (1984); M. Yu. Khlopov and A. Linde, *Phys. Lett.* **138B** 265 (1984); E. Holtmann and M. Kawasaki and K. Kohri and T. Moroi, hep-ph/9805405.

[67] B. A. Campbell, S. Davidson and K. A. Olive, Nucl. Phys. **B399**, 111 (1993).

[68] L. Covi, E. Roulet and F. Vissani, Phys. Lett. **B384**, 169 (1996).

[69] M. Plumacher, hep-ph/9704231.

[70] P. Di Bari, hep-ph/0211175; W. Buchmuller, P. Di Bari and M. Plumacher, Nucl. Phys. **B643**, 367 (2002) [hep-ph/0205349].

[71] The factor 0.7 in Eq. (21) [instead of 1 in Eq. (13) of the first paper in Ref. [70]] is an estimate that incorporates the modification needed for SUSY corresponding to a doubling of N_1-decay width owing to the presence of both $l + H$ and $\tilde{l} + \tilde{H}$-modes and an increase of g^* from 106 for the standard model to 228 for SUSY.

[72] One can verify that $K \equiv (\Gamma(N_1)/2H)_{T=M_1} \approx (0.37)[M_{\rm Pl}/(1.66\sqrt{g^*}(8\pi v^2))]\tilde{m}_1 \approx 234(\tilde{m}_1/{\rm eV})$, where 0.37 denotes the usual time-dilation factor, g^*(for SUSY) ≈ 230 and $v \approx 174$ GeV. For comparison, we note that if one includes only inverse decays (thus neglecting $\Delta L \neq 0$-scatterings) in the Boltzmann equations, one would obtain: $\kappa \approx 0.3/[K(\ln K)^{0.6}]$ for $K > 10$ [65], and $\kappa \approx 1/2K$ for $1 \lesssim K \lesssim 10$. As pointed

out in Ref. [70], these expressions, frequently used in the literature, however, tend to overestimate κ by nearly a factor of 7. In what follows, we will therefore use Eq. (21) to evaluate κ.

[73] M. Plumacher, Z. Phys. **C74**, 549 (1997).

[74] For a specific scenario of inflation and leptogenesis in the context of SUSY G(224), see R. Jeannerot, S. Khalil, G. Lazarides and Q. Shafi, JHEP **010**, 012 (2000) (hep-ph/0002151), and references therein. As noted in this paper, with the VEV's of $(1, 2, 4)_H$ and $(1, 2, \bar{4})_H$ breaking G(224) to the standard model, and also driving inflation, just the COBE measurement of $\delta T/T \approx 6.6 \times 10^{-6}$, interestingly enough, implies that the relevant VEV should be of order 10^{16} GeV. In this case, the inflaton made of two complex scalar fields (i.e., $\theta = (\delta\tilde{\nu}^c_H + \delta\bar{\tilde{\nu}}^c_H)/\sqrt{2}$, given by the fluctuations of the Higgs fields, and a singlet S), each with a mass $\sim 10^{12}$-10^{13} GeV, would decay *directly* into a pair of heavy RH neutrinos – that is into N_2N_2 (or N_1N_1) if $m_{\rm infl} > 2M_2$ (or $2M_1$). The subsequent decays of N_2's (or N_1's), thus produced, into $l + \Phi_H$ and $\bar{l} + \bar{\Phi}_H$ would produce lepton-asymmetry *during the process of reheating.* I will comment later on the consistency of this possibility with the fermion mass-pattern exhibited in Sec. 2. I would like to thank Qaisar Shafi for a discussion on these issues.

[75] K. Kumekawa, T. Moroi and T. Yanagida, Prog. Theor. Phys. **92**, 437 (1994); G. F. Giudice, M. Peleso, A. Riotto and T. Tkachev, JHEP **9908**, 014 (1999) [hep-ph/9905242]; T. Asaka, K. Hamaguchi, M. Kawasaki and T. Yanagida, Phys. Lett. **B464**, 12 (1999) [hep-ph/9906366].

[76] Incorporating such an effective superpotential in accord with the assignment of flavor-charges suggested in Refs. [53] and [57] would involve two additional singlets with appropriate charges. The $(\text{VEV})^2$ of one or both of these may represent M^2. Derivation of such a picture with appropriate flavor-charge assignments from an underlying (string/M) theory still remains a challenge.

[77] Note that for this non-thermal case, since the gravitino-constraint is relaxed, N_1 can be chosen heavier than for the case considered before (the thermal case), still in accord with Eq. (17). Since $Y_B \propto \epsilon_1 T_{\rm RH}/m_{\rm infl}$, while $\epsilon_1 \propto (M_1/M_2)$, $T_{\rm RH} \propto M_1(m_{\rm infl})^{1/2}$ and $m_{\rm infl} \propto \lambda$, we see that $Y_B \propto (M_1^2/M_2)/\sqrt{\lambda}$, for a constant M, for the case of non-thermal leptogenesis.

[78] N. Sakai and T. Yanagida, Nucl. Phys. **B 197**, 533 (1982); S. Weinberg, Phys. Rev. **D 26**, 287 (1982).

[79] K. S. Babu and J. C. Pati, "*Radiative Processes ($\tau \to \mu\gamma$, $\mu \to e\gamma$, and $(g-2)_\mu$) as Probes of ESSM/SO(10)*", hep-ph/0207289.

[80] K. S. Babu and J. C. Pati, "*Neutrino Counting, NuTeV Measurements, Higgs Mass and V_{us} as Probes of Vectorlike families in ESSM/SO(10)*", hep-ph/0203029.

[81] J.C. Pati, Phys. Lett. **B228**, 228 (1989); K.S. Babu, J.C. Pati and H. Stremnitzer, Phys. Rev. **D 51**, 2451 (1995); K.S. Babu, J.C. Pati and X. Zhang, Phys. Rev. **D46**, 21990 (1992).

[82] C. Kolda and J. March-Russell, Phys. Rev. **D 55** 4252 (1997); R. Hempfing, Phys. Lett **351**, 206 (1995); M. Bastero-Gil and B. Brahmachari, Nuc. Phys. **B575**, 35, (2000).

[83] SuperK Collaboration: Y. Hayato, Proc. ICHEP, Vancouver (1998); M. Earl, NNN2000 Workshop, Irvine, Calif (Feb, 2000); Y. Totsuka (private comm. May, 2001); M. Vagins, Report on SuperK Results presented at WHEPP-7 meeting, Allahabad, India (January 6, 2002).

[84] S. Dimopoulos, S. Raby and F. Wilczek, Phys. Lett. **B112**, 133 (1982).

[85] J. Ellis, D.V. Nanopoulos and S. Rudaz, Nucl. Phys. **B 202**, 43 (1982).

[86] P. Nath, A.H. Chemseddine and R. Arnowitt, Phys. Rev. **D 32**, 2348 (1985); P. Nath and R. Arnowitt, hep-ph/9708469.

[87] J. Hisano, H. Murayama and T. Yanagida, Nucl. Phys. **B 402**, 46 (1993). For a recent estimate of the lifetime for the $d = 6$ gauge boson mediated $e^+\pi^0$-mode, see J. Hisano, hep-ph/0004266.

[88] K.S. Babu and S.M. Barr, Phys. Rev. **D 50**, 3529 (1994); **D 51**, 2463 (1995).

[89] V. Lucas and S. Raby, Phys. Rev. **D55**, 6986 (1997); R. Darmisek, A. Mafi and S. Raby, hep-ph/0007213, V2.

[90] H. Murayama and A. Pierce, hep-ph/0108104.

[91] S. Aoki *et al.*, JLQCD collaboration, hep-latt/9911026; Phys. Rev. **D 62**, 014506 (2000).

[92] K. Turznyski, hep-ph/0110282, V2.

[93] J. Arafune and T. Nihei, Prog. Theor. Phys. **93**, 665.

[94] J.L. Feng, K.T. Matchev and T. Moroi, Phys. Rev. **D 61**, 75005 (2000), hep-ph/9909334.

[95] SuperK bound on $\Gamma^{-1}(p \to e^+\pi^o)$; See e.g. Ref. [83].

[96] L. I. Ibanez and C. Munoz, Nucl. Phys. **B245**, 425 (1984).

[97] M. Claudson, M. B. Wise and L. J. Hall, Nucl. Phys. **B195**, 297 (1982); S. Chadha and M. Daniels, Nucl. Phys. **B229**, 105 (1983); S. J. Brodsky, J. Ellis, J. S. Hagelin and C. Sachrajda, Nucl. Phys. **B238**, 561 (1984).

[98] B. Bajc, P. Fileviez Perez and G. Senjanovic, hep-ph/0204311 and hep-ph/0210374.

[99] D. Emmanuel-Costa and S. Wiesenfeldt, hep-ph/0302272.

[100] See e.g. L. Hall and Y. Nomura, hep-ph/0103125; hep-ph/0205067. These papers provide examples of higher dimensional GUTs which lead to prominent d=6 proton decay via the $e^+\pi^o$ mode. Alternative cases lead to dominance of the μ^+K^o-mode: See e.g. Y. Nomura, hep-ph/0108070; A. Hebecker and J. March-Russell, hep-ph/0204037.

[101] T. Friedman and E. Witten (hep-th/0211269).

[102] See e.g. Candelas et al in Ref. [15]; Ref. [17] and [38].

[103] For attempts of this type leading to standard model-like gauge symmetry in 4D, see e.g. Y. Kawamura, Prog. Theor. Phys. **105**, 999 (2001); hep-ph/0012125; L.J. Hall and Y. Nomura, Phys. Rev. **D64**, 055003, hep-ph/0103125; G. Altarelli and F. Feruglio, Phys. Lett. **B511**, 257 (2001); hep-ph/0102301; A. B. Kobakhidze, hep-ph/0102323; M. Kakizaki and M. Yamaguchi, hep-ph/0104103; A. Hebecker and J. March-Russell, hep-ph/0107039; hep-ph/0204037; L.J. Hall and Y. Nomura, hep-ph/0111068; T. Asaka and W. Buchmuller and L. Covi, *Phys. Lett.*, **B523** (2001); L.J. Hall, Y. Nomura, T. Okui and D. Smith, hep-ph/0108071.

[104] For attempts based on higher dimensional GUTs leading to the G(224) symmetry in 4D, see Ref. [18]. For a short discussion of such attempts and relevant references, see Sec. 6.7 of Ref. [31].

[105] International Workshop on Neutrino and Subterranean Science (NeSS 2002 Meeting), Washington, D.C., September (2002). For transparencies of various talks see http://www.physics.umd.edu/NeSS02.

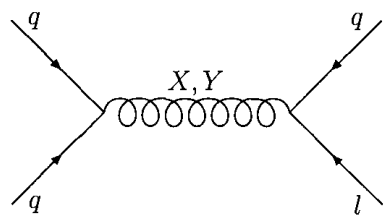


Figure 1: $d = 6$ proton decay operator.

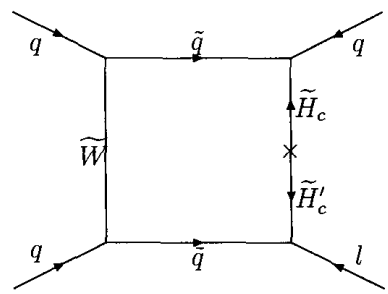


Figure 2: The standard $d = 5$ proton decay operator. The $\widetilde{H}'_c$ ($\widetilde{H}_c$) are color triplet(anti-triplet) Higgsinos belonging to $5_H(\overline{5}_H)$ of $SU(5)$ or 10_H of $SO(10)$.

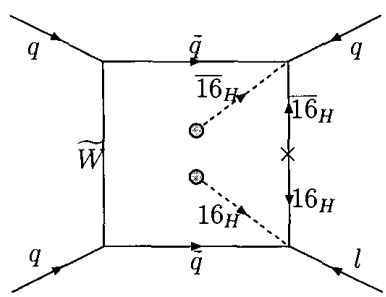


Figure 3: The "new" $d = 5$ operators related to the Majorana masses of the RH neutrinos. Note that the vertex at the upper right utilizes the coupling in Eq.(9) which assigns Majorana masses to ν_R's, while the lower right vertex utilizes the g_{ij} couplings in Eq.(8) which are needed to generate CKM mixings.

CHAIRMAN: J. PATI

Scientific Secretaries: A. Torrielli, M. Papucci

DISCUSSION

- *Brooijmans:*

In your lecture, you mentioned that neutrino oscillation signals, while compatible with G(224) model, disfavour large extra dimension. Could you elaborate on this?

- *Pati:*

As I discussed in the morning, within the string-unified (G224) or SO(10) symmetry, the Dirac mass of ν_τ is fixed by the $SU(4)^c$ symmetry to be equal to the top mass at the unification-scale, while the Majorana mass of ν_R^τ is determined (within the minimal Higgs system of 45_H, 16_H, $\overline{16}_H$ and 10_H) to be about 10^{15} GeV. This in turn yields (using see-saw and 2-3 family mixing) a mass for ν_R^τ given by $m(\nu_3) \approx (1/24\ \text{eV})(1/2 - 2)$ and $m(\nu_2) \sim 1/10\ m(\nu_3)$, in good agreement with the SuperK data for $\sqrt{\Delta m_{23}^2} \approx 1/20$ eV. By contrast, with TeV-scale large extra dimensions, one way people have tried to get a light neutrino mass is to introduce a singlet (right-handed) neutrino by hand and place it in the bulk. Thereby the neutrino acquires a Dirac mass, which is suppressed and has a value $K\nu_{EW}\ M^*/M_{Pl}$. However, for natural values of the Yukawa coupling $K \lesssim 1$ and the cut-off scale $M^* \sim 1$TeV, with $\nu_{EW} \approx 246$ GeV and $M_{Pl} \approx 10^{19}$ GeV, one gets: $m_\nu \lesssim 2 \times 10^{-5}$ eV, which is too small, by a factor of 2500 compared to the SuperK value. To raise m_ν to about 1/20 eV, one would need $K \sim 2\ 10^3$ and/or too large a value for $M^* \gtrsim 1000$ TeV, which would, however, face the gauge-hierarchy problem. One can find variants of this approach in the literature which attempt to *accommodate* the observed value of Δm_{23}^2 but, to my knowledge, none of them provide a compelling reason for the same. This is why I said that the neutrino oscillation observed at SuperK seems to clearly favor the string-unified G(224)/SO(10) route to higher unification with GUT/String-size extra dimensions, as opposed to large extra dimensions or even SUSY grand unification based on symmetries like SU(5). [I have explained this in more detail in the written version of my talk.]

- *Haidt:*

What is the mass of the top quark in your model?

- *Pati:*

Strictly speaking, I have used the observed top quark mass as one of the input parameters. But I should mention that there exist promising three-family string solutions (for example those obtained by A. Faraggi) which related the top Yukawa coupling to the gauge coupling. These attempts predicted the top mass to be about 170

GeV before the discovery of the top quark. There is also another reason why the top mass should be heavy. This is based on the idea of radiative origin for electroweak symmetry-breaking within a supersymmetry/supergravity theory, which suggested that the top mass should be heavy. Both of these approaches are, of course, fully compatible with the approach I have presented here, that presumes the validity of an effective G(224) or SO(10)-symmetry near the conventional string/GUT-scale.

- *Haidt:*
  So in some sense for you a high mass is natural?

- *Pati:*
  If you mean high mass for the top quark, the answer is yes. If you mean high GUT-scale and string-scale, the answer is also yes.

- *Ramtohul:*
  Does $SU(4)^c$ have the same phenomenological predictive power as $SU(3)^c \times U(1)$?

- *Pati:*
  Since $SU(4)^c$ contains $SU(3)^c \times U(1)_{B-L}$, it of course has all the predictions of the latter, but it has more. For instance, (a) $SU(4)^c$ leads to two mass relations: (i) $m_b(M_x) \approx m_\tau$ and (ii) $m(\nu^\tau_{Dirac}) = m_t(M_x)$, both of which are empirically favored; and (b) $SU(4)^c$ has the leptoquark gauge bosons which give rise to new interactions that are important in the early universe. These features are absent in $SU(3)^c \times U(1)_{B-L}$.

- *Bechtle:*
  Is there the possibility to derive a low energy effective SUSY Lagrangian for your G(224) model that could be used to derive predictions about SUSY particles at colliders?

- *Pati:*
  The answer to your question depends primarily on the mechanism of SUSY breaking which determines the soft parameters (i.e. the scalar squark/slpeton masses, gaugino masses, Higgs and Higgsino masses, $B\mu$-term and the A-terms), and less on the nature of the effective symmetry near the GUT-scale. Some of the promising ideas of SUSY-breaking – in particular those of gaugino-mediated, anomaly-mediated, dilaton and anomalous U(1) D-term mediated and gravity-mediated – would apply more or less in the same manner, irrespective of whether the effective symmetry near the GUT-scale is SO(10), SU(5) or G(224). There can, however, be one difference. The relationship between the gluino, wino and bino-masses at the GUT and therefore at the electroweak scales can be different for SO(10) and SU(5) versus G(224).

- *Levell:*

You advocate a model with gauge group SU(4) x SU(2) x SU(2), which is a product of three groups like the Standard Model. Why is this "unification"?

- *Pati:*

This is a very good question. The point is that even if the effective symmetry in 4D emerging from a string theory is non-simple, like G(224), string theory can still ensure familiar gauge coupling unification near the string-scale. The need (minimally) for a G(224) symmetry in 4D in accounting for neutrino oscillations, fermion masses and leptogenesis can be preserved, and at the same time observed gauge coupling unification at $M_X \sim 2x10^{16}$ GeV can be accounted for, if the string-scale is not far above the GUT scale (we need $M_{St} \approx$ (2-3) M_{GUT} (say). I have discussed how such a picture can emerge through the ideas of string-duality or semi-perturbative unification. As a practical matter, a string-unified G(224) solution in 4D seems to have an advantage over the SO(10) solution in that it neatly avoids the generic doublet-triplet splitting problem of SUSY SO(10), because the dangerous color triplets are projected out through compactification. It remains to be seen whether a four-dimensional string SO(10)-solution can give rise to the desired doublet-triplet splitting. For the present, I keep an open mind as to whether the effective symmetry in 4D is G(224) or SO(10). They both have essentially the same advantages as regards an understanding of fermion masses, neutrino oscillations and leptogenesis. [These points will be discussed in more detail in the written version of the talk.]

- *Markov:*

How many free parameters do you have in SUSY SO(10)?

- *Pati:*

I have not counted the number but I will tell you roughly. You have to ask: "What kind of SUSY breaking are you using?" I have not given to you any definite theory of SUSY breaking yet. If you ask what I prefer, then, for example, with the gaugino-mediated SUSY breaking, I would have only 5 parameters for the SUSY breaking plus the parameters I showed for the fermion masses and the gauge coupling. Including CP violating phases, this will add up to about 20.

Status of Super String Theory

E. Verlinde
Princeton University

Only the discussions are reported here,
since the camera-ready copy of the contribution
did not arrive in time.

CHAIRMAN: E. VERLINDE

Scientific Secretaries: E. Kohlprath, T. Maillard

DISCUSSION I

- *Wendland:*
  Do strings interact with the gauge fields? Or is the gauge field itself a string?

- *Verlinde:*
  A gauge particle corresponds to a string in its lowest ground state. Strings interact with each other by joining and splitting, which as a special case contains the interaction of strings with gauge fields. In addition to the massless states, a string has many excited states that are massive. These states all interact with each other and with the gauge field.

- *Bechtle:*
  Can one visually explain what makes a string a boson or a fermion, a gauge boson or a spinor?

- *Verlinde:*
  Whether a string is a boson or a fermion depends on whether the spin is integer or half-integer. For bosonic strings, one can visualize spin as coming from angular momentum. The string is an extended object and hence can carry orbital angular momentum. When you then think of the string as a point-like object, this orbital angular momentum is re-interpreted as internal spin. In this way one gets only integer spins, and hence a boson. To get half-integer valued spin, one has to introduce fermionic coordinates on the world sheet of the string.

- *Gromov:*
  Is it true that the quantum theory of strings corresponds to the classical field theory of particles?

- *Verlinde:*
  No, the quantized string leads to a quantum theory of particles at low energies. I do not know a way in which a quantum theory leads to a classical theory of particles. On the other hand, string theory has many more features than just the set of point particle states. The way the strings interact shows that they are different from point particles.

- *Golovnev:*

How can string theory describe excitations of many fields of different spins? The string can move only as a whole entity and fields are independent...

- *Verlinde:*

One string describes one bosonic or fermionic particle. E.g., in the diagram with four external open strings, each external string corresponds to one gauge particle. It is not the case that we try to describe all gauge particles by a single string. To consider many particle states, it is not sufficient to quantize a single string. You have to take the product of Hilbert spaces. Also the highly excited string states correspond just to one massive state. There is only one momentum and therefore only one moving object for every string.

- *Maas:*

What are D-branes, what is their origin and is there a Lagrangian governing their dynamics?

- *Verlinde:*

Open strings end on D-branes and therefore there are gauge fields on these branes. In addition, there are scalars corresponding to the transversal coordinates, which can be used to describe the position of the D-brane. We can find a low energy action - the Born-Infeld action : It is non-polynomial and is of the Yang-Mills type plus some string corrections. The tension of the D-brane is proportional to $1/g_s$. If g_s is small - in the perturbative limit - the D-brane is very heavy and can almost not move. If we e.g. compare D1 brane to fundamental strings, branes are similar to strings. At weak coupling, the branes are heavy and the fundamental object is the string. In the dual picture - the strong coupling limit - the fundamental objects are the D-branes.

- *Ramtohul:*

Can string couplings be related to coupling constants in the standard model?

- *Verlinde:*

Because the gauge fields are given by open strings, the relevant coupling g_o is related to the string coupling by $(g_o)^2 = g_s$. This is the 10d coupling for open strings. To relate it to the 4d coupling, you have to take into account the volume of the internal manifold. The square of the inverse of the 4d coupling is given by $1/g_s$ times the volume of the internal manifold. In string theory, the coupling is given by the dilaton field and therefore dynamical. Our hope is to fix the string coupling constant by finding the correct dilaton potential.

- *Wendland:*

Where is the field going? It has diverged for a while, developing various theories. Then some were found to be dual. Is the field converging? Is there any hope to find the correct theory soon?

- *Verlinde:*

It is still believed that all string theories can be combined in one theory called M-theory. Some progress has been made to find a formulation of M-theory using ideas from D-branes, but in recent years not much more has been done. The magic trick is still missing. Another direction of great importance is the duality between gauge theory and strings. I will say more about this in my next lecture, in particular concerning the Maldacena conjecture.

- *Markov:*

Does the cluster decomposition principle hold in the string theory? What about the unitarity that has been a problem for the Veneziano models?

- *Verlinde:*

String theory is an S-matrix theory, i.e. it gives rules to calculate the S-matrix. This S-matrix has to be unitary. Unitarity was a problem for the Veneziano amplitude as it is only a tree level result. If we include all loop contributions, we expect unitarity to be recovered. I find your question regarding the cluster decomposition a bit more tricky, since I think the answer depends on which situation.

- *Cata:*

String theory is nowadays better understood with supersymmetry. Is it feasible to think of string theory without supersymmetry?

- *Verlinde:*

I do not think so. Recent work has shown that you cannot get rid of tachyons without supersymmetry. This has to do with the high number of excited string states that contribute in loop amplitudes. One can show that, in order to have no tachyons or infinities, the bosonic and fermionic excited states have to cancel each other, which only happens when you have supersymmetry.

- *'t Hooft:*

Erik, this is a point I usually bring forward in my discussions with string theorists. The perturbative string theory starts off with a perfectly clear picture of what happens locally: a string, consisting of string-bits attached together. But when you discuss non-perturbative features such as branes, duality, etc... this local picture appears to get lost completely. I heard you say that we have an S-matrix theory only. So my problem is: How do we understand the bookkeeping of the physically relevant degrees of freedom?

- *Verlinde:*

It is true that if the coupling constants are finite, the local picture breaks down. The reason is that space-time need not be local itself. I think that field theory with its concept of locality may not be the correct way. In black hole physics, you cannot have

an arbitrary number of degrees of freedom in a volume. There must be some holographic bounds on the number of states inside a volume. String theory needs to make contact with holography. It suggests that space and time and the local degrees of freedom at a point are not the correct way to describe if the coupling constants are finite as also the Planck scale is finite. Another way to see this is related to the question of how to describe D-branes. The scalar fields describe the embedding of D-branes becomes matrices. Hence, coordinates become non-commuting variables, and the notion of a space-time point become fuzzy. At finite g_s, I think it may eventually remove the picture of space-time.

- *Haidt:*

Is it possible to view a neutrino as a string attached to a D-brane and acquiring its mass by moving in the bulk?

- *Verlinde:*

It is certainly not something that has been considered before.

- *Korthals-Altes:*

I want to come back to the question: How to get the Standard Model from string theory. One of the hallmarks of the SM is chiral fermions. On the other hand, the D-branes are like domain walls on which - as we know from field theory - chiral fermions are localized. So my question is: are D-branes the natural locus to formulate the SM?

- *Verlinde:*

Even without branes, we can have chiral fermions in four dimensions e.g., by compactifying on a manifold like a Calabi-Yau space. Certainly branes are not the only way to get chiral fermions in string theory. But indeed we can get chiral fermions in four dimensions by intersecting higher dimensional branes at a certain angle so that the intersection is four-dimensional.

- *Nobbenhuis:*

As follow-up to the answer to 't Hooft's question: It seems that when you start talking about 'fuzziness' of space-time, that space-time itself has become a dynamical variable. Is this right? Is it the same as, for example, in perturbative quantum gravity?

- *Verlinde:*

In string theory, the coordinates become dynamical fields on the world volume of the brane. This is in contrast to ordinary quantum gravity where the metric is dynamical, whereas the coordinates are still commuting variables. After a choice of the metric, space-time is continuous. In string theory, we lose the continuous picture of space-time because the coordinates are matrices.

- *Englert:*

The effective action for the background implies that in the 'original' string theory, one cannot integrate over such backgrounds. However, quantum mechanics tells us that such backgrounds could be superposed. How is this problem conceptually solved in string theory?

- *Verlinde:*

This is a difficult point. Under fixed boundary conditions at infinity, the classical equations of motion can have more than one solution. You expect that, in a semi-classical approach to quantum gravity, one has to sum over these solutions. That this indeed works in string theory, can indeed be shown in simple examples when it is possible to describe the full theory by a dual gauge theory. But I think this question is worth thinking about further.

CHAIRMAN: E. VERLINDE

Scientific secretaries: F. Bigazzi, A. Torrielli

DISCUSSION II

- *Kohlprath*:
  What has string theory to say about the cosmological constant problem?

- *Verlinde*:
  The cosmological constant problem is clearly one of the most important questions that string theory needs to solve. It is a problem of scales: the energy scale associated with the cosmological constant is 30 orders of magnitude smaller than the Planck or String scale. We can understand some things about hierarchies in string theory using compactifications with warp factors in a similar way as in the Randall-Sundrum scenario. However, this may be sufficient to explain the hierarchy between the weak scale and the string scale, but it does not solve the cosmological constant problem. Supersymmetry helps by cancelling the quantum contributions from bosons and fermions. This may explain why the scale of the cosmological constant is much smaller than the string scale. But we know that supersymmetry is broken near or above the weak scale, which is still 15 orders of magnitude larger than the scale of the cosmological constant. There is still no good explanation of this fact in string theory.

  I believe that in order to solve the cosmological constant problem we have to get a better understanding of cosmology in string theory in general.

- *Bechtle*:
  What would it mean for string theory if at LHC and NLC one of the following possibilities is true:
  1) A rich supersymmetric spectrum is found: which sort of susy breaking mechanism and compactifications?
  2) No low energy supersymmetry is found?
  3) Not even a Higgs is found?

- *Verlinde*:
  It would of course be great if at the LHC or the NLC one finds signatures of supersymmetry. This would certainly support string theory, which in its present form needs supersymmetry for its consistency. When a rich supersymmetric spectrum is found. it will of course tell us a lot about the possible compactification and supersymmetry

breaking scenarios in string theory. If no signal of supersymmetry is discovered at LHC or the NLC, it does not mean that string theory is dead. It is possible that supersymmetry is broken at higher energy scales. Finally, if no Higgs particle is found, but some other mechanism of electro-weak symmetry breaking turns out to be correct, then also string theorists will start thinking on how to incorporate this in string theory.

- *Maillard*:

It has been proposed to fix some moduli by introducing some flux. Which moduli cannot be fixed by this method? And why?

- *Verlinde*:

String compactifications in general have many moduli. There are moduli corresponding to the overall size of the compactification manifold, and to the dilaton. Other moduli, describing the shape of the internal manifold, can be fixed by turning on fluxes. This generates a potential, which can be minimized and determines the shape in terms of the integer fluxes. But a combination of the dilaton and size cannot be fixed, because the potential induced by the flux is flat in that direction.

- *Nobbenhuis:*

After compactification on 6D Calabi Yau space, you are left with a 4D space. In your talk this was 4D Minkowski space. Then you discussed warped geometries similar to the Randall- Sundrum scenario, where you need very specific spacetime metrics. The form of these metrics seem to come out of the blue. Are they derivable from string theory? Are they unique?

- *Verlinde*:

In general, string backgrounds need to be solutions of the low energy effective action, which is given by a 10 (or in the case of M-theory 11) dimensional supergravity action. When we study compactifications to four dimensions, we impose that the solution is Poincare invariant, so that the four-dimensional world is described by Minkowski space. One way to do this is to assume that the metric is a product of a six-dimensional manifold with 4D Minkowski space. Requiring that the four-dimensional theory has N=1 supersymmetry imposes conditions on the six-dimensional manifold, which are solved by Calabi-Yau Manifolds. But these are not the most general solutions. In general, one can allow metrics with warp factors, and with compact internal spaces on which tensor and scalar fields have non-zero values.

This leads to a much larger class of backgrounds which have the form I presented in the lecture. These compactifications still have undetermined parameters, and hence in that sense they are not unique.

- *Maas*:

In your lecture you said that all string theories stem from one theory, M-theory. You also said that there are problems concerning its formulation. Can you detail this?

- *Verlinde*:

When M-theory was first proposed, it was known that at low energies it is described by 11D supergravity, and that upon compactification on a circle it reduces to type IIA string theory. From this one can conclude that M-theory must also have extended objects, which are the membrane and the fivebrane. When M-theory is compactified on a circle, it has Kaluza-Klein excitations, which turn out to correspond to the D0 branes in the IIA string theory. This fact was used to arrive at a concrete proposal for a formulation of M-theory in terms of a kind of Matrix quantum mechanics of these D0-branes. One can check that this theory indeed gives type IIA strings after compactification on a circle, but still has a problem: it is not Lorentz invariant. Other problems are that no one understands how to compactify this theory to four dimensions, because the dynamics of the D0 zero branes becomes very complicated in this situation. So, at the moment we are still waiting for new ideas.

- *Englert*:

You presented the AdS/CFT correspondence as an example of holographic theories. If quantum gravity were holographic, we should expect the boundary theory to be non local. So the question is: is there a hidden non locality in the AdS/CFT correspondence which is not apparent in the quantum field theory of the boundary itself?

- *Verlinde*:

I agree with you that it is surprising that the theory on the boundary is a local theory. In the bulk, one does expect to have some kind of locality, or causality, and this may be the reason that the boundary theory is local. Also the fact that the isometries of the AdS space correspond to conformal invariance on the boundary may have something to do with that. There is, however, indeed a non-locality hidden in AdS/CFT, namely in the map from the boundary to the bulk degrees of freedom. A small object in the bulk is represented by something on the boundary, which in general is not small. The holographic image on the boundary even grows when the objects move further in to the bulk away from the boundary. Because the bulk-to-boundary map is non-local, it is not immediately clear how to relate locality in the boundary to causality in the bulk. This is an interesting question that certainly needs to be looked at further.

- *Golovnev*:

1) How is it possible to check that string excitations will behave like ordinary quantum fields that are used to describe the many particles that we see in

experiments? Does one need to consider some worldwide ocean of strings for this?

2) It is possible to deduce all the standard lagrangians from the superstring theory?

- *Verlinde*:

It costs energy to stretch a string, so at low energy a string really behaves like a point particle. Experimentally it is not possible to distinguish between a string and point particle in this regime. By considering many strings in the low energy limit, we can describe many particles, just as one can using quantum field theory. Also the interactions are similar as in QFT, which is sufficient to show that in the limit in which the string tension becomes very big (i.e. α' goes to zero), one reproduces QFT.

The effective low energy action describes gravity and gauge theory, because string theory incorporates gauge invariance and general coordinate invariance. In this way one can derive standard Lagrangians from string theory, similar to the ones used in describing the Standard Model, and General Relativity.

- *Nobbenhuis*:

There has been some excitement recently about new insights string theory could give about black holes. Could you say a few words about that?

- *Verlinde*:

Using D-branes, it is possible to construct objects in string theory that at finite coupling behave exactly like charged black holes. In string theory it is possible to count the number of quantum states of a black hole. The number of microscopic states is in exact agreement with the entropy of a black hole that was calculated more than 25 years ago by Hawking using methods of general relativity and quantum field theory. In this way the area law for entropy was verified in string theory. However, even in string theory we do not really understand why the number of states grows like the exponent of the area. That is a question that still needs to be solved.

- *Kohlprath*:

Why has the string scale to be larger (or of the order) of the GUT scale?

- *Verlinde*:

The string scale is usually derived from the known values of the Planck scale and the value of the gauge coupling at the unification scale. In most models one finds that the string scale differs from the Planck scale by at most one or two orders of magnitude, and hence is larger than the unification scale. This assumes, however, that the size of the compactification manifold in string theory is of order one in string scale units. It is

possible to lower the string scale by allowing large manifolds, which is what is being used in the large extra dimension scenarios.

- *Catà*:

People working with the Randall-Sundrum model and similar claim that these models are supported by string theory. However, on the one hand, D3 branes can only live in some of the five string theories we know today, and on the other hand, they are dynamical entities, which are able to attract and repel each other. As far as I know, the R-S model does not work with dynamical D-branes. So how justified is this claim?

- *Verlinde*:

The branes that are used in the Randall-Sundrum scenario could naively be thought of as D3-branes in type IIB (although in principle one could relate this by duality to other string theories). The AdS background that is used in the RS scenario can be obtained in string theory from the near horizon geometry of these D3 branes. However, to get gravity on the TeV brane, one needs to introduce a Planck brane with negative tension. Such branes are not present in string theory. The closest way to reproduce the scenario is to compactify the theory, and allowing the internal geometry to develop singularities, where some branes can be put. In this way the Planck brane is effectively replaced by the compactification manifold.

- *Stamen*:

What information is gained from AdS/CFT about gauge theories? Are you able to predict fermion masses or coupling constants?

- *Verlinde*:

AdS/CFT teaches us a lot about gauge theories for which we have a dual string theory, such as N=4 supersymmetric Yang-Mills, and certain other supersymmetric gauge theories. Let me first talk about the N=4 Yang-Mills case. This theory has two parameters: g and N. In the dual picture, the radius of the AdS space is given by the 't Hooft coupling g^2N. When the 't Hooft coupling is large, we can use classical supergravity to calculate various non-perturbative quantities in the dual gauge theory which we would not be able to calculate otherwise. The approximation you obtain is reliable. Some examples are glueball spectra, Wilson loops, instanton spectra, etc. So in this way you can make highly nontrivial predictions. You can also check the correspondence by comparing quantities that are calculable on both sides.

- *Nobbenhuis*:

How justified is the claim that string theory contains gravity? It apparently is highly nontrivial to get the right space-time metric, you can only somehow construct objects that

"look like" black holes and I think that you cannot take the large distance limit in order to get General Relativity. Basically, you only have scattering amplitudes for gravitons and nobody has ever seen those.

- *Verlinde*:

The low energy limit of superstrings is supergravity, the Einstein action is contained in the low energy effective action. In this way, classical General Relativity is naturally encoded in string theory.

Perturbative QUANTUM GRAVITY

Gerard 't Hooft

Institute for Theoretical Physics
Utrecht University, Leuvenlaan 4
3584 CC Utrecht, the Netherlands

and

Spinoza Institute
Postbox 80.195
3508 TD Utrecht, the Netherlands

e-mail: g.thooft@phys.uu.nl
internet: http://www.phys.uu.nl/~thooft/

ABSTRACT

A good understanding of Perturbative Quantum Gravity is essential for anyone who wishes to proceed towards any kind of non-perturbative approach. This lecture is a brief resumé of the main features of the perturbative regime.

1. INTRODUCTION
Perturbative Quantum Gravity as a gauge theory.

The Einstein-Hilbert action describing General Relativity is

$$S = \int \mathcal{L}(x)\,\mathrm{d}^4x\ ; \qquad \mathcal{L}(x) = \sqrt{-g}\left(\frac{R}{16\pi G} + \mathcal{L}^{\text{matter}}\right)\ . \tag{1.1}$$

R is the Ricci scalar curvature. g is the determinant of the metric tensor $g_{\mu\nu}$. The rule is that the matter Lagrangian must be made completely covariant by inserting the metric tensor $g_{\mu\nu}(x)$ or its inverse, $g^{\mu\nu}(x)$ wherever needed. $g_{\mu\nu}$, with its proper Minkowski signature, is promoted to being a dynamical variable. The variational principle with $g_{\mu\nu}$ and the matter fields as dynamical variables gives us the classical field equations obeyed by these variables. We assume here that the most essential principles of General Relativity are known[1]; let us recapitulate the most basic features that we need.

The "gauge transformation" in this theory is the space-time dependent coordinate transformation,

$$x^\mu \to x^\mu + \varepsilon \eta^\mu(x), \tag{1.2}$$

where ε is infinitesimal, and $\eta^\mu(x)$ is the space-time dependent generator of this transformation. The metric tensor transforms as

$$g_{\mu\nu} \to g_{\mu\nu} + \varepsilon\left(\eta^\alpha \partial_\alpha g_{\mu\nu} + g_{\alpha\nu}\partial_\mu \eta^\alpha + g_{\mu\alpha}\partial_\nu \eta^a\right) . \tag{1.3}$$

The last two terms here tell us that $g_{\mu\nu}$ transforms as a tensor. In perturbation theory, we will write (using Euclidean notation):

$$g_{\mu\nu} = \delta_{\mu\nu} + \varepsilon h_{\mu\nu} , \tag{1.4}$$

where $h_{\mu\nu}$ is taken to be infinitesimal. The transformation rule for $h_{\mu\nu}$ can be written as

$$h_{\mu\nu} \to h_{\mu\nu} + D_\mu \eta_\nu + D_\nu \eta_\mu , \tag{1.5}$$

where we used the notion of a covariant derivative:

$$D_\mu \eta_\nu \equiv \partial_\mu \eta_\nu - \Gamma^\alpha_{\mu\nu}\eta_\alpha . \tag{1.6}$$

It adds to the two gradients of η_ν in Eq. (1.3) not only the first term in Eq. (1.3), but also the extra terms one gets by lowering the index of the η^α field using the metric $g_{\alpha\nu}$.

The expressions giving R in terms of the metric tensor $g_{\mu\nu}$ are quite non-linear:[1]

$$\Gamma_{\alpha\mu\nu} = \tfrac{1}{2}(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\alpha\mu} - \partial_\alpha g_{\mu\nu}) \; ; \qquad \Gamma^\lambda_{\mu\nu} = g^{\lambda\alpha}\Gamma_{\alpha\mu\nu} . \tag{1.7}$$

$$R^\lambda{}_{\alpha\mu\nu} = \partial_\mu \Gamma^\lambda_{\alpha\nu} - \partial_\nu \Gamma^\lambda_{\alpha\mu} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\alpha\nu} - \Gamma^\lambda_{\nu\sigma}\Gamma^\sigma_{\alpha\mu} \; ; \tag{1.8}$$

$$R = g^{\alpha\nu} R^\mu{}_{\alpha\mu\nu} . \tag{1.9}$$

Substituting (1.4) and writing

$$g^{\mu\nu} = \delta^{\mu\nu} - \varepsilon h_{\mu\nu} + \varepsilon^2 h_{\mu\alpha}h_{\alpha\nu} + \cdots , \tag{1.10}$$

we can expand the action (1.1) in powers of $h_{\mu\nu}$. This results in an expression that we can write as

$$\mathcal{L} = \tfrac{1}{2} h_{\alpha\beta} V_{\alpha\beta\mu\nu} h_{\mu\nu} + (\text{higher orders}) , \tag{1.11}$$

where $V_{\alpha\beta\mu\nu}$ is a fairly complicated expression. The Euler-Lagrange equations following from varying this Lagrangian do not have unique solutions unless we impose a gauge condition. To understand what will happen physically, it is best first to consider the *radiation gauge*:

$$\sum_{i=1}^{3} \partial_i h_{i\mu} = 0 \; ; \qquad \mu = 1, \cdots, 4. \tag{1.12}$$

[1]There is a way to make these equations look *nearly* linear, by using a more sophisticated choice of variables[2], but the physics remains the same, and interactions due to non-linearity remain present.

Choosing

$$\varepsilon = \sqrt{16\pi G}\,, \tag{1.13}$$

and going to Fourier space,

$$f(x) = \tfrac{1}{(2\pi)^2}\int \mathrm{d}^4k\, e^{ikx}\hat{f}(k)\,, \tag{1.14}$$

one finds for $V_{\alpha\beta\mu\nu}$

$$V_{\alpha\beta\mu\nu} = \tfrac{1}{2}k^2(\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\beta}\delta_{\mu\nu}) + k_\mu k_\nu\delta_{\alpha\beta} - k_\beta k_\nu\delta_{\alpha\mu} + b^2\vec{k}_\beta\vec{k}_\nu\delta_{\alpha\mu}\,, \tag{1.15}$$

where $\vec{k}$ is k with its time component replaced by 0, and the parameter b^2 is sent to infinity, so as to impose Eq. (1.12).

These expressions look complicated, but they become a lot more transparent if we rotate $\vec{k}$ into the z-direction,

$$\vec{k}_\mu = (0,\ 0,\ \kappa,\ 0)\ . \tag{1.16}$$

To find the propagator in this gauge, we first have to symmetrize $V_{\alpha\beta\mu\nu}$ with respect to interchanges $\alpha \leftrightarrow \beta$, $\mu \leftrightarrow \nu$ and $(\alpha\beta) \leftrightarrow (\mu\nu)$. The propagator $\mathbb{P}$ is solved from

$$\mathbb{V}\cdot\mathbb{P} = \mathbb{I}\ ; \qquad \mathbb{I} = \tfrac{1}{2}(\delta_{\alpha\mu}\delta_{\beta\nu} + \delta_{\alpha\nu}\delta_{\beta\mu})\ . \tag{1.17}$$

The solution to this tensor equation is

$$P_{\mu\nu\alpha\beta} = \frac{1}{k^2}\left(\hat{\delta}_{\alpha\mu}\hat{\delta}_{\beta\nu} + \hat{\delta}_{\alpha\nu}\hat{\delta}_{\beta\mu} - \frac{2}{n-2}\hat{\delta}_{\alpha\beta}\hat{\delta}_{\mu\nu}\right) + \text{terms containing only } \vec{k}^2 \text{ in their denominators,} \tag{1.18}$$

where $\hat{\delta}$ is defined as

$$\hat{\delta}_{\mu\nu} \equiv \mathrm{diag}(1,\ 1,\ 0,\ 0)\,, \tag{1.19}$$

and n is the number of space-time dimensions, $n = 4$ being the physical value. Only the part explicitly written in Eq. (1.18) represents excitations that actually propagate. One sees first of all that only the completely transverse components of the field $h_{\mu\nu}$ propagate: $\mu, \nu = 1$ or 2. Secondly, the diagonal component (the trace) drops out:

$$P_{\mu\mu\,\alpha\beta} = 0 \qquad \text{since} \quad \hat{\delta}_{\mu\mu} = n - 2\ . \tag{1.20}$$

Since traceless, symmetric 2×2 matrices have only two independent components, we read off that there are only two propagating modes, the two helicities of the graviton. The propagator (1.18) propagates a graviton with the speed of light.

For practical calculations of Feynman diagrams and divergences, the radiation gauge (1.12) is not so suitable, since it violates Lorentz invariance. Let us again consider the quadratic term of the Lagrangian (1.1) prior to fixing the gauge. It can be written as:

$$\mathcal{L} = \tfrac{1}{8}(\partial_\sigma h_{\alpha\alpha})^2 - \tfrac{1}{4}(\partial_\sigma h_{\alpha\beta})(\partial_\sigma h_{\alpha\beta}) + \tfrac{1}{2}A_\mu^2 - \tfrac{1}{2}T_{\mu\nu}h_{\mu\nu} \tag{1.21}$$

$$+\ (\text{total derivative})\ +\ (\text{higher orders in } h)\ +\ \mathcal{L}^{\text{gauge fix}}\,, \tag{1.22}$$

where

$$A_\mu \equiv \partial_\sigma h_{\sigma\mu} - \tfrac{1}{2}\partial_\mu h_{\sigma\sigma} \ , \tag{1.23}$$

$$\text{and} \qquad \mathcal{L}^{\text{gauge fix}} = -\tfrac{1}{2}C_\sigma^2 + \mathcal{L}^{\text{ghost}} \ . \tag{1.24}$$

To fix the gauge, we can choose any non-gauge invariant function C_μ. It obviously is convenient to choose

$$C_\mu = A_\mu \ , \tag{1.25}$$

because then the gauge fixing term cancels out a similar term in the Lagrangian (1.22), and the remainder is easy to invert in order to obtain a smooth propagator for $g_{\mu\nu}$ that looks renormalizable — the theory however is still not renormalizable because of the derivatives in the interaction terms.

The ghost Lagrangian is obtained, as usual, by determining how the gauge-fixing term transforms under a gauge transformation:

$$A_\mu \to A_\mu + \partial^2\eta_\mu + \ \text{(higher orders)} \ ; \tag{1.26}$$

this leads to

$$\mathcal{L}^{\text{ghost}} = -\partial_\alpha\overline{\varphi}_\mu\partial_\alpha\varphi_\mu + \text{(higher orders)}. \tag{1.27}$$

2. DIVERGENCES

From this point, one may proceed exactly as in a Yang-Mills theory. Dimensional renormalization is quite convenient; we choose a continuously variable number of dimensions, n, and express the singularities that arise as poles of the form $(n-4)^{-r}$, where r is an integer varying from one to the number of irreducible loops in the diagram. We cancel these poles by inserting counter terms $\Delta\mathcal{L}$ in the Lagrangian:

$$\mathcal{L} \to \mathcal{L} + \Delta\mathcal{L} \ . \tag{2.1}$$

What kind of counter term Lagrangians can we expect?[3]

From unitarity, causality and dispersion relations, one deduces that $\Delta\mathcal{L}$ must be a *local* Lagrangian. What else can we say?

1. Dimensionality. The expansion parameter here, ε^2, as given by Eq. (1.13), is essentially Newton's constant, G. It has the dimension of a length squared (after putting $c = \hbar = 1$). For purely dimensional reasons then, we expect two extra derivatives at every consecutive order in G.

2. Gauge-invariance. Because the gauge is fixed by a gauge-fixing term, we do *not* expect the infinities to be gauge-invariant, but only physically observable quantities have to be handled in such a way that the infinities cancel out. There *is* a trick to limit the infinities to only gauge-invariant expressions; it is called the background field method[4]. So, we limit ourselves to gauge-invariant expressions for $\Delta\mathcal{L}$.

3. Only those infinities have to be considered that do not vanish on mass shell, for the following reason:

There is a theorem: if, at a given order, a term in $\Delta\mathcal{L}$ vanishes 'on mass shell' (which means that $\Delta\mathcal{L} = 0$ whenever the field equations of motion are substituted in the fields that occur in $\Delta\mathcal{L}$), then that term is unphysical at that order, or, to be precise, that term can be transformed away by a field transformation.[5]

The proof of the theorem goes as follows. The Euler-Lagrange equations read

$$\frac{\delta\mathcal{L}}{\delta\varphi_i} - \partial_\mu \frac{\delta\mathcal{L}}{\delta\partial_\mu\varphi_i} = 0 \ , \tag{2.2}$$

where φ_i simply stand for all conceivable dynamical fields that occur in $\mathcal{L}$, which include the metric tensor $g_{\mu\nu}$. Assume that $\Delta\mathcal{L}$ vanishes as soon as these equations are satisfied. This means that there must exist field combinations that we call $\delta\varphi_i$, being functions of the existing fields φ, $\partial\varphi, \cdots$, such that

$$\Delta\mathcal{L} = \delta\varphi_i \left(\frac{\delta\mathcal{L}}{\delta\varphi_i} - \partial_\mu \frac{\delta\mathcal{L}}{\delta\partial_\mu\varphi_i} \right) . \tag{2.3}$$

This implies that, at lowest order, we can write the action S as

$$S = \int \mathrm{d}^4x(\mathcal{L} + \Delta\mathcal{L}) = \int \mathrm{d}^4x\mathcal{L}(\varphi_i + \delta\varphi_i) . \tag{2.4}$$

This is a field redefinition, such as $\varphi \to Z\varphi + F$. Such field redefinitions have no physically observable effects on the predictions of a theory; they just define what our fields φ are. If, after such field redefinitions, an infinity disappears, then this infinity is not in any observable quantity such as the magnetic moment of a particle.

Knowing all these restrictions, which independent counter terms can one expect to encounter?

A In the case of pure gravity, $\mathcal{L} = \sqrt{-g}\,R$. Consider the counter terms needed for the infinities in the one-loop diagrams. Conditions 1 and 2 imply that the only possible terms to expect are

$$\Delta\mathcal{L} = \sqrt{-g}\,(\alpha R^2 + \beta R_{\mu\nu}^2 + \gamma R_{\alpha\beta\mu\nu}^2) \ . \tag{2.5}$$

Here, $R_{\alpha\beta\mu\nu}$ is the Riemann tensor (1.8), $R_{\mu\nu}$ is the Ricci tensor, which is the Riemann tensor with two indices contracted, and R is the Ricci scalar (1.9). To convince oneself that there is only one variety for the last term in Eq. (2.5), one uses the known symmetry features of the Riemann tensor.

Condition 3 tells us that, since there is no matter field, the first two terms in (2.5) are unphysical, because $R = 0$ and $R_{\mu\nu} = 0$ due to Einstein's equations. However, it so happens that the combination

$$\int \mathrm{d}^4x\sqrt{-g}(R^2 - 4R_{\mu\nu}^2 + R_{\mu\nu\alpha\beta}^2) , \tag{2.6}$$

is a *topological* invariant. Being a pure derivative, the integral (2.6) is completely determined by the fields on the boundary, and therefore such a term in the Lagrangian does not affect the field equations. This implies that also the third term in (2.5) is unphysical. We conclude that pure gravity has no infinity at all at the one-loop level! It is one-loop renormalizable.

B What about pure gravity in two loops? Which independent invariants may we expect? Dimensional analysis tells us that the terms may be of the form $DDDDR$, $RDDR$ or RRR. Here, D stands for a (covariant) derivative and R stands for a non-contracted Riemann tensor. The first terms are pure derivatives, so they can be ignored. Next, one observes that the Bianchi identities may be used to show that also the second set of terms vanish on shell. The third set is harder. Let us use representation theory in Euclidean 4-space, writing the various possible components as representations of $SO(4) = SU(2)_{\rm L} \otimes SU(2)_{\rm R}$, we see that a Lorentz index μ stands for a $2_{\rm L} \otimes 2_{\rm R}$, and an antisymmetric combination of two indices, $\mu\nu$ splits up into a self-dual and an anti self-dual part: $6 = 3_{\rm L} \oplus 3_{\rm R}$. The Riemann tensor itself, $R_{\mu\nu\ \alpha\beta}$ is a symmetric combination of two 6's, or

$$((3_{\rm L} + 3_{\rm R}) \times (3_{\rm L} + 3_{\rm R}))_{\rm Symm} \ . \tag{2.7}$$

Since in $SO(3)$, $(3 \times 3)_{\rm Symm} = 5 + 1$, we can write (2.7) as

$$5_{\rm L} + 1 + 5_{\rm R} + 1 + 3_{\rm L} \times 3_{\rm R} = 21 \, . \tag{2.8}$$

One of the 1 representations is the pseudoscalar which vanishes due to a symmetry equation for the Riemann tensor:

$$\varepsilon_{\lambda\nu\alpha\beta} R_{\mu\nu\alpha\beta} = 0 \ . \tag{2.9}$$

This leaves 20 terms. Here, $3_{\rm L} \times 3_{\rm R} + 1$ are the 10 components of the symmetric, contracted Ricci tensor $R_{\mu\nu}$, the 1 being its trace R. These vanish on shell. We are left with a $5_{\rm L}$ and a $5_{\rm R}$, which are the self-dual and the anti self-dual parts of the Weyl tensor.

How many invariants are there of the form $(5_{\rm L} + 5_{\rm R})^3$? In SO(3), we have $5^2 = 9+7+5+3+1$. Only the 5 in here contributes to a scalar in ${5_{\rm L}}^3$. So, this gives one scalar. Similarly, we expect one scalar out of ${5_{\rm R}}^3$. The cross terms cannot be scalar. The two terms we get are related by parity (they form a scalar and a pseudoscalar) So, eventually, since our theory is symmetric under parity, we can have only one infinite counter term from ${5_{\rm L}}^3 + {5_{\rm R}}^3$. *Pure gravity without matter only requires one new counter term at the two loop level.* It goes associated with one new constant of Nature. Considering the accuracy of the order of G^3 that we reached, this is an impressive result. Pure Gravity is pseudo-renormalizable.

C Pure gravity plus one scalar field φ. The Lagrangian is taken to be

$$\mathcal{L} = \sqrt{-g}\left(R - \tfrac{1}{2}(\partial_\mu \varphi)^2\right) \ . \tag{2.10}$$

This requires just one counter term at the one loop order, which has been calculated to be:[5]

$$\Delta\mathcal{L} = \frac{203}{640\pi^2(n-4)}\sqrt{-g}R^2 \ . \tag{2.11}$$

Again, having just one extra adjustable parameter seems to be not so bad a price to pay for a theory with one-loop accuracy.

Thus, perturbative gravity generates new infinities at higher orders, requiring counter terms whose finite parts are arbitrary, uncalculable coefficients, at each order in G. This does not seem to be the worst property of the theory. Much worse features are:

- The perturbation expansion in powers of G *diverges.*
- Therefore, we have absolutely no idea about the behaviour at distance scales comparable to, or shorter than, the Planck scale.

There actually are many good features of this theory, when compared with much more ambitious, so-called non-perturbative theories of gravity.

- The gauge-fixing procedure leads to a well-defined *foliation* of space-time: the definition of time does not lead to new difficulties. Our theory just behaves as any other non-Abelian gauge theory.[5]
- Its symmetry structure completely determines all finite parts of the amplitudes. There are no unwarranted assumptions.
- The theory can be used as a starting point for any more ambitious approach.
- The analytic structure of the amplitudes is well-defined. The Wick rotation can be performed without any difficulty, *in spite* of the fact that the classical action in Euclidean space does not seem to be bounded from below!

3. THE WICK ROTATION

Let us concentrate a bit more on this Wick rotation, explain the problem, and its resolution, as dictated precisely simply by studying perturbative Quantum Gravity. The Wick rotation corresponds to the replacement

$$t \to i\tau \ , \tag{3.1}$$

which turns space-time into a Euclidean space. For a scalar theory with action

$$S = \int \mathrm{d}^4x \left(-\tfrac{1}{2}(\partial\varphi)^2 - V(\varphi)\right) \ , \tag{3.2}$$

the amplitudes in Euclidean space are described by functional integrals of the type

$$\int \mathcal{D}\varphi(x) \exp \int \mathrm{d}^4x \left(-\tfrac{1}{2}(\partial\varphi)^2 - V(\varphi)\right) , \tag{3.3}$$

which can be approximated by decently convergent Gaussian integrals.

In QED, the Euclidean functional integral is

$$\int \mathcal{D}A_\mu(x) \exp\left(-\tfrac{1}{4}\int F_{\mu\nu}F_{\mu\nu}\mathrm{d}^4x\right) , \tag{3.4}$$

and since this integrand is minus a pure square, one here also has properly convergent Gaussian integrals. In gravity, however, there is no upper or lower bound on the value of $\int \mathrm{d}^4x\sqrt{g}R(x)$ in Euclidean space (that is, a space where the metric tensor field $g_{\mu\nu}(x)$ is real valued and has signature $(++++)$), so the functional integral is ill-defined.

Some authors concluded that, therefore, the functional integral will be dominated by "non-perturbative effects", sometimes referred to as "space-time foam"[6]. Here, we will show that the problem not only also shows up in perturbative gravity, but it clearly has a valid resolution here; no such thing as space-time foam is needed.

First, why does the Wick rotation work so well in theories such as QED? In QED, the Lagrangian is

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - \lambda\partial_i A_i + J_0 A_0 . \tag{3.5}$$

Here, λ is a Lagrange multiplier that fixes for us a radiation gauge condition, $\partial_i A_i = 0$. It is convenient, again, to work in Fourier space, where we again rotate the vector $\vec{k}$ into the z-direction:

$$\vec{k} = (0,\ 0,\ \kappa)\ ; \qquad \vec{A} = (A_1,\ A_2,\ 0)\,. \tag{3.6}$$

The Lagrangian then becomes

$$\mathcal{L} = -\tfrac{1}{2}\vec{k}^2\vec{A}^2 + \tfrac{1}{2}\dot{\vec{A}}^2 + \tfrac{1}{2}\kappa^2\vec{A}_0^2 + J_0 A_0 . \tag{3.7}$$

The first two terms here describe the two helicities of the photon, and the last just generates the Coulomb force between the sources $J_0(\vec{x})$. The Coulomb force is obtained by extremizing the action as a function of A_0. In Euclidean space, $t = i\tau$, the action will be bounded from above if we choose A_0 to be imaginary, $A_0 = -iA_4$. Indeed, in that case, we see that in Euclidean space $-\frac{1}{4}\int F_{\mu\nu}F_{\mu\nu}$ is real and bounded from above. We find that the Euclidean action integral, $\int \Delta A \exp(-\frac{1}{4}\int F_{\mu\nu}F_{\mu\nu})$ then converges properly. This is the standard Wick rotation.

In (perturbative) Quantum Gravity, the situation is more complicated. Take the radiation gauge:

$$\mathcal{L} = -\tfrac{1}{2}h_{\alpha\beta}V_{\alpha\beta\mu\nu}h_{\mu\nu} - \tfrac{1}{2}T_{\mu\nu}h_{\mu\nu} + \text{(higher orders)} \tag{3.8}$$

In Eq. (1.15), we again rotate the vector $\vec{k}$ into the positive z-direction, so that

$$\vec{k} = \begin{pmatrix} 0 \\ 0 \\ \kappa \end{pmatrix} \; ; \qquad h_{\mu\nu} = \begin{pmatrix} \frac{1}{2}h + h_1 & h_2 & 0 & h_{10} \\ h_2 & \frac{1}{2}h - h_1 & 0 & h_{20} \\ 0 & 0 & 0 & 0 \\ h_{10} & h_{20} & 0 & h_{00} \end{pmatrix} \tag{3.9}$$

In this gauge, the kinetic part of the Lagrangian splits up into three parts:

$$\mathcal{L} = \mathcal{L}_I + \mathcal{L}_{II} + \mathcal{L}_{III} \; , \tag{3.10}$$

$$\mathcal{L}_I = \tfrac{1}{2}(\dot{h}_1^2 + \dot{h}_2^2) - \tfrac{1}{2}\kappa^2(h_1^2 + h_2^2) - T_1 h_1 - T_2 h_2 \; ; \tag{3.11}$$

$$\mathcal{L}_{II} = \tfrac{1}{2}\kappa^2 h_{oa}^2 + h_{0a} T_{oa} \; ; \tag{3.12}$$

$$\mathcal{L}_{III} = -\tfrac{1}{8}\dot{h}^2 + \tfrac{1}{8}\kappa^2 h^2 - \tfrac{1}{2}\kappa^2 h h_{00} - \tfrac{1}{2}h_{00}T_{00} - \tfrac{1}{4}hT_{aa} \; . \tag{3.13}$$

The first Lagrangian, $\mathcal{L}_I$, describes the propagating modes, and it allows a Wick rotation into Euclidean space as usual, just like what we do with scalar fields in flat space-time. The second Lagrangian yields an instantaneous repulsive force between the Poynting currents T_{0a} of the form

$$V_P = -T_{0a}^2/2\kappa^2 \, , \tag{3.14}$$

provided that we integrate h_{0a} along the imaginary axis. This component can be handled in Euclidean space after Wick rotation without complications. (In Euclidean space, h_{4a} can be taken to be a real valued field.)

It is the third term where something unusual happens. In $\mathcal{L}_{III}$, we see that h_{00} acts as a Lagrange multiplier field, yielding the constraint

$$h = -\frac{T_{00}}{\kappa^2} \; . \tag{3.15}$$

In Euclidean space, this constraint would be obtained from the functional integral

$$\int \mathcal{D}h_{44} \exp\left(-\int \mathrm{d}^4x \, \tfrac{1}{2}h_{44}(\kappa^2 h + T_{44})\right) \; . \tag{3.16}$$

Note that this only gives the correct constraint, (3.15) if either h or h_{44} are taken to be imaginary. Normally, one would be inclined to give only real values to the fields in Euclidean space, but if in Eq. (3.16) all quantities inside the exponent were kept real, the integral would diverge badly; only complex exponential integrals yield delta functions. In fact, this is the way in which we encounter the fact that the Einstein Hilbert action, Eq. (1.1) is not properly bounded when rotated to Euclidean space. Here, we now see what has to be done: some of the dynamical fields, even after the Wick rotation, are not allowed to take on real values.

From a perturbative point of view, it is obvious that all functional integrals in Euclidean space-time are duly convergent, provided that the integration contours are chosen appropriately. The message for any non-perturbative formalism is quite clear: also in a

non-perturbative gravity theory, the functional integrals must converge. The functional integrand formed by the Einstein-Hilbert action in Euclidean space is unbounded in the space of *real* metrics $g_{\mu\nu}$; it is the exponent of a quantity that can have any sign. From Eq. (3.16), we concluded that either h or h_{44} must be imaginary. These quantities would form the *conformal factor* of the metric. Thus, by adjusting the integration contours, particularly in the space of the conformal factors, one can obtain properly defined functional integrals in Euclidean space.

There is an even more transparent way to formulate these conditions. Apparently, in Euclidean space-time, one must always rotate the contours in function space such that the action stays bounded from above. We can formulate this condition non-perturbatively. Write

$$g_{\mu\nu} = \Omega(x)\hat{g}_{\mu\nu}(x)\,, \tag{3.17}$$

where we define the extra factor $\Omega(x)$ in such a way that the Ricci scalar with respect to the metric $\hat{g}_{\mu\nu}$ vanishes:

$$R(\hat{g}_{\mu\nu}) = 0\ . \tag{3.18}$$

Writing $\Omega(x) = e^{i\theta(x)}$, we find that the gravitational part of the action becomes

$$S = \sqrt{g}\,R \to \sqrt{g}\left(-\tfrac{3}{2}(\partial_\mu\theta)^2\right)\ . \tag{3.19}$$

This is bounded from above provided that we choose θ to be real rather than imaginary. We conclude: the metric cannot be integrated over real values with positive signature; one must choose a complex conformal factor, or some similar revision of the functional integration contours.

4. THE SPECULATIVE PART OF QUANTUM GRAVITY

Clearly, perturbative Quantum Gravity cannot answer the question as to what really happens at the Planck scale. Whenever the gravitational field becomes so strong that perturbative procedures no longer apply, new theoretical approaches are required, and indeed, new laws of physics may have to be searched for. In all other cases, one might be able to extrapolate from what is already known in perturbative terms. Which are the fundamentally new situations where the gravitational fields become too strong?

The answer to this is that the gravitational force harbors a fundamental instability. If large amounts of matter converge into a small region of space, the attractive gravitational force will cause a further contraction, and eventually this may result into an explosion. The laws of General Relativity leave little doubt as to what this must lead to: a black hole. It is the black hole that creates the strongest possible gravitational fields. If we wish to know how to do non-perturbative gravity, it is the black holes that we must study. The author's approach to black holes is described in Refs.[7]. Some tendency of pure gravity to generate string-like structures could be observed, although the string action deduced from gravity does not exactly coincide with the starting points employed in string theory.

Applying purely logical arguments takes us very far, but eventually, we have to use some intuition to make further progress, and it is here that opinions on the way to proceed diverge most. The suspicion advocated here is that we must involve the discussions concerning the foundations of Quantum Mechanics. It is quite conceivable that what we are really searching for is a theory that combines hidden variables with superstrings or black holes. Such ideas are further displayed in Ref.[8].

References

[1] For introductions into General relativity, see: R. Adler, M. Bazin and M. Schiffer, *Introduction to General Relativity*, Mc.Graw-Hill 1965; R.M. Wald, *General relativity*, Univ. of Chicago Press 1984; G.'t Hooft, *Introduction to General Relativity*, Rinton Press, Inc., Princeton, ISBN 1-58949-000-2; also to be downloaded from `www.phys.uu.nl\~thooft\`

[2] A. Ashtekar, *Phys. Rev.* **D36** (1987) 1587; A. Ashtekar et al, Class. Quantum Grav. **6** (1989) L185.

[3] G. 't Hooft and M. Veltman, *DIAGRAMMAR*, CERN Report 73/9 (1973), reprinted in "Particle Interactions at Very High Energies", Nato Adv. Study Inst. Series, Sect. B, vol. 4b, p. 177.

[4] B.S. DeWitt, Phys. Rev. **162** (1967) 1195, 1239; J. Honerkanp, *Nucl. Phys.* **B 48** (1972) 269.

[5] G. 't Hooft and M. Veltman, *Ann. Inst. Henri Poincaré,* **20** (1974) 69.

[6] S. Hawking, *Spacetime foam*, D.A.M.T.P. preprint (1977).

[7] G. 't Hooft, The scattering matrix approach for the quantum black hole: an overview. *J. Mod. Phys.* **A11** (1996) 4623; `gr-qc/9607022`; id., TransPlanckian Particles and the Quantization of Time, *Class. Quant. Grav.*, **16** (1999) 395; gr-qc/9805079.

[8] G. 't Hooft, Quantum Gravity as a Dissipative Deterministic System, *Class. Quant. Grav.* **16** (1999) 3263; also published in: Fundamental Interactions: from symmetries to black holes (Conference held on the occasion of the "Emèritat" of François Englert, 24-27 March 1999, ed. by J.-M. Frère et al, Univ. Libre de Bruxelles, Belgium, p. 221; id., Determinism beneath Quantum Mechanics, presented at "Quo vadis Quantum Mechanics?", Temple University, Philadelphia, September 25, 2002, `quant-ph/0212095` ?

CHAIRMAN: G. 't HOOFT

Scientific Secretaries: O. Catà, S. Nobbenhuis

DISCUSSION

- *Haidt:*

What is the relation between your version of a theory of Quantum Gravity and String Theory?

- *'t Hooft:*

Actually, String Theory is purported to be the successor of perturbative Quantum Gravity, so what I discussed here is basically a pre-string theory description of Quantum Gravity. Surely, we all desperately want to do better. Having a theory which, as you go towards higher accuracies, generates more and more free parameters that have to be adjusted to your observations, having a theory which is not described properly at distance scales at or smaller than the Planck scale; that is unsatisfactory. As it is so difficult to obtain a good theory, it is not unreasonable to suspect that, if eventually we do succeed to produce a theory that works, this theory will be truly spectacular, giving us a lot more than just Quantum Gravity. It might give us the entire Standard Model. This is what many of us have been trying to do for the last 30 years or so. Presently, String Theory is the main competitor and it is far ahead of most other attempts. String Theory is a very powerful and interesting approach. It claims to generate not only a theory that looks like Quantum Gravity as I described it, in the perturbative regime, but in addition it generates all the free parameters that I mentioned. I had stated that with or without matter, you just get so many free parameters. With pure gravity you just have one or two parameters at the two-loop order. If you add scalar fields you get more free parameters, but string theory generates gravity, scalars, spinors and vectors and in the end it somehow avoids the generation of new free parameters. In addition, it claims that we do know what the theory does at the Planck length. The only thing is, String Theory is still basically perturbative; it has not yet completely succeeded in getting beyond the perturbation expansion. It is qualitatively better certainly in the sense that there is an idea behind it. I have not presented any further idea besides saying: I want a quantum theory and I want gravity, no more than that. String Theory has a new idea behind it: the whole idea of having particles replaced by strings and all the consequences of that; so it is more powerful. However, we should remember that it has not yet achieved its ultimate goal, which is to give a completely coherent description of physics at the Planck scale and also to give the bonus that we are all waiting for: to predict what kind of matter Lagrangians we are going to have. The new idea is suggestive, but we have no real clue as to whether it is also correct. This should not be forgotten.

- *Haidt:*

Would it be right to say that String Theory is basically a topological theory?

- *'t Hooft:*

No, I would not call it that. It is a dynamical theory, which has as its basic constituents something else than what we normally do. Normally we have particles described by fields and we may have bound states of these particles, or other more complicated structures, like solitons, but the basic starting point is fields. The basic starting point in strings is another idea, but it is a dynamical theory. I would not directly call it topological, although there are certainly many topological features in the theory.

- *Bechtle:*

What are the quantization conditions of Quantum Gravity, i.e. what corresponds to the quantum mechanical commutation relations?

- *'t Hooft:*

The perturbative theory that I discussed is exactly a quantum theory such as any non-Abelian gauge theory. Exactly as in a non-Abelian gauge theory, we replace commutation relations by functional integral expressions for the amplitudes, which is basically equivalent. We know that functional integrals are essentially equivalent to Schrödinger equations. Mathematically you can transform one representation - or one paradigm - into the other. So there are no differences between them. Just like in gauge theories, we can generate operators and we have the difficulty that only gauge invariant operators are physically observable. In gravity, however, there is a complication in the sense that writing down completely gauge invariant operators is quite difficult. That is because everything depends on space-time, normally. In gauge theory that was fine, an operator like $G_{\mu\nu}G_{\mu\nu}$ is gauge invariant, it depends on x and t, but that is fine. In gravity that is not fine because now x and t are themselves gauge dependent parameters. If you make a transformation to another coordinate frame, your x and t will be something else.

So everything depending on x and t is strictly speaking not gauge invariant. However, and that is the nice thing about perturbative Quantum Gravity, you can write down sequences in perturbation expansion which will be invariant up to any given order in perturbation expansion. Then you are back into a familiar domain of quantizing things just like in a gauge theory. There we have a Hilbert space which, depending on the gauge fixing procedure we choose, will be enlarged with ghost states. The operators may or may not lead you from physical states into ghost states but these operators have completely conventional commutation rules. So, in that respect, gravitation will not be different from gauge theories.

There is a problem in gravity though, which is that, because of its non-renormalizable nature, the commutators in gravity will generate more and more singular terms as you go to higher orders. Where on the lightcone the commutators of ordinary field theories have

moderate singularities, in gravity those singularities will become more bothersome and eventually things run out of control if you try to go to very high orders in perturbation theory.

- *Maas:*

In non-perturbative Yang-Mills theories, the Gribov problem complicates the non-perturbative treatment and perturbatively you are only saved since the perturbative expansion stays within the first Gribov horizon. Is there something similar in perturbative Quantum Gravity?

- *'t Hooft:*

In my opinion, the Gribov problem in the case of gauge theories is quite manageable. It is a nuisance and, depending on the choice of gauge, there are problems of over-counting, but I consider the Gribov problem as a purely technical problem in the case of gauge theory. I think that in the case of gravity it is a much more severe problem. For instance, the statement now is that you have to identify a physical configuration in Quantum Gravity as a given sort of curved space-time but then leave away the coordinate frame that you chose because you want to have a gauge-independent description of a curved space-time, so as soon as you introduce coordinates you have a gauge redundancy. Fixing the gauge then surely will leave all sorts of ambiguities that are probably much more harmful than simple Gribov ambiguities in gauge theories. I think, yes, there will be rather severe counting problems. You can see this problem in all its magnificence and glory if you remove one dimension and take gravity in 2+1 dimensions. That theory is thought by some people to be trivial, others say it is solvable, it is easy, you can quantize it … . However, the Gribov problem there is extremely difficult. And I do not think in 3+1 dimensions it will be any easier. But as you pointed out yourself, it is a non-perturbative feature. As long as you stick to perturbation expansion you simply do not encounter that difficulty. But I think that in a full theory, whether it is String Theory, gravity or anything else, there will be tremendous complications of a similar nature as the Gribov problem.

- *Stamen:*

What are you going to calculate that can be measured by experimentalists?

- *'t Hooft:*

We cannot do direct experiments that check Quantum Gravity effects, as far as we know now. There is a remote possibility in the theory of large extra dimensions that you may have heard about. In that case, it is conceivable that Quantum Gravity effects show up at a much lower energy than we now expect. We now expect Quantum Gravity effects to become important only if energies reach the order of the Planck energy. If you can accelerate two protons such that their energy reaches the Planck mass, then you are in

business, because you will encounter Quantum Gravity. But for that you will need accelerators the length of several light-years. Those kinds of experiments are out of the question. However, the theory could predict for instance the fine structure constant. It could be that ultimately all sorts of features of the Standard Model are predicted. It could be that certain predictions will be made about the Standard Model that can also be checked experimentally. This is for the distant future. Right now we can only say that the Standard Model should come out of our theory. String Theory has some moderate successes in this respect, but a true answer as to why our world is described by the Standard Model has not been given. As theorists, we are unable to predict from Quantum Gravity the kind of GUT that will be the true description of our world. At present, we have no way to test this theory experimentally.

- *Amelino Camelia:*

Would perturbative Quantum Gravity be able to describe situations in which the large-scale structure of space-time depends on the solution of the quantum mechanical problem (so that we cannot specify a background beforehand)?

- *'t Hooft:*

I am not sure I understand the question well because in the world as we know it, the background metric is very firmly established to be 4-dimensional Minkowski space.

Now there could be extra dimensions that are compactified. It could also be that there is, as some people have suggested, "space-time foam". I just argued against that, but it could be that at very tiny distance scales there could be a sort of bubbles in space-time or topologically non-trivial features, which you could call gravitational instantons.

Besides, also the cosmological constant problem is very much on our minds, because it tells us something about the infrared limit of gravitational forces that we do not understand. If you look at quantum fluctuations at distance scales tinier than a millimetre or so, they appear already to contribute to the cosmological constant more than the effects we see in the universe.

My conclusion is that there is something very basically wrong with our theory in that domain, which could indicate something of the kind you say; that we do not have completely the right background description. However I should insist that of course working with Minkowski space in the Standard Model was extremely successful and we do not want to give that away. So we should stick to Minkowski space, but keep in mind that perhaps the Einstein equations or the dimensionality of space-time at very tiny distance scales could be different from what we are familiar with. How exactly is difficult to answer.

The other question is at the extreme large scales. I sometimes like to speculate the universe might have a non-trivial topology, like the shape of a torus or a 3-sphere, or something more complicated in higher dimensional space-time. That is where the question of the boundary conditions of the universe comes in. It probably would be

irrelevant to physics at our scales; we want of course to understand how Quantum Gravity affects the cutoff at the Planck scale and at first sight it seems that that has nothing to do with the boundary conditions of the universe. But it does have something to do with the boundary conditions of the universe at the moment the universe was formed, at the moment of the big bang. There, all these discussions come together. I am afraid that this is all I can say about this question.

- *Markov:*
  What are the prospects of loop gravitation?

- *'t Hooft:*
  You mean the idea by Ashtekar, Rovelli and Smolin. Yes, these are several attempts to try to do better than what I have done now. String Theory I mentioned, but there is indeed also another approach, which is actually quite ingenious. Ashtekar and his followers have started with exactly the scheme of Quantum Gravity, but tried to formulate the commutators-operator-structure of the theory more rigorously. Moreover, they also tried to avoid the need to introduce coordinates from the very beginning. Then at some point they made a rather big step in their analysis. This big step is that from the operator structure they attempt to get a representation of the algebra, whereby they observe that a natural representation of the degrees of freedom of gravity would be the degrees of freedom of knots and links.

  If the gravitational degrees of freedom are represented this way, then you have the right algebraic structure as a starting point of a gravity theory. This is a very good attempt, except for a complication, which is how to write down the dynamical rules. Once you have a large 'knot', how would the knot evolve in time? What is time in this theory anyway? This is something not quite understood, and it is what I see as the main weakness of the theory.

  String Theory is in better shape with respect to this point. There you do have a Lagrangian and a prescription as to how things move.

- *Korthals-Altes:*
  So in a way Ashtekar takes out all redundant variables from gravity. Does he then not get the same difficulties as we get in unitary gauge in ordinary gauge theories?

- *'t Hooft:*
  No, Ashtekar does have a natural space-time cutoff in his theory, which results from the fact that you have only the loop variables and there is simply nothing in between. So there is no longer a space-time continuum. He was right in thinking that he could afford this. That part of the theory is, I think, the beautiful part: it has a natural cutoff, a natural

length scale, everything is natural, except for the dynamical evolution. And that is where the theory requires a lot more thinking.

- *Englert:*

I am not at ease with the mathematical trick that damps the fluctuations of gravity in Euclidean space by using complex conformal factors. The reason for the uneasiness is that these fluctuations have their counterparts in Minkowski space as large fluctuations of the (non-dynamical) conformal factor as a conformal class. So what is the rationale of suppressing these fluctuations, which, at least theoretically, seem to exist in the conventional form of the theory?

- *'t Hooft:*

I do not quite understand your question. I think a theory should have an inherent stability in it. If you say that the theory is described by degrees of freedom which are approximately flat, then you should stick to that, whereas if the theory is not fundamentally unstable you get conformal deformations and you would say: How come that the universe does not show any conformal fluctuations, why are we sitting here peacefully in this room, without this room being oscillating in a conformal sense? We do not see any of these oscillations happening around us. That is presumably because at the boundary there is some constraint that fixes this. More to the point, if you carefully do the perturbative gravity, there is no such thing as a conformal instability; everything is completely stable, in fact the gravitons all have positive energies. So all degrees of freedom that actually propagate in gravity have perfectly positive energy. Thus the conformal factor is not a propagating thing. In other words: if you somehow fix this at the boundary of the universe by some appropriately chosen boundary condition, there will be no conformal instability of this universe. If there had been such instabilities, we would not have been sitting here. The instability that we do see in the Euclidean continuation is unphysical and due to a bad attempt to get away from Minkowski space by a false Wick rotation.

The boundary plays an important role in gravitation. Gravitational radiation is the example where you have a flat space-time with any kind of radiation in it that you like, and it is completely stable. There is no sign of any conformal instability.

Moreover, you might be referring to black holes. But black holes are not sick. A black hole is a perfectly reasonable object from a certain point of view. It only says that if you collide many particles, they will make this little bowl, which may emit Hawking radiation. But that is not sick, that is just a peculiar property of matter.

In the end we see that gravity appears to be a perfectly stable theory, and the instability is only in our description.

- *Bechtle:*

Is there a purely theoretical reason to call a part of the Lagrangian, like the ghost part, "unphysical"?

- *'t Hooft:*

You can better ask this question in terms of ordinary non-Abelian gauge theories where you have the same situation. There you have the gauge invariance as a starting point; you start with the gauge theory as a classical concept and then you quantize it. Classically you already say that you are free to choose any gauge you like, it does not affect the physical features of the solution so you can say that choosing the gauge is an unobservable procedure. In gravity the same situation applies: we can choose the coordinates anyway we like. If you are to believe Einstein, then the way to understand gravity is to observe that one may choose the coordinates. Coordinates are man-made, artifacts, to enable us to describe what we see. You see that they are not physical features that exhibit dynamical properties. Making this observation and carrying it over to the quantum description of the theory, we have the same situation. So in non-Abelian theories what happens is that like in the classical case, you have to choose the gauge-fixing procedure and you can do that in many different ways. Quantum theory allows you many more different ways to fix the gauge than a classical theory. In a quantum theory, you fix the gauge such that all surviving operators directly correspond to physically observable things. In gravity, this is a little more subtle and complicated but you can do the same thing. In a quantized theory, it is often much more convenient to work with a 'renormalization gauge' or 'renormalizable' gauges, but the price you pay is that you work in a Hilbert space which appears to contain much more than ordinary physical observables; these we call ghosts. If you say that this is only a mathematical transformation from a theory that you started off with, these ghosts are not physical. If you stick to that description of the situation, you generate all the amplitudes that you want, and those will be the same amplitudes you would compute if you stick to a unitary gauge, where there is nothing superfluous. In contrast, you might decide to have a theory where the ghosts are given some physical meaning. That would be very hard because the ghosts have unphysical features in their Lagrangians and unphysical relations between spin and statistics and so, if you would take them seriously, they would generate, for example, negative probabilities. You would end up in awful difficulties, such as probabilities that are no longer conserved. The Schrödinger equation does not do what you want it to do, and you have a theory that you cannot really use, to describe particles and fields in a way we are not used to. You do not want to do that. As a theorist, you can always try to do something novel, but I do not think it will help to give a physical meaning to the ghosts.

- *Maas:*

Since the coordinates are now involved in the gauge transformation, is stochastic quantization still feasible?

- *'t Hooft:*

Stochastic quantization is being tried. As far as the procedure called stochastic quantization is realistic, it is just an alternative way to formulate what quantization is. It gives you a kind of interpretation that I have some difficulties with. For instance, it basically only works for the Euclidean case and I have not really seen a good way to completely generate the stochastic quantization procedure for Minkowski space. In gravity, to go to Euclidean space is somewhat problematic and in that respect stochastic quantization would be more problematic in gravity. Sticking to Euclidean space, I find that what is called stochastic quantization is not very helpful in understanding what is going on to my taste.

- *Markov:*

In the theorem you just proved you use the property valid for S-matrix elements, but which is not valid for Green's functions. Does this theorem therefore only apply to S-matrix elements?

- *'t Hooft:*

Well, in gauge theories this problem has been handled, so what you say about Green's functions happens precisely because you fix the gauge. If you want to handle gauge theories just like Quantum Gravity in a manner that is very similar to what we do in normal gauge theories, then you have to fix the gauge. Fixing the gauge means that you are imposing extra conditions, extra constraints on the amplitudes, which are really not physical but they are only there for our convenience. And as soon as you do that, you get Green's functions that do not have a direct physical interpretation. Only the S-matrix elements have a direct physical interpretation. Green's functions do exist and they may obey conditions of unitarity and causality, and you can write down what it means for a theory to be gauge invariant in terms of the Green's functions: the so-called BRS relations. BRS symmetry is the prime example of the fact that Green's functions obey a symmetry because of the structure of the theory. We can understand that the theory is only consistent if also the renormalization procedure and everything else is such that the symmetry is kept fully intact, not broken by anything. In other words, we can check whether the Green's functions obey the BRS conditions. Ultimately, only the S-matrix elements are physically observables so the Green's functions are auxiliary quantities that you need in order to ultimately produce and extrapolate the S-matrix elements. They are physical and there should be a way to compute those. The Green's functions help you to explain where the S-matrix elements come from, what the local dynamics is of the theory.

You can always ask whether or not Green's functions have to be finite. We want the S-matrix elements to be finite. Maybe the Green's functions have some intrinsic infinities in them. They are unphysical, so, who cares? Well, we do care; we want to use intermediate results that are finite and meaningful, otherwise you do not know how to do the computations properly. This is the reason why we introduce counterterms such that also the Green's functions are finite. This is only for our convenience, the requirement is not physical. So when I talk about physical parameters, I mean parameters that, in principle, can be observed in experiments; real experiments are probably impossible, but you can do thought experiments. The counterterms with which you finally end up are just these coefficients for the two-loop diagrams, that are the ones that will survive in the S-matrix. All the other ones are connected to field transformations; they will be needed to make all Green's functions finite, but they will not show up in the S-matrix. This is why they are just field transformations. Of course, I am interested ultimately in a theory that tells you what you see when you look at it, which are the S-matrix elements, and the counterterms that survive all the way into the S-matrix are the only ones that correspond to physically observable quantities.

- *Haidt:*

Our dream is to have gravity and quantum theory unified. Is it possible to say why quantum theory has proven to be so exceedingly successful?

- *'t Hooft:*

Well, we are all actually surprised by the enormous success of quantum theory and it is obviously a very central theme in whatever the ultimate theory is going to be. In my last lecture I explain some of my ideas about quantum mechanics, why it is successful, and why I suspect nevertheless that quantum mechanics is not part of the ultimate theory of Nature, whatever that might be. It could be that Quantum Mechanics is an outflow of an ultimate theory when one tries to do statistical computations. Yes, this is the old dream of the so-called hidden variable theories: the idea that there are many variables that are difficult to detect directly. Stochastic quantization is a theory that attempts to introduce hidden variables of some sort that show stochastic behaviour. I would like to have a theory that inherently generates stochastic behaviour at very small scales. We can imagine a theory that generates lots of stochastic, haphazard features at small distance scales, and where the only way to integrate this towards larger distance scales is to accept that things behave in a way that we cannot foresee or control. This is what I think Quantum Mechanics is. It is a description of a theory in which it is hopelessly complicated to compute anything with infinite accuracy, and only parts of the information of how things behave at the Planck scale will transpire all the way to our scale. The remaining unpredictable features of the theory are what we perceive as quantum mechanical fluctuations.

Now this is exactly the philosophy behind the hidden variable theories, but it has never been materialized in workable models. I also have not found a workable model, but I have not given up the suspicion that the ultimate theory of Nature is of such a sort. Also when you look at the whole structure of quantum theory, it appears to be perfectly suitable for such an interpretation, except for some very annoying bugs that stop you from just making good models of this kind. It could be that the quantum mechanical features of Nature cannot be approximated in any toy model, but that you have to combine everything we know, including the gravitational force, and that only then will you understand the peculiarities of Quantum Mechanics. If my suspicions are right, Quantum Mechanics is certainly in some sense an 'exact' theory. In the sense that it describes statistical phenomena, Quantum Mechanics is an exact feature of our world, not an approximate feature and that should explain its successes.

Proton structure and its flavor decomposition

H. Abramowicz[a*]

[a]School of Physics and Astronomy
The Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Israel

In spite of the tremendous progress achieved in establishing the validity of the Standard Model of weak, electromagnetic and strong interactions, there are still many mysteries. In particular, the structure of the proton, whose mass is in 98% generated by strong interactions [1], cannot be calculated from first principles and has to be investigated experimentally. In this lecture, the present status of our knowledge on the parton decomposition of the proton is shortly summarized.

1. Introduction

A very successful classification of hadron properties is obtained within the Quark-Parton Model (QPM) in which it is assumed that hadrons are bound states of free valence quarks, which carry their quantum numbers. There are six known quarks, the up – u, down – d, strange – s, charm – c, bottom – b and top – t quarks. No bound states associated with the top quark have been observed as yet.

In the presence of Quantum Chromodynamics (QCD), the naive QPM picture of hadrons has to be altered to take into account the radiation and absorption of gluons by quarks as well as the creation of $q\bar{q}$ pairs by gluons. Thus, in effect, hadrons consist of various partons, quarks and gluons. We know today that about 50% of the proton momentum is carried by gluons.

The interactions of quarks and gluons are described by QCD, a non-abelian gauge theory based on the SU(3) color symmetry group. Color constitutes the equivalent of the electric charge in electromagnetic interactions. The quarks, each in three colors, interact by exchange of electrically neutral vector bosons - the gluons, which form a color charge octet. The gluons are not color neutral and thus they themselves interact strongly. A consequence of this property is asymptotic freedom, which states that the interaction strength of two colored objects decreases the shorter the distance between them. The effective strong coupling constant α_S depends on the scale at which the QCD process occurs. The solution of the renormalization group equation, in leading order, leads to

$$\alpha_S(Q^2) = \frac{4\pi}{\beta_0 \ln(Q^2/\Lambda^2)} , \tag{1}$$

*Partly supported by the Israel Science Foundation, The German Israeli Foundation and the Israel Ministry of Science.

where Q^2 denotes the scale at which α_S is probed, and Λ is a QCD cut-off parameter. The parameter β_0 depends on the number of quark flavors in the theory, N_f,

$$\beta_0 = 11 - \frac{2}{3}N_f \ . \tag{2}$$

Since the known number of flavors is six, $\beta_0 > 0$, and the coupling constant becomes smaller the larger the scale Q^2. The property of asymptotic freedom has been proven rigorously and allows to make predictions for the properties of strong interactions in the perturbative QCD regime, in which α_S is small.

Another property of QCD, which has not been proven rigorously, is confinement, which keeps quarks bound into colorless hadrons and prevents the observations of free quarks.

The distribution of partons bound in hadrons cannot be calculated from first principles. The calculations would have to be performed in a regime of QCD where the perturbative approach breaks down. However the QCD factorization theorem [2] states that for hard scattering reactions the cross section can be decomposed into the flux of universal incoming partons and the cross section for the hard subprocess between partons. The measurement of parton distributions in hadrons becomes an essential element in investigating the properties of QCD.

2. Deep Inelastic Scattering

Among hard scattering processes, the most suitable for studying the internal proton structure is deep inelastic scattering (DIS) of a point-like lepton on a proton target. The point-like, partonic substructure of the nucleon was first firmly established in the pioneering SLAC experiment [3,4] in which the spectrum of electrons scattered off a nucleon target was measured. This experiment was very similar in its essence to the famous Rutherford experiment which established the structure of atoms.

In the most general case, the lepton-nucleon interaction proceeds via the exchange of a virtual vector boson as depicted in Fig. 1. Since the lepton number has to be conserved, we expect the presence of a scattered lepton in the final state, while the nucleon fragments into a hadronic final state X,

$$lN \to l'X \ . \tag{3}$$

Assuming that k, k', P, P' are the four vectors of the initial and final lepton, of the incoming nucleon and of the outgoing hadronic system, respectively (see Fig. 1), the usual variables describing the lepton nucleon scattering are

$$\begin{aligned} Q^2 &= -q^2 = -(k-k')^2 \ , \\ s &= (k+P)^2 \ , \\ W^2 &= (q+P)^2 \ , \\ x &= \frac{Q^2}{2P\cdot q} \ , \end{aligned} \tag{4}$$

$$y = \frac{q\cdot P}{k\cdot P} \ , \tag{5}$$

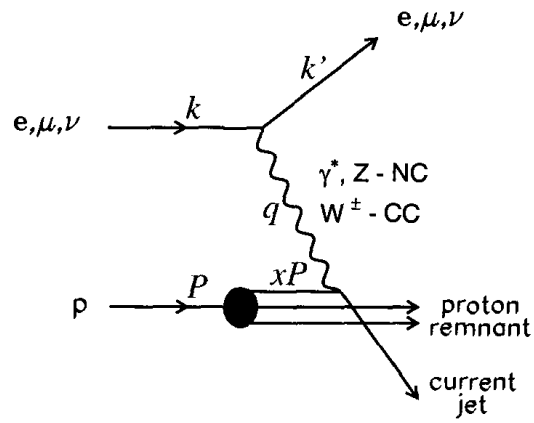


Figure 1. *Schematic diagram describing deep inelastic lepton nucleon scattering. The four vectors of the particles, or particle systems, are given in parenthesis.*

The square of the four momentum transfer (the mass squared of the virtual boson), $q^2 < 0$, determines the hardness of the interaction, or in other words, the resolving power of the interaction. The exchanged boson plays the role of a "partonometer" with a resolution given by its Compton wavelength of $\hbar/\sqrt{Q^2} = 0.197/\sqrt{Q^2}\,\mathrm{GeV\,fm}$. Thus for different values of Q^2 the interaction is sensitive to structures at different scales.

The variables s and W^2 are the center-of-mass energy squared of the lepton-nucleon and intermediate boson-nucleon systems, respectively. The meaning of y is best understood in the rest frame of the target, in which it measures the inelasticity of the interaction and its distribution reflects the spin structure of the interaction. The variable x is the dimensionless variable introduced by Bjorken. In the QPM, x is the fraction of the proton momentum carried by the massless struck quark.

The inclusive differential cross section, integrated over all possible hadronic final states, is a function of two variables which uniquely determine the kinematics of the events. These variables are most easily recognizable as the energy and production angle of the scattered lepton. However, in anticipation of the partonic structure of hadrons, the differential cross section is usually expressed in terms of two Lorentz invariant variables, x and Q^2, defined in Eq. (4),

$$\frac{d^2\sigma^{l(\bar{l})N}}{dxdQ^2} = A\left\{\frac{y^2}{2}2xF_1(x,Q^2) + (1-y)F_2(x,Q^2) \pm (y-\frac{y^2}{2})xF_3(x,Q^2)\right\}\ , \qquad (6)$$

where, for $Q^2 \ll M^2_{W,Z}$ (the mass squared of the intermediate vector bosons), $A = G_F^2/2\pi x$ for neutrinos and anti-neutrinos, with G_F the Fermi constant, and $A = 4\pi\alpha^2/xQ^4$ for charged leptons, with α the electromagnetic coupling constant.

The structure functions, F_i, depend on the kinematics of the scattering and the usual chosen variables are x and Q^2. The structure functions, F_1, F_2, and F_3, are process dependent. The F_3 structure function is non-zero only for weak interactions and is generated by the parity violating interactions.

3. Structure functions and QCD

In deep inelastic scattering (i.e. $Q^2 \gg 1\,\mathrm{GeV}^2$), the nucleon is viewed as composed of point like free constituents - quarks and gluons. In the QPM, the lepton nucleon interaction is described as an incoherent sum of contributions from individual free quarks.

The electroweak gauge bosons couple to quarks through a mixture of vector (v) and axial-vector (a) couplings. The structure functions can then be expressed in terms of quark distributions $q_i(x)$, where i stands for individual quark types. For non-interacting partons, as is the case in QPM, Bjorken scaling (no Q^2 dependence) is expected.

$$\begin{aligned} F_1(x) &= \frac{1}{2}\sum_i q_i(x)(v_i^2 + a_i^2)\,, \\ F_2(x) &= \sum_i x q_i(x)(v_i^2 + a_i^2)\,, \\ F_3(x) &= 2\sum_i q_i(x)(v_i a_i)\,. \end{aligned} \tag{7}$$

The index i runs over all flavors of quarks and anti-quarks which are allowed, by conservation laws, to participate in the interaction. For the simplest case of electromagnetic interactions, $v_i = e_i$, where e_i is the charge of quark i in units of the electron charge, and $a_i = 0$. For charged currents, $v_i = a_i = 1$ for quarks and $v_i = -a_i = 1$ for anti-quarks. For neutral current interactions mediated by the Z^0, $v_i = T_{i3} - 2e_i \sin^2\Theta_W$ and $a_i = T_{i3}$, where T_{i3} is the third component of the weak isospin of quark i and Θ_W denotes the Weinberg mixing angle, one of the fundamental parameters of the Standard Model. The couplings have a more complicated structure for neutral current interactions in which the interference between the Z^0 and the γ play an important role. A direct consequence of formulae (7), derived for spin 1/2 partons, is the Callan-Gross relation [5], i.e.

$$2xF_1(x) = F_2(x)\,. \tag{8}$$

For universal parton distributions in the proton, expected in the QPM and QCD, formulae (7) can be used to relate DIS cross sections obtained with different probes. In fact, many more relations and sum rules can be derived assuming $SU(3)$ or $SU(4)$ flavor symmetry for hadrons. Inversely, the validity of these assumptions can be tested experimentally. A detailed discussion of these issues is beyond the scope of this lecture. The naive QPM approach, which allows the construction of structure functions from quark distributions, has to be altered to take into account some dynamical features predicted by QCD, such as violation of scaling and of the Callan-Gross relation.

In perturbative QCD, the violation of scaling is ruled by the parton evolution equations, derived on the basis of the factorization theorem, and are known as the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [6–8]. The DGLAP equations describe the way the quark q and gluon g momentum distributions in a hadron evolve with the scale of the interactions Q^2.

$$\frac{\partial}{\partial \ln Q^2}\begin{pmatrix} q \\ g \end{pmatrix} = \frac{\alpha_S(Q^2)}{2\pi}\begin{bmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{bmatrix} \otimes \begin{pmatrix} q \\ g \end{pmatrix}\,, \tag{9}$$

where both q and g are functions of x and Q^2. The splitting functions P_{ij} provide the probability of finding parton i in parton j with a given fraction of parton j momentum. This probability will depend on the number of splittings allowed in the approximation. Given a specific factorization and renormalization scheme, the splitting functions P_{ij} are obtained in QCD by perturbative expansion in α_S,

$$\frac{\alpha_S}{2\pi} P_{ij}(x, Q^2) = \frac{\alpha_S}{2\pi} P_{ij}^{(1)}(x) + \left(\frac{\alpha_S}{2\pi}\right)^2 P_{ij}^{(2)}(x) + \dots . \tag{10}$$

The truncation after the first two terms in the expansion defines the next to leading order (NLO) DGLAP evolution.

It should also be noted that beyond leading order (LO) the splitting functions depend on the factorization scale and thus the definition of parton distributions is not unique. This affects the simple relation (7) between quarks and structure functions. The relation (7) is preserved in LO, but the parton distribution functions acquire a Q^2 dependence. In NLO, the Callan-Gross relation is violated and its violation is measured in terms of a longitudinal structure function $F_L = F_2 - 2xF_1$. Formula (6) for the deep inelastic cross section remains valid to all orders.

4. Scaling violation in F_2

The predications of the DGLAP evolution equations are that F_2 at large x will decrease with increasing Q^2, while it will increase at low x. This is a natural consequence of the radiation process, which at the expense of lowering the momenta of fast partons, produces many low momenta partons. This is indeed observed in the data, as shown in Fig. 2, where measurements of F_2 [9–11] are presented as a function of Q^2 in the range $0.013 < x < 0.65$.

The tendency observed in the data is well reproduced by the solution of the evolution equations, with appropriate initial conditions fitted to the data.

The initial conditions that, once evolved, describe best the F_2 measurements, are in fact parameterizations of the parton distributions, quarks and gluons, as a function of x, at an initial Q_0^2. Obviously, it is not possible to derive 13 different distributions (six for quarks, six for anti-quarks and one for gluons) in x from just fitting the F_2 data. Prior knowledge and extra assumptions are required. This is the main subject of this lecture.

Before a more detailed discussion of input conditions, it is worthwhile to note that experimental data require that the perturbative expansion of the splitting function P_{ij} be at least performed to order α_S^2. The most convincing argument comes from dedicated measurements of the longitudinal structure function F_L [12–14,11]. This is shown in Fig. 3, where a compilation of F_L measurements is compared to predictions of NLO QCD calculations, with an appropriate input.

5. x dependence of parton distributions

The shape of the x distributions of quarks and gluons have to be guessed from physics considerations. This is then expressed in a functional form, with free parameters to be deduced from the fits to the data.

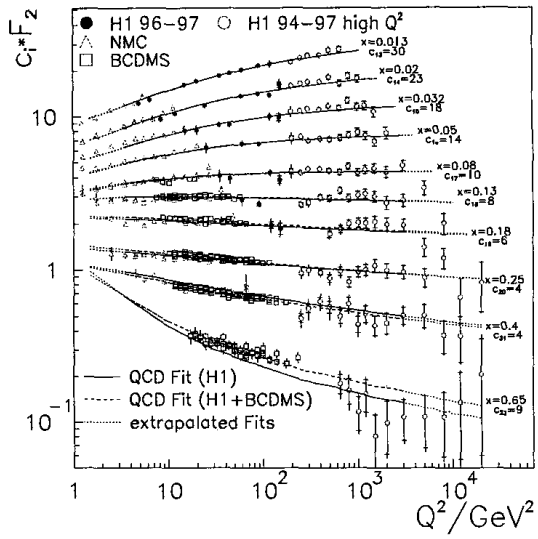


Figure 2. *The Q^2 dependence of F_2 for a range of x values. The full and shaded line represent the results of fitting QCD expectations to the data.*

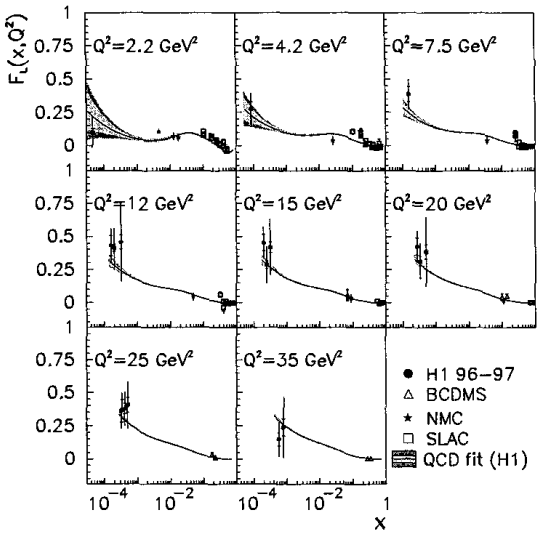


Figure 3. *The x dependence of F_L for a range of Q^2 values. The data are compared to expectations of QCD.*

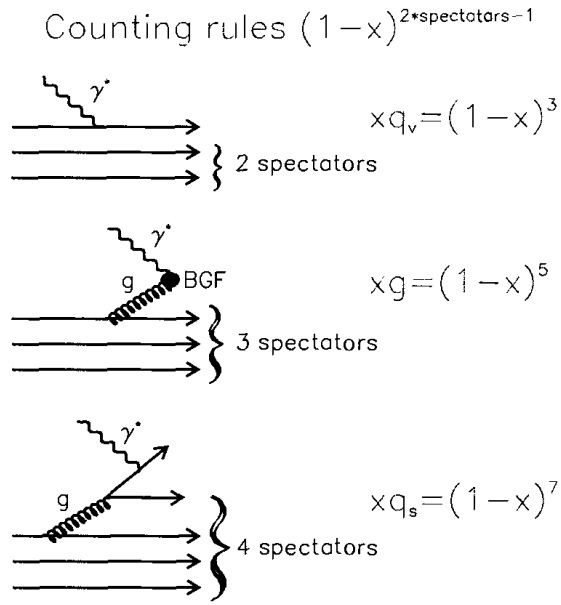


Figure 4. *Schematic representation of the counting rules applied to a proton consisting of three quarks, for valence quarks, q_V, gluons, g and sea quarks, q_S.*

5.1. $x \to 1$

It is easy to guess that as x tends to unity, the probability, $f(x)$, of finding partons goes to zero, be it because of momentum conservation. A more exact shape may be derived from the so-called counting rules [15,16], based on dimensional arguments. The counting rules predict that

$$\lim_{x\to 1} xf(x) = (1-x)^{2*\text{spectators}-1}\,, \tag{11}$$

where the number of spectators is given by the number of partons not participating in the DIS interaction. This is depicted in Fig. 4. For the photon to interact with the gluon, a boson gluon fusion (BGF) has to take place leading to a $q\bar{q}$ pair in the final state. The sea quarks, on the other hand, originate from a gluon splitting into a $q\bar{q}$ pair.

The $(1-x)^b$ parameterization works very well for $x \simeq 1$, with values of b close to the ones expected, especially in the valence case.

5.2. $x \to 0$

There is no simple counting rules which can be used for low x. However, the F_2 structure function is directly related to the interaction cross section of a virtual photon with the proton, $\sigma(\gamma^* p)$. At low x, this relation is particularly simple,

$$F_2 = \frac{Q^2}{4\pi^2\alpha}\sigma(\gamma^* p)\,. \tag{12}$$

One may try then to invoke the properties of hadronic cross sections, for which the W dependence is well described within the Regge phenomenology [17] by the exchange of

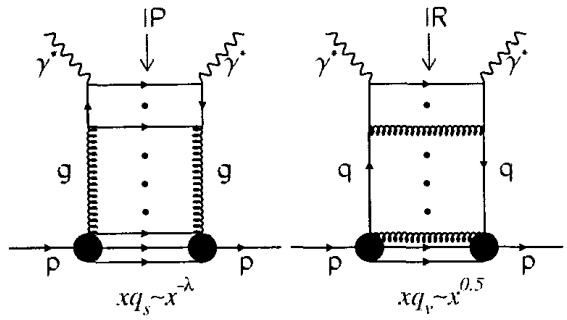


Figure 5. *Schematic representation of Pomeron, $I\!P$, and Reggeon, $I\!R$, exchange in partonic language.*

the Reggeon, $I\!R$, and the Pomeron, $I\!P$, trajectories [18],

$$\sigma(\gamma^* p) \sim A_{I\!P}(Q^2)(W^2)^{\alpha_{\mathbf{P}}(0)-1} + A_{I\!R}(Q^2)(W^2)^{\alpha_{\mathbf{R}}(0)-1}\,, \tag{13}$$

where $\alpha_{\mathbf{P}}(0)$ is the intercept of the $I\!P$ trajectory, while $\alpha_{\mathbf{R}}$ is that of the $I\!R$. A_i give the overall normalization as well as the relative contribution of the two components. $I\!R$ is responsible for interactions with exchange of quantum numbers in the t channel, while $I\!P$ corresponds to the exchange of quantum numbers of the vacuum. The hadron-hadron interactions are well described by $\alpha_{\mathbf{P}}(0) = 1 + \lambda \simeq 1.1$ and $\alpha_{\mathbf{R}}(0) \simeq 0.5$ [18,19]. The relation between the Regge language and the parton language is depicted in Fig. 5.

In the parton language, $I\!P$ exchange corresponds to reactions induced by sea quarks, while $I\!R$ exchange corresponds to reactions induced by valence quarks. Since $x = Q^2/(Q^2 + W^2 - m_p^2)$, the W dependence of the $\sigma(\gamma^* p)$ may be translated into the x dependence of sea and valence quarks,

$$xq_S \sim x^{-\lambda}\,, \tag{14}$$

$$xq_V \sim x^{0.5}\,. \tag{15}$$

The xF_3 structure function, that can only be measured directly in weak interactions, is proportional to the valence quark distributions. It has been measured in charged current neutrino interactions [20] and its x distribution for various vales of Q^2 is shown in Fig. 6. As expected the valence quark distribution tends to zero when $x \to 0$ and this behavior is preserved by the QCD evolution.

At low x, F_2 is dominated by the contribution of sea quarks. The expectations for the sea quarks is more controversial. From hadron-hadron interactions one expects a mild rise of the structure function F_2 as x becomes small. On the other hand, the properties of the DGLAP evolution equations are such that one expects a sharp rise of F_2 with decreasing x. For example, in the double logarithmic approximation (DLL), valid for low x and high Q^2, F_2 can be calculated and

$$F_2^{DLL} \propto \exp\left(2\sqrt{\frac{C_A \alpha_s}{\pi} \ln\frac{1}{x} \ln\frac{Q^2}{Q_0^2}}\right)\,. \tag{16}$$

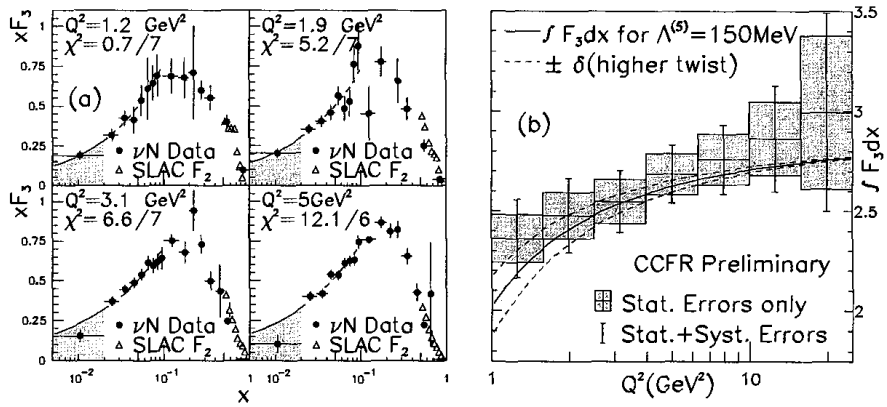


Figure 6. *Left: The x dependence of xF_3 for different Q^2 values as measured by the CCFR experiment. Right: Integral of F_3 over the full x range.*

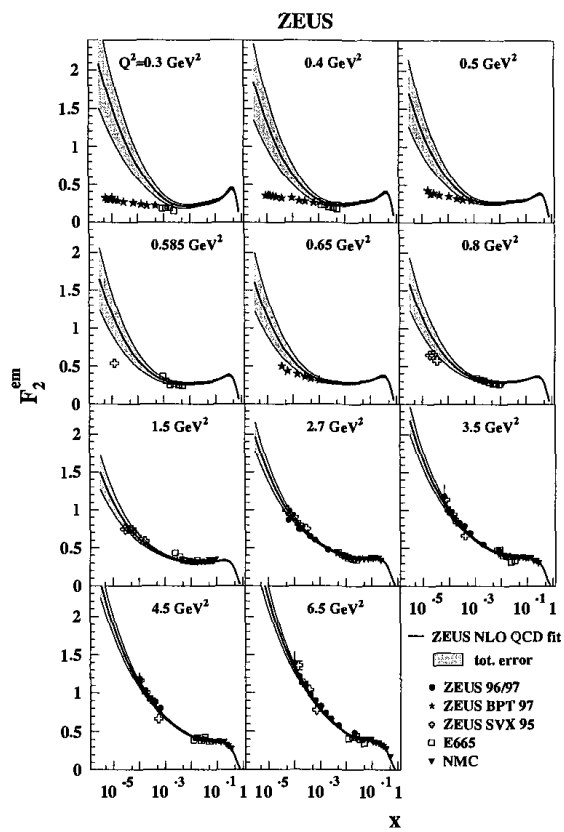


Figure 7. *The x dependence of F_2 as a function of Q^2 as depicted in the figure. The shaded area represents the results of QCD evolution fitted to higher Q^2 data, also calculated for low Q^2 where perturbative QCD is not expected to be valid.*

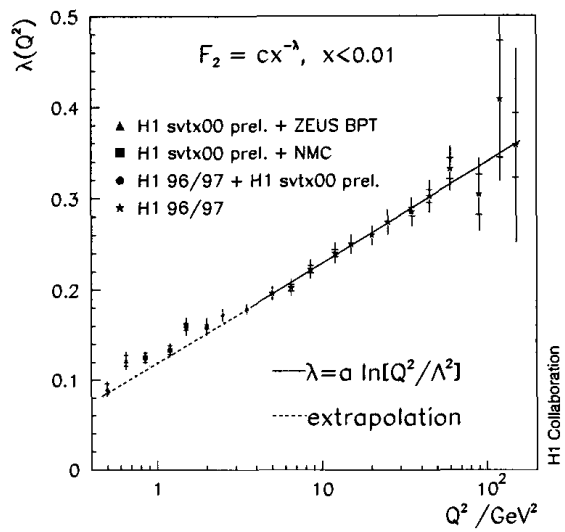


Figure 8. *The Q^2 dependence of the power of $1/x$ fitted to low x F_2 data.*

The controversy has been solved by the HERA data [21,22]. Fig. 7 shows that at low Q^2, the rise of F_2 with decreasing x is very mild, and becomes more pronounced as Q^2 increases. When the F_2 measurements are fitted at low x with a form $cx^{-\lambda}$, the value of λ is found to steadily increase with Q^2 as shown in Fig. 8 [22].

To conclude, the typical distribution expected for parton distributions is expected to be of the form

$$xf(x) = p_1 x^{p_2}(1-x)^{p_3} f'_{\text{smooth}}(x)\,, \tag{17}$$

where $f'_{\text{smooth}}(x)$ is some kind of polynomial function, which allows to smoothly interpolate between the low and large x regions. The parameters p_i are obtained by fitting the solution of the DGLAP evolution equations to the data.

6. Flavor content of the proton

In this discussion, for the sake of simplicity, we will assume the the proton consists only of light quarks, d, u and s. The valence quarks are obtained by subtracting from the quarks q the anti-quarks $\bar{q}$ contribution,

$$q_V = q - \bar{q}\,. \tag{18}$$

The sea quarks, q_S are then

$$q_S = q - q_V = \bar{q} \tag{19}$$

Since the gluons are flavor blind, one might expect that

$$\bar{u} = u_S = \bar{d} = d_S = \bar{s} = s\,, \tag{20}$$

where for the strange quark the subscript S has been dropped, as there are no strange valence quarks in the proton. The $SU(3)$ symmetry of the sea may be broken at low scales,

due to non-perturbative effects. For example, the strange quark is known to be heavier than the u and d quarks and its contribution may be suppressed. Non-perturbative effects, like intermediate bound states, may also affect the symmetry of the $\bar{u}$ and $\bar{d}$. Therefore relations 20 have to be checked experimentally.

6.1. Flavor information from structure functions

Assuming universality of parton distributions, it is possible to extract flavor information from structure functions measured in neutrino induced and charged lepton induced DIS on protons and neutrons. The relation between parton distribution in the proton and the neutron is given by the strong isospin symmetry,

$$\begin{array}{ll} u_n = d_p = d & \bar{u}_n = \bar{d}_p = \bar{d} \\ d_n = u_p = u & \bar{d}_n = \bar{u}_p = \bar{u} \\ \overset{(-)}{s}_n = \overset{(-)}{s}_p = \overset{(-)}{s} & . \end{array}$$

The following equations can easily be derived in the QPM:

$$F_2^{lp} = \frac{x}{9}\left[4(u+\bar{u}) + (d+\bar{d}) + (s+\bar{s})\right] \tag{21}$$

$$F_2^{ln} = \frac{x}{9}\left[4(d+\bar{d}) + (u+\bar{u}) + (s+\bar{s})\right] \tag{22}$$

$$F_2^{\nu N} = x\left[(u+\bar{u}) + (d+\bar{d}) + (s+\bar{s})\right] \tag{23}$$

$$F_3^{\nu} = \frac{1}{2}\left(F_3^{\nu N} + F_3^{\bar{\nu} N}\right) = \left[(u-\bar{u}) + (d-\bar{d}) + (s-\bar{s})\right] , \tag{24}$$

where lp (ln) stands for structure functions measured in charged lepton scattering on the proton (neutron) and νN ($\bar{\nu}N$) stands for neutrino (anti-neutrino) scattering on a nucleon, understood as the average of the proton and the neutron.

Assuming that $s = \bar{s}$, the following can be further derived:

$$F_3 = q_V \tag{25}$$

$$F_2^{lN} = \frac{5}{18}F_2^{\nu N} - \frac{6}{18}xs . \tag{26}$$

From equation (26), the strange sea may be derived. It is easy to see however, that the system remains under-constrained, unless one also assumes that $\bar{u} = \bar{d}$.

7. Constraints from DIS data

Further constraints can be derived from existing sum rules or from dedicated measurements, as discussed below.

7.1. Gross-Llewellyn Smith sum rule

One of the important constraints imposed on the valence quarks of the proton is that their total number should be equal to three. Since we can directly measure only F_3, this requirement is modified by QCD,

$$\int_0^1 F_3 dx = 3\left(1 - \frac{\alpha_S(Q^2)}{\pi} + \ldots\right) . \tag{27}$$

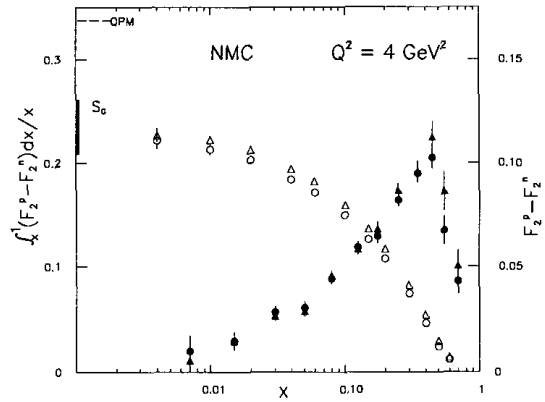


Figure 9. *The measurement of $F_2^{lp} - F_2^{ln}$ as a function of x and the corresponding integral I_{GS}.*

This relation, called the Gross-Llewellyn-Smith sum rule [23], has been verified experimentally [20], as shown in Fig. 6 and a good agreement with QCD expectations was found.

7.2. Gottfried sum rule

The Gottfried sum rule [24] postulates that

$$I_{GS} = \int_0^1 \left[F_2^{lp} - F_2^{ln}\right] \frac{dx}{x} = \frac{1}{3} + \frac{2}{3}\int_0^1 (\bar{u} - \bar{d})dx \overset{\bar{u}=\bar{d}}{=} \frac{1}{3}. \tag{28}$$

The sum rule has been tested by the NMC experiment [25,26], as shown in Fig. 9, and found to be violated. The value measured at $Q^2 = 4\,\mathrm{GeV}^2$, $I_{GS} = 0.235 \pm 0.026$, implies that in fact $\bar{d} > \bar{u}$.

7.3. F_2^{ln}/F_2^{lp}

The ratio F_2^{ln}/F_2^{lp} can be expressed as

$$\frac{F_2^{ln}}{F_2^{lp}} = \frac{4\frac{d+\bar{d}}{u+\bar{u}} + 1}{4 + \frac{d+\bar{d}}{u+\bar{u}}}. \tag{29}$$

The region of $x \to 0$ is dominated by sea quarks, and if the sea were to be flavor symmetric, $\bar{u} = \bar{d}$, the ratio $F_2^{ln}/F_2^{lp} = 1$. In the data [25,27] the ratio is significantly below one, as shown in Fig. 10. At large x, dominated by valence quarks, the ratio decreases with increasing x, suggesting that $d_V/u_V \to 0$ as $x \to 1$.

7.4. Strange sea

The suppression of strange sea relative to the light sea quarks has been measured directly in di-muon production in $\overset{(-)}{\nu}\, N$ interactions [28,29]. The wrong sign muon originates from the decay of a charm particle produced from the s or d sea in charged current interactions, as depicted in Fig. 11.

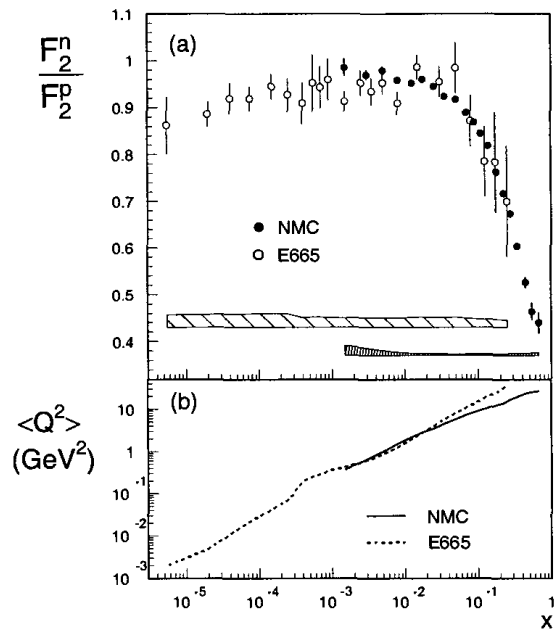


Figure 10. *The ratio F_2^{ln}/F_2^{lp} as a function of x. The bands correspond to systematic errors. The average Q^2 for each point in x is also plotted.*

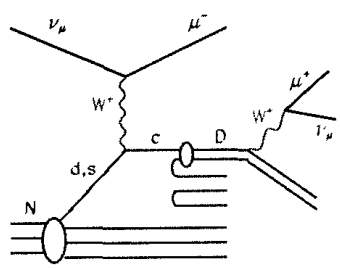


Figure 11. *The diagram for charm quark production from the nucleon sea in νN charged current interactions.*

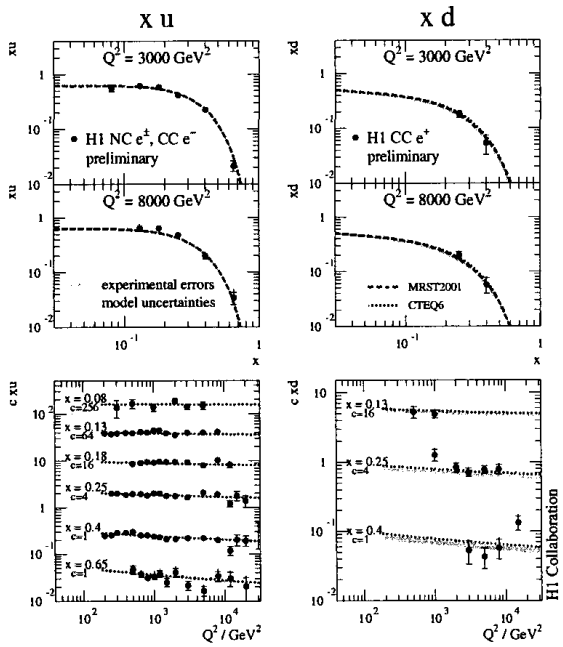


Figure 12. *The xu and xd distributions as a function of x and Q^2 extracted from charged current $e^{\pm}p$ interactions at HERA. The measurements are compared to existing parton parameterizations.*

The data can be interpreted in terms of

$$\kappa_{\overset{(-)}{\nu}} = \frac{\int 2x \overset{(-)}{s}\, dx}{\int x[\bar{u} + \bar{d}]dx} \,. \tag{30}$$

Experimentally, one finds a suppression of the strange sea relative to the light sea. The exact number depends on the model assumed to extract κ from the data [29], but typically

$$\kappa_{\nu} \quad < 0.45 \,, \tag{31}$$

and within uncertainties there is no asymmetry between s and $\bar{s}$. Moreover, the x dependence comes out to be the same as that of the light sea.

7.5. $e^{\pm}$ charged current interactions

In charged currents e^+p and e^-p collisions at large x, the cross section is dominated by the reactions $e^+d \to \bar{\nu}u$ and $e^-u \to \nu d$, and therefore the measured cross section is directly related to the d and u distributions. The results obtained by the H1 experiment [30] are shown in Fig. 12, and compared to expectations based on QCD fits.

8. Constraints from hadronic collisions

While it is more difficult to obtain reliable theoretical predictions for hard scattering processes in hadron-hadron collisions, some of the measurements turned out to be crucial

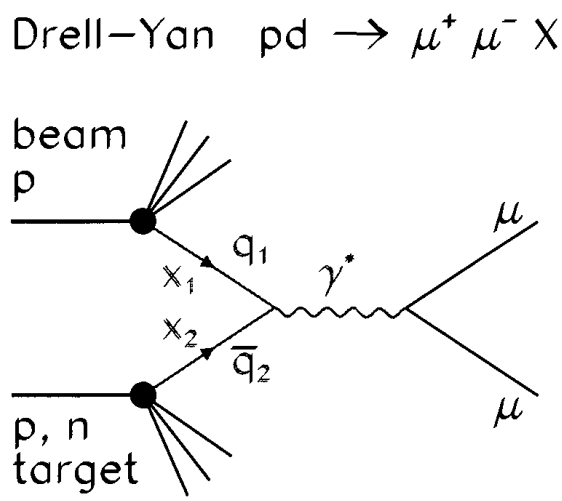


Figure 13. *Diagram for the production of a $\mu^+\mu^-$ pair as a result of annihilation of a quark and anti-quark.*

in constraining the shapes of the u and d flavors in the proton.

8.1. Measurement of the $\bar{d}/\bar{u}$ in Drell-Yan process

The cross section for the Drell-Yan production of two muons in pp or pd collisions (see Fig. 13) can be expressed in the following way:

$$\frac{d^2\sigma}{dx_1 dx_2} = \frac{4\pi\alpha^2}{9M^2}\sum_i e_i^2\left[q_i(x_1)\bar{q}_i(x_2) + \bar{q}_i(x_1)q_i(x_2)\right] , \tag{32}$$

where M^2 is the invariant mass of the two muons, and x_1 and x_2 are the fractions of the beam and target momenta respectively, as depicted in Fig. 13.

For a hard process $\sigma(pd) = \sigma(pn) + \sigma(pp)$ and for $x_1 \gg x_2$ the cross section formulae can be simplified,

$$\sigma^{pp} \propto \frac{4}{9}u(x_1)\bar{u}(x_2) + \frac{1}{9}d(x_1)\bar{d}(x_2) \tag{33}$$

$$\sigma^{pn} \propto \frac{4}{9}u(x_1)\bar{d}(x_2) + \frac{1}{9}d(x_1)\bar{u}(x_2), \tag{34}$$

and, since we know by now that at large x, the contribution of d is negligible compared to u, it is easy to show that

$$\frac{\sigma_{pd}}{2\sigma_{pp}}\,|_{x_1 \gg x_2} \simeq \frac{1}{2}\left(1 + \frac{\bar{d}(x_2)}{\bar{u}(x_2)}\right) . \tag{35}$$

Thus the ratio $\sigma_{pd}/2\sigma_{pp}$ is directly proportional to $\bar{d}/\bar{u}$. The latter is shown in Fig. 14 as a function of x [31]. In the figure, one can also appreciate the impact of these measurements on the parameterizations of parton distributions [32–36].

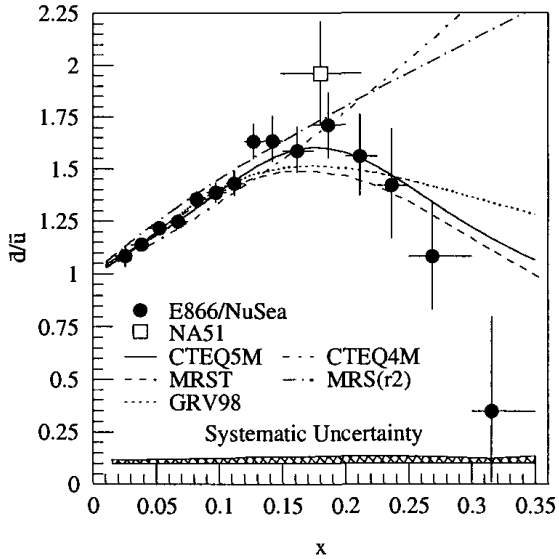


Figure 14. *The ratio $\bar{d}/\bar{u}$ as a function of x, as measured in Drell-Yan production of di-muons in pp and pd collisions. Also shown are predictions of various parameterizations of parton distributions.*

8.2. Charge asymmetry in $p\bar{p} \to W^{\pm}$ production

The charge asymmetry in $W^{\pm}$ production in $p\bar{p}$ interactions is sensitive to the differences in the d and u quark distributions, as can be inferred from Fig. 15.

Experimentally, one observes the leptonic decays of the W, and it is the charged lepton asymmetry that is being measured [37]. The charge asymmetry, A_W, is defined as

$$A_W(y) = \frac{\frac{d\sigma^+}{dy} - \frac{d\sigma^-}{dy}}{\frac{d\sigma^+}{dy} + \frac{d\sigma^-}{dy}}, \tag{36}$$

where y is the rapidity of the decay lepton and the sign stands for its charge. The measured leptonic asymmetry is shown in Fig. 16 and compared to expectations based on various parton distributions [34,35,40]. The impact of these data on the parton distributions is

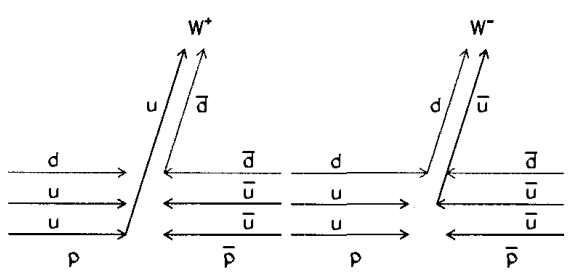


Figure 15. *Schematic representation of the production of $W^p m$ in $p\bar{p}$ collisions.*

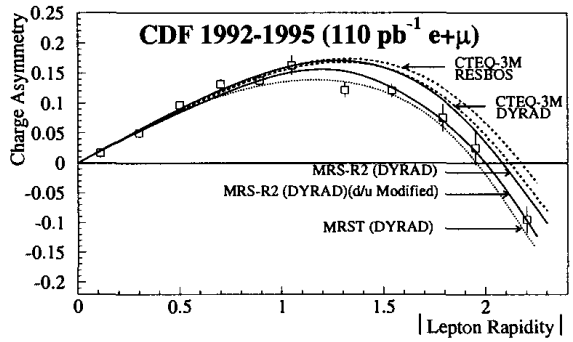


Figure 16. *The lepton charge asymmetry measured in the production and leptonic decay of $W^{\pm}$ in $p\bar{p}$ interactions as a function of the lepton rapidity. The lines correspond to expectations based on various parton parameterizations.*

readily visible.

9. Summary of flavor decomposition

The measurements of F_2, in charged lepton induced DIS, and xF_3, in neutrino induced DIS, constrain the u and d, sea and valence quarks. In addition, the dimuon production in pp and pd constrain the $\bar{d}/\bar{u}$ ratio, while charge asymmetry in $W^{\pm}$ production in $p\bar{p}$ interactions constrains the d/u ratio. The suppression of the strange quarks relative to the light sea is measured in charm neutrino-production. The quarks and anti-quarks carry about 50% of the proton momentum. The rest is carried by gluons.

10. Gluons

The remaining partons that need to be unraveled are the gluons. There are no interactions from which the gluon distribution can be determined directly. They can be extracted from the scaling violation of the F_2 structure function measured in DIS, and from hadron-hadron interactions in which the gluons participate in hard scattering.

The only constraint on the gluon distribution comes from the momentum sum rule, which requires that

$$\sum_f \int_0^1 x(q+\bar{q})(x,Q^2)dx + \int_0^1 xg(x,Q^2)dx = 1\,. \tag{37}$$

Moreover, asymptotically, as Q^2 goes to infinity, QCD predicts that

$$\sum_f \int_0^1 x(q+\bar{q})(x,Q^2)dx \;\rightarrow\; \frac{9}{25}\,, \tag{38}$$

$$\int_0^1 xg(x,Q^2)dx \;\rightarrow\; \frac{16}{25}\,. \tag{39}$$

Therefore, at high Q^2 the gluons are expected to carry more than 60% of the proton momentum.

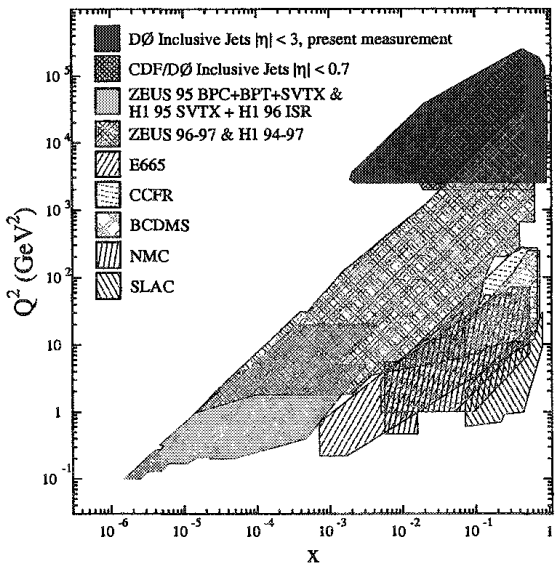


Figure 17. *The x and Q^2 kinematic plane covered by existing DIS measurements. Also shown is the area covered by measurements of high transverse momentum jets in $p\bar{p}$ collisions by the CDF and D0 experiments.*

10.1. Scaling violation

The evolution equation for F_2 (see eq. (9)) can be written as

$$\frac{\partial F_2(x,Q^2)}{\partial \ln Q^2} = \frac{\alpha_S(Q^2)}{2\pi} P_{qq} \otimes F_2(x,Q^2) + D P_{qg} \otimes xg(x,Q^2)\,, \tag{40}$$

where $D = \sum_i e_i^2$ for charged lepton DIS, and $D = 2N_f$ for neutrino DIS. The logarithmic derivative can be measured from the data and the convolution $P_{qq} \otimes F_2(x,Q^2)$ can be calculated, also using the measurements of F_2. This gives access to $P_{qg} \otimes xg(x,Q^2)$ from which xg can be extracted. In practice, this approach is difficult as F_2 is known only in a restricted range of x for a fixed Q^2. This can be seen in Fig. 17, where the kinematic plane of the existing measurements is shown [38].

The usual approach is to postulate a parametric form for the x dependence of F_2 and xg, at a predefined Q_0^2, and to predict the expected values of $F_2(x,Q_2)$ in the kinematic plane of interest by solving the DGLAP evolution equations. The parametric forms are then adjusted as to best reproduce the measurements. In reality, F_2 is reconstructed from assumed quark distributions, which are required to fulfill the flavor constraints discussed earlier. The typical number of free parameters used in this procedure is large, but there is sufficient information to determine all of them. This procedure is referred to as global QCD fit.

As an example, the results for the gluon distribution obtained by the ZEUS Collaboration [21] are shown in Fig. 18. Also shown are the uncertainties derived from the global fit.

At low Q^2 and low x the relative uncertainty is still large. This is due to the singular

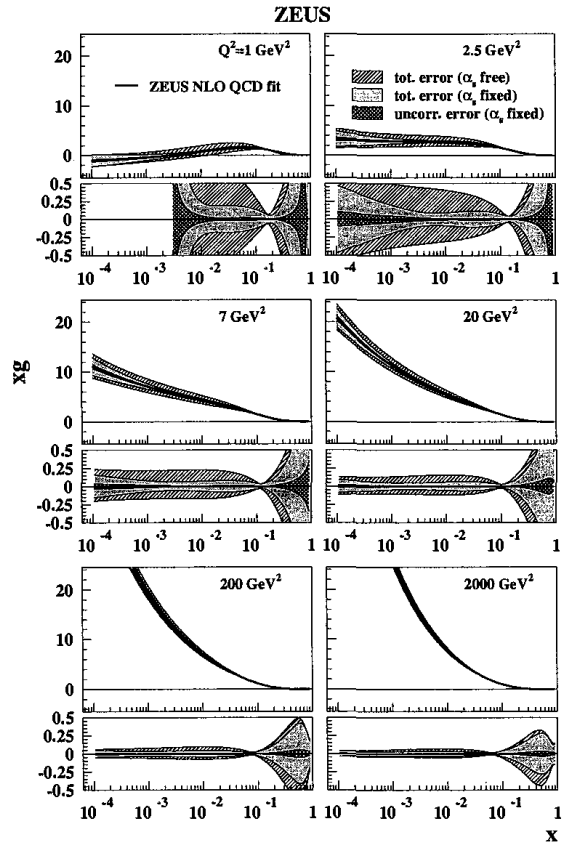


Figure 18. *The gluon distribution, xg as a function of x as obtained from a global fit to the ZEUS and fixed target data (see [21] for details). Also shown are the relative uncertainties derived from the fits.*

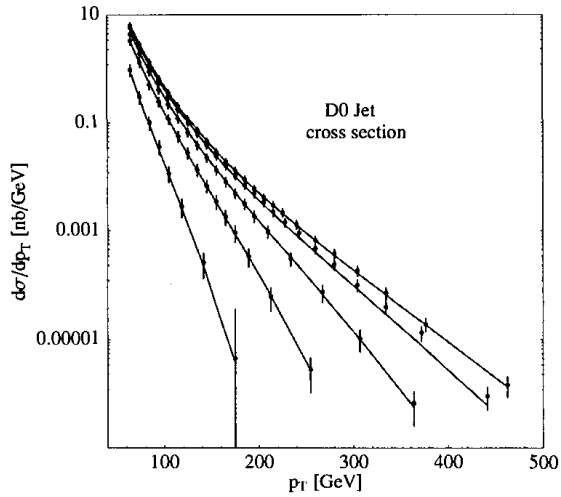


Figure 19. *Jet production cross section in $p\bar{p}$ interactions as a function of transverse momentum p_T and for different production angles compared to QCD expectations.*

nature of $P_{qg} \propto 1/x$, which amplifies the experimental errors on F_2. The small error around $x \simeq 0.1$ is due to the momentum sum rule. At low x, the precision with which the gluon distribution can be determined improves substantially. At high x, it remains poor throughout the Q^2 range, as the scaling violation at large x is dominated by quark contributions.

10.2. Hard hadron-hadron scattering

The precision on xg at high x may be improved by including in the fits measurements of high transverse momentum, p_T, jets in $p\bar{p}$ collisions. The jets originate from a parton-parton scattering and the cross section can be calculated in perturbative QCD. The hard scale is related to p_T. A comparison of QCD expectations and jet cross section measurements, as a function of p_T, for different production angles by the D0 experiment [38], is shown in Fig. 19. An impressive agreement, over close to five orders of magnitude, is observed.

The jets are produced in quark-(anti)quark, gluon-gluon or gluon-(anti)quark scattering. The relative fraction of cross section, produced by these three channels, is shown, as a function of p_T, in Fig. 20 for centrally produced jets [32]. The larger the p_T, the larger the x value of interacting partons.

The highest p_T jets originate predominantly from quark-anti-quark scattering, which is well constrained by fits to the DIS data. Therefore the data remain sensitive to gluons. In fact, the measurements performed by the CDF experiment [39] require more gluons at high x than obtained from fits to DIS data. An example of the modification [43] required to accommodate the CDF results is shown in Fig. 21.

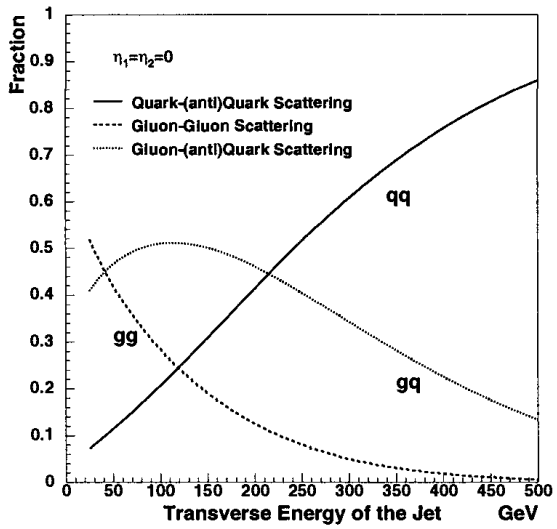


Figure 20. *Fraction of quark-(anti)quark, gluon-gluon or gluon-(anti)quark scattering in central jet production as a function of* p_T.

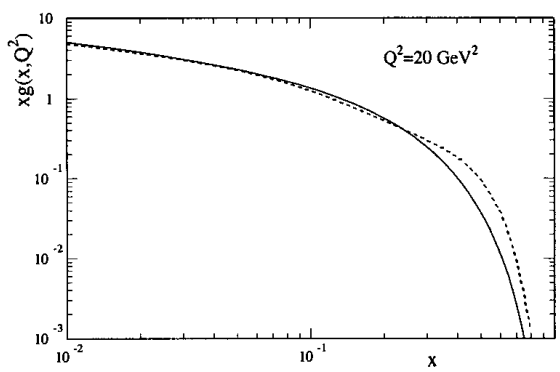


Figure 21. *Comparison between the gluon distribution obtained from QCD fits to DIS data (full line) and the one necessary to describe the production of high* p_T *jets in* $p\bar{p}$ *collisions (dashed line), as obtained by the MRST group [43] and shown at* $Q^2 = 20\,\mathrm{GeV}$.

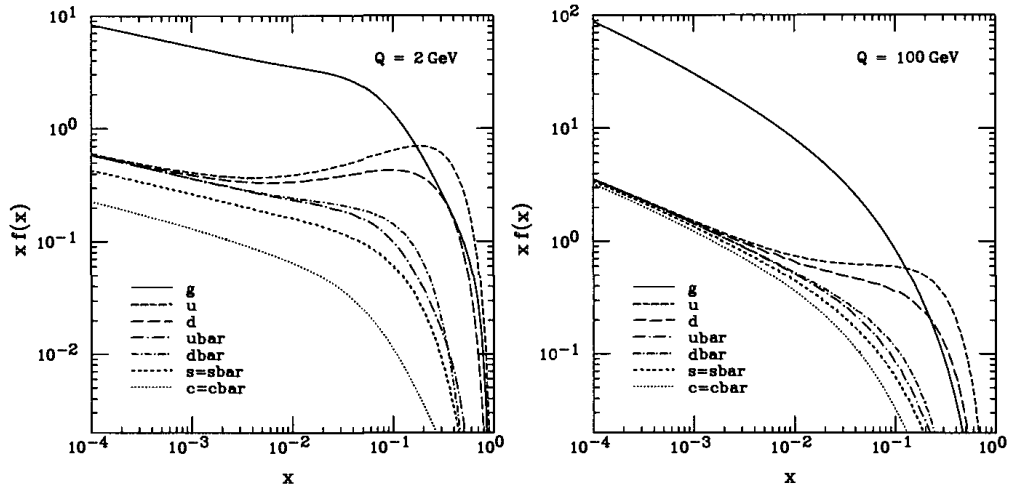


Figure 22. Parton distributions in the proton at two Q^2 scales as obtained in the CTEQ6 parameterization.

Table 1
Fractions of the proton momentum carried by partons as denoted in the table.

Q^2 (GeV^2)	u_v	d_v	$2\bar{u}$	$2\bar{d}$	$2\bar{s}$	$2\bar{c}$	$2\bar{b}$	g
2	0.310	0.129	0.058	0.075	0.037	0.001	0.000	0.388
20	0.249	0.103	0.063	0.077	0.046	0.017	0.000	0.439
$2 \cdot 10^4$	0.178	0.074	0.070	0.080	0.058	0.036	0.026	0.472

11. Proton structure

There are two groups that specialize in global QCD fits to the data, CTEQ [32,33,40,41] and MRST [34,35,42–44]. Out of these analyses, full parton distributions in the proton are available, including heavy flavors, which are derived from the evolution of gluons. Since there are experimental and theoretical uncertainties, there are usually many sets available, which account for the need of making different assumptions. Also the fits are regularly updated, usually when new precise data become available. The sets are available from the respective web pages of the two groups and in the CERN computer library [45].

The proton structure, as seen through the latest parameterizations of the CTEQ group, CTEQ6 [41], is shown in Fig. 22. The parton distributions are plotted as a function of x for two values of Q^2, at a low scale of $4\,\mathrm{GeV}^2$ and at a higher scale of $10^4\,\mathrm{GeV}^2$.

Typical fractions of the proton momentum carried by the various quark flavors and gluons are summarized in table 1.

The best way to assess the precision with which the proton has been mapped into partons, is by looking at the uncertainties on the calculation of the cross sections for

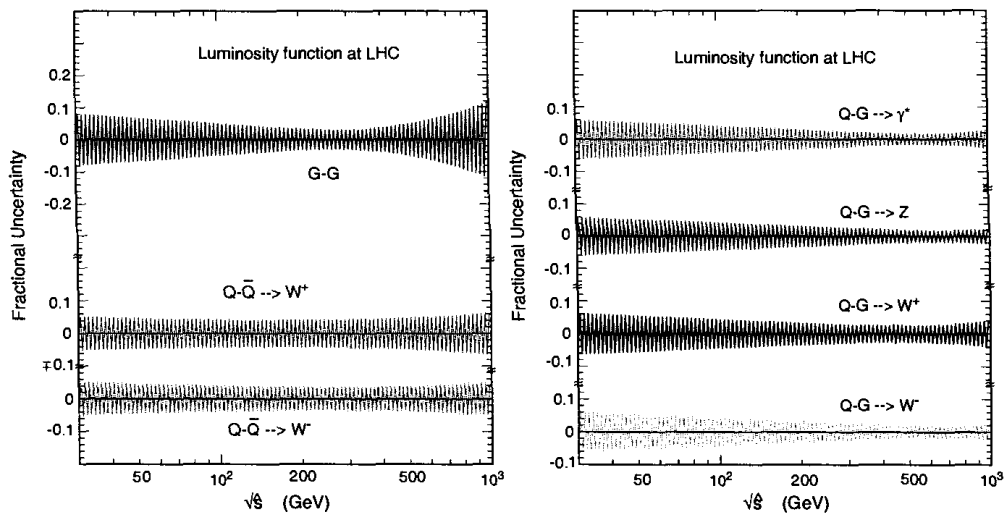


Figure 23. Fractional uncertainties on contributions from various parton-parton scatterings as a function of energy.

hard processes in the Large Hadron Collider kinematic range, where 7 TeV protons will collide in the near future. The fractional uncertainties on individual hard subprocesses, estimated by CTEQ [41], are shown in Fig. 23.

Typically, the uncertainties are below 10%. As expected the largest uncertainty comes from scatterings involving gluons. The latter is partly due to limitations in the theory. At the highest energies of the subprocesses, which correspond to large x, the largest uncertainty comes from the knowledge of the strong coupling constant, α_S.

12. Future

One of the uncertainties on the parton distributions, especially at low x, comes from the fact that theory does not predict at which Q^2 it is safe to use the NLO DGLAP formalism for the evolution. Numerically, the global fits seem to describe the data down to Q^2 as low as $\sim 1\,\mathrm{GeV}^2$, which is comparable to the size of the proton, as if the proton was a loosely bound state of partons. However, a closer look at the solutions obtained from the fit, shows that at these scales the gluons tend to be negative. While it is widely disputed whether the gluon distribution, which is not a physical observable, may turn negative, it is clear that F_L, which represents a cross section, has to be positive. The F_L expected from the gluons, shown in Fig. 18 [21], is presented in Fig. 24 as a function of x, for different values of Q^2. At $Q^2 = 1\,\mathrm{GeV}^2$ the F_L becomes negative for $x < .01$, which obviously is not a physical solution.

Another indication that the starting scale for safe NLO-DGLAP evolution should be closer to $10\,\mathrm{GeV}^2$ than to $1\,\mathrm{GeV}^2$, comes from a comparison between a prediction of the QCD global fits, obtained from MRST99 [42] and a calculation based on a dynamical model of low x interactions, the so-called colour-dipole model [46]. This is shown in Fig. 25.

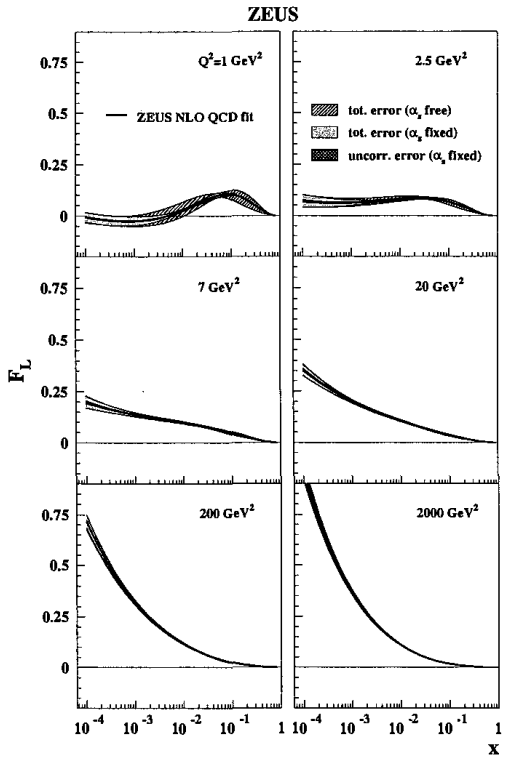


Figure 24. *The F_L structure function as derived from gluons presented in Fig. 18.*

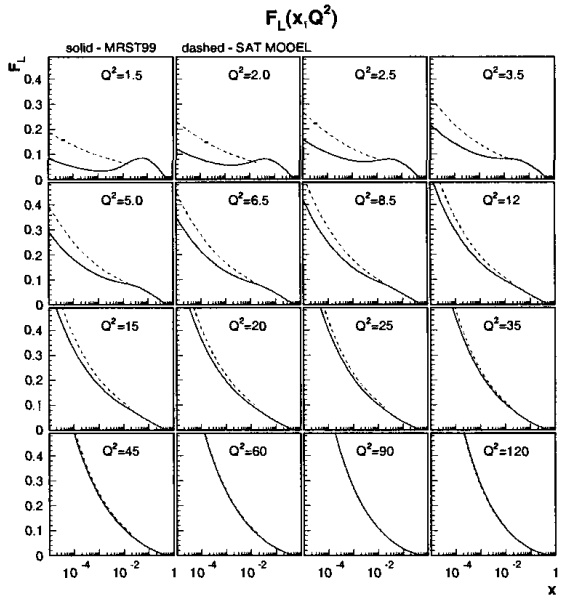


Figure 25. *The F_L structure function as derived from the MRST99 parameterization of parton distributions (full line) compared to the expectations of a colour dipole model (dashed line) [47].*

The colour-dipole model views the deep inelastic scattering as a three step process: first the photon is emitted, then it fluctuates into a $q\bar{q}$ pair and then the pair, which forms a colour-dipole, interacts with the proton. This approach is in fact a representation of DIS at low x in the proton rest frame. The dipole-proton cross section depends on the transverse size of the dipole. The larger the size, the larger the cross section, till the dipole becomes of a hadronic size, in which case the cross section saturates. The size of the dipole is given by the known wave function of the photon, while the dipole cross section is subject to modelling. The dipole model is very successful in describing many low x phenomena in DIS [48], therefore one may consider seriously its prediction for F_L.

The comparison shown in Fig 25 indicates that the two approaches, in the framework of DGLAP and of the colour-dipole model, come close only when $Q^2 > 10\,\mathrm{GeV}^2$. Since the dipole approach was specifically developed to take into account the many interesting effects observed in the low x region [49], one may wonder whether parton distributions at low x can be trustworthy. The answer to this question has to await dedicated measurements of F_L.

The discussion of low x phenomena is beyond the scope of this lecture. It is worthwhile to note however, that the studies of ep scattering at the HERA collider, where 820 GeV protons collide with 27.5 GeV positrons/electrons, have uncovered a new class of exclusive hard processes, like the exclusive production of vector mesons, V, or real photons γ, $ep \to epV(\gamma)$. The generic QCD diagram for these processes is shown in Fig. 26. At sufficiently high Q^2, or when the mass of the $q(\bar{q})$ is large, the interaction proceeds through an exchange of two-gluons. Factorization has been proven for these processes [50], and

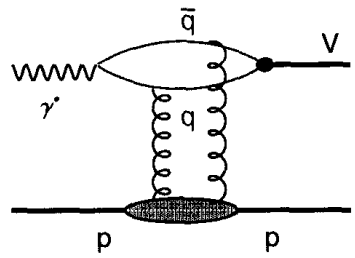


Figure 26. *Generic diagram for exclusive vector meson production, V, or real photon, γ in DIS through two-gluon exchange.*

therefore the gluon distribution should be that of the proton.

The W dependence of the cross section for exclusive V production [51] is shown in Fig. 27, when the virtuality of the exchanged photon is close to zero and the mass of the V is varied, and in Fig 28 for ρ^0 production, for different values of Q^2. While for light vector mesons, such as ρ^0, ω and ϕ, the W dependence for $Q^2 \simeq 0\,\mathrm{GeV}^2$ is mild, it becomes much stronger when charm quarks are involved. This would indicate that for large enough quark masses, the reaction becomes sensitive to the rise of gluons with decreasing x (equivalent with increasing W). Similarly, for ρ^0 production, at low Q^2, a mild W dependence is observed, while it becomes stronger when Q^2 increases, again hinting that gluons in the proton are involved.

These new class of exclusive hard processes has opened a whole new way of studying the dynamical structure of the proton [52].

Acknowledgements

It has been a pleasure and a priviledge to participate in the Erice Summer School. It is as interesting and enjoyable as I remember it from the time I was myself a student at the School. I would like to express my gratitude to Prof. Zichichi for giving me this opportunity. Many thanks go to my colleagues, experimentalists and theorists, from the HERA community which have kept the investigation of the proton structure alive and as fascinating as ever. I would also like to thank Aharon Levy for his critical help in preparing this writeup.

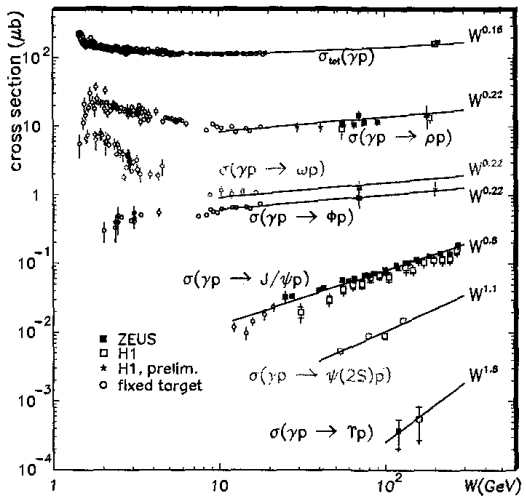


Figure 27. *W dependence of exclusive vector meson production for* $Q^2 \simeq 0\,\mathrm{GeV}^2$ *for different vector mesons, as denoted in the figure. The lines corresponding to* W^δ*, with the value of* δ *marked in the figure are shown to guide the eye.*

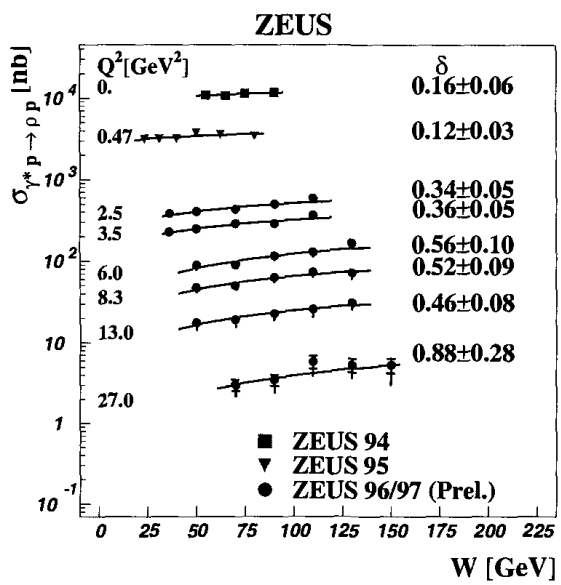


Figure 28. *W dependence of exclusive* ρ^0 *production for diferent* Q^2 *values as denoted in the figure. The lines are the result of fits of* W^δ *dependence to the measurements.*

REFERENCES

1. P. Hoyer, Nucl. Phys. A **711** (2002) 3 [arXiv:hep-ph/0208181].
2. J. C. Collins, D. E. Soper and G. Sterman, Nucl. Phys. B **261** (1985) 104.
3. E. D. Bloom *et al.*, Phys. Rev. Lett. **23** (1969) 930.
4. M. Breidenbach *et al.*, Phys. Rev. Lett. **23** (1969) 935.
5. C. G. Callan and D. J. Gross, Phys. Rev. Lett. **22** (1969) 156.
6. V. N. Gribov and L. N. Lipatov, Yad. Fiz. **15** (1972) 1218 [Sov. J. Nucl. Phys. **15** (1972) 675].
7. Y. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641 [Zh. Eksp. Teor. Fiz. **73** (1977) 1216].
8. G. Altarelli and G. Parisi, Nucl. Phys. B **126** (1977) 298.
9. A. C. Benvenuti *et al.* [BCDMS Collaboration], Phys. Lett. B **223** (1989) 485.
10. M. Arneodo *et al.* [New Muon Collaboration], Nucl. Phys. B **483** (1997) 3 [arXiv:hep-ph/9610231].
11. C. Adloff *et al.* [H1 Collaboration], Eur. Phys. J. C **21** (2001) 33 [arXiv:hep-ex/0012053].
12. A. C. Benvenuti *et al.* [BCDMS Collaboration], Phys. Lett. B **223** (1989) 485.
13. L. W. Whitlow, S. Rock, A. Bodek, E. M. Riordan and S. Dasu, Phys. Lett. B **250** (1990) 193.
14. E. Rondio [NMC Collaboration], Nucl. Phys. Proc. Suppl. **54A** (1997) 139.
15. R. Blankenbecler and S. J. Brodsky, Phys. Rev. D **10** (1974) 2973.
16. G. R. Farrar and D. R. Jackson, Phys. Rev. Lett. **35** (1975) 1416.
17. P. D. B. Collins, "An Introduction to Regge Theory and High Energy Physics", Cambridge University Press England, 1977.
18. A. Donnachie and P. V. Landshoff, Phys. Lett. B **296** (1992) 227 [arXiv:hep-ph/9209205].
19. J. R. Cudell, K. Kang and S. K. Kim, arXiv:hep-ph/9712235.
20. D. A. Harris *et al.* [CCFR-NUTEV Collaboration], arXiv:hep-ex/9506010.
21. S. Chekanov *et al.* [ZEUS Collaboration], Phys. Rev. D **67** (2003) 012007 [arXiv:hep-ex/0208023].
22. J. Gayler [H1 Collaboration], Acta Phys. Polon. B **33** (2002) 2841 [arXiv:hep-ex/0206062].
23. D. J. Gross and C. H. Llewellyn Smith, Nucl. Phys. B **14** (1969) 337.
24. K. Gottfried, Phys. Rev. Lett. **18** (1967) 1174.
25. P. Amaudruz *et al.* [New Muon Collaboration], Phys. Rev. Lett. **66** (1991) 2712.
26. M. Arneodo *et al.* [New Muon Collaboration], Phys. Rev. D **50** (1994) 1.
27. M. R. Adams *et al.* [E665 Collaboration], Phys. Rev. Lett. **75** (1995) 1466.
28. A. O. Bazarko *et al.* [CCFR Collaboration], Z. Phys. C **65** (1995) 189 [arXiv:hep-ex/9406007].
29. M. Goncharov *et al.* [NuTeV Collaboration], Phys. Rev. D **64** (2001) 112006 [arXiv:hep-ex/0102049].
30. [H1 Collaboration], "Measurement and QCD Analysis of Inclusive Deep Inelastic Scattering at high Q^2 and large x", abstract 978 submitted to the 31st International Conference on High Energy Physics, ICHEP02, Amsterdam, 2002.

31. R. S. Towell *et al.* [FNAL E866/NuSea Collaboration], Phys. Rev. D **64** (2001) 052002 [arXiv:hep-ex/0103030].
32. H. L. Lai *et al.*, Phys. Rev. D **55** (1997) 1280 [arXiv:hep-ph/9606399].
33. H. L. Lai *et al.* [CTEQ Collaboration], Eur. Phys. J. C **12** (2000) 375 [arXiv:hep-ph/9903282].
34. A. D. Martin, W. J. Stirling and R. G. Roberts, Phys. Lett. B **354** (1995) 155 [arXiv:hep-ph/9502336].
35. A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C **4** (1998) 463 [arXiv:hep-ph/9803445].
36. M. Gluck, E. Reya and A. Vogt, Eur. Phys. J. C **5** (1998) 461 [arXiv:hep-ph/9806404].
37. F. Abe *et al.* [CDF Collaboration], Phys. Rev. Lett. **81** (1998) 5754 [arXiv:hep-ex/9809001].
38. B. Abbott *et al.* [D0 Collaboration], Phys. Rev. Lett. **86** (2001) 1707 [arXiv:hep-ex/0011036].
39. T. Affolder *et al.* [CDF Collaboration], Phys. Rev. D **64** (2001) 032001 [Erratum-ibid. D **65** (2002) 039903] [arXiv:hep-ph/0102074].
40. H. L. Lai *et al.*, Phys. Rev. D **51** (1995) 4763 [arXiv:hep-ph/9410404].
41. J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky and W. K. Tung, JHEP **0207** (2002) 012 [arXiv:hep-ph/0201195].
42. A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C **14** (2000) 133 [arXiv:hep-ph/9907231].
43. A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C **23** (2002) 73 [arXiv:hep-ph/0110215].
44. A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, arXiv:hep-ph/0211080.
45. H. Plothow-Besch, "Proton, Pion and Photon Parton Density Functions, Parton Density Functions of the Nucleus and α_S Calculations", User's Manual, W5051.
46. K. Golec-Biernat and M. Wusthoff, Phys. Rev. D **59** (1999) 014017 [arXiv:hep-ph/9807513].
47. K. Golec-Biernat, private communication.
48. M. F. McDermott, arXiv:hep-ph/0008260, and references therein.
49. For a review see H. Abramowicz and A. Caldwell, Rev. Mod. Phys. **71** (1999) 1275 [arXiv:hep-ex/9903037].
50. J. C. Collins, L. Frankfurt and M. Strikman, Phys. Rev. D **56** (1997) 2982 [arXiv:hep-ph/9611433].
51. For a discussion and references see A. Levy [ZEUS Collaboration], Acta Phys. Polon. B **33** (2002) 3547 [arXiv:hep-ex/0206048].
52. For a discussion and references see M. Diehl, Eur. Phys. J. C **25** (2002) 223 [arXiv:hep-ph/0205208].

CHAIRWOMAN: H. ABRAMOWICZ

Scientific Secretaries: T. Lastovicka, R. Stamen

DISCUSSION

- *Chakrabarti:*

In experiments, how do people separate the contribution of the valence quarks and the sea quarks?

- *Abramowicz:*

In the proton case, the valence quarks are defined as the difference between quarks and anti-quarks. This is a "statistical" definition. The valence quarks carry the quantum numbers of the proton. The only experiments in which one can probe the valence quarks and the sea quarks are neutrino and anti-neutrino charged current scattering experiments. As shown in the lecture, the xF_3 structure function averaged over the proton and neutron is proportional to the valence quarks in the proton. The valence quarks are dominating at high x, therefore measurements performed at high x can be thought of as being sensitive to valence quarks. On the other hand, the probability density function of valence quarks tends to zero when x goes to zero. The sea quarks, which mostly originate from gluon splitting, are expected to populate the low x region. Here again, the only experiments that are really sensitive to the sea quarks are (anti)neutrino experiments.

Also, charged current electron/positron-proton interactions are sensitive to valence quarks, but not as directly as charged current neutrino scattering on an isoscalar target. In addition, since at large x we expect only valence quarks, any structure function measurement at large x is sensitive to valence quarks.

- *Ramtohul:*

Could you explain the differences between the MRST and the CTEQ parton distributions?

- *Abramowicz:*

There are a few reasons why they are different: first of all, the groups do not assume exactly the same shapes as the input distributions. This leads to slight differences. Secondly, they do not always use the same data sets. This is especially true for data sensitive to gluon distributions. Thirdly, they use different schemes to treat threshold effects. And finally, there are different procedures to handle systematic errors. The most recent parton distributions differ mainly in the gluon density. The quark distributions are pretty well constrained by 'direct' data – so there is not much space for substantial differences.

- *Bechtle:*

What is the experimental set-up (machine and detector) conceived for the F_L measurement at HERA 3?

- *Abramowicz:*

At this stage we are discussing different options. The dream design would be the one optimized for measurements of F_L. The principle of measuring F_L has been worked out by the SLAC experiment some 20 years ago. The idea is to measure the differential cross-section for fixed x and Q^2 as a function of y. Since $y = Q^2/(xs)$, we have to change the beam energy. For reasons of systematics, it is most advantageous to change the energy of the proton beam. Then the final state electron, which determines fully the kinematics of the event, is at a similar detector location for a given x and Q^2.

To measure F_L with reasonable precision, we need a large lever-arm in y. This also means that we need to measure electrons down to low momenta. This is the region where it may be easy to misidentify a charged particle or a photon from photoproduction background for the scattered electron. To prevent it, we need a good background rejection, i.e. a good identification of the photoproduction background. This may be possible, if we design an electron spectrometer (tracking + calorimetry) down to the lowest Q^2.

- *Lastovicka*

The use of different proton energies was already tried by the H1 experiment, because in 1997 HERA was running with 820 GeV protons while in 1999/2000 the proton energy was increased to 920 GeV. The problem is that these energies are so close to each other that one cannot get anything precise, the error was sort of 100%.

- *Abramowicz:*

As for the initial proton energies, they have to be optimized, taking into account the luminosity of the accelerator as a function of proton energy. At this stage, we can estimate what effect we are after and therefore to what precision we need to measure F_L. For example, there are two approaches, DGLAP-QCD and the saturation model, that predict quite different values for F_L (up to 30% difference) although they both describe the F_2 data very well. Possible deviation of the measurement from the QCD prediction would indicate where the DGLAP evolution equations break down.

- *Shuryak:*

The idea that $F_2(Q^2)$ measurements may not be providing a clear place where DGLAP validity stops, while those of F_L could, reminds me of the situation I discussed in my first lecture. Among many studied correlation function, $\langle 0|j(x)j(0)|0\rangle$, the one for vector currents is remarkably smooth, so that boundaries of pQCD cannot be seen. However, in many other functions, where j is scalar or pseudoscalar, this is very clearly seen. So, maybe F_L is really different from F_2, as we have seen similar things before.

- *Abramowicz:*

At this point it might be interesting to point out one clear hint that DGLAP formalism may be breaking down at low Q^2. There is an outstanding difference between the measurements of the neutrino-nucleon and charge lepton-nucleon structure functions, and this at relatively low x and low Q^2. This difference might be partly due to nuclear effects (in the case of neutrino measurements), heavy flavour threshold effects or implementation of radiative corrections. However, following the HERA example, the neutrino measurements have been extended to lower Q^2 and it is clear that, as expected, the limiting behaviour of the neutrino F_2 as Q^2 goes to zero is different from that of the electromagnetic F_2 (because of the axial current contribution in the neutrino case, absent in the electromagnetic case). So it might well be that the observed difference is in fact due to non-perturbative effects.

- *Levy:*

Why is it important to know where exactly DGLAP evolution equations stop working?

- *Abramowicz:*

As you might remember, one of the plots I have shown is that the property of the splitting function is such that we are accumulating a lot of partons at very low x. Eventually what might happen is that we will have so many of them that, due to overcrowding, we expect to see coherent effects. In particular, there are so many gluons that the gluons will start to interact with each other and this will invalidate the DGLAP regime. We are sort of moving towards strong interactions but hopefully in the region where α_s is still small. Then maybe this will allow theorists to test certain ideas of how to do calculations in the non-perturbative region when going from the perturbative region. The other issue is that, maybe, when we are sitting in this very dense environment, new matter effects will appear. It may turn out that some of the effects observed at HERA are relevant to heavy ion physics at RHIC. The question whether DGLAP evolution equation is valid in a particular region of x and Q^2 has two important implications. If one evolves the parton distributions toward higher Q^2, the question is whether the input was correct. And if one evolves toward lower Q^2, the question is whether we observe new effects, beyond the DGLAP evolution.

- *Dvergsnes:*

Could you comment on the last sentence in your presentation, about the photon structure and that there is still a long way to go?

- *Abramowicz:*

The photon is an extremely interesting particle. It is a gauge particle and therefore a very fundamental particle of nature. We have known for quite some time that in interactions with matter, the photon develops a hadronic structure and this structure can be probed, in a manner very similar to the probing of the nucleon structure. Originally, Witten calculated the asymptotic photon $F_2(x,Q^2)$ from first principles. That gave hope

that α_s could be determined very precisely and QCD evolution tested with a known input. Unfortunately, when the photon develops hadronic structure and, if it is not particularly virtual, it likes to turn into hadrons, for example a ρ meson. This starts to be an unknown input to the evolution equations. This is particularly important for the low x region. There are some measurements from LEP and also HERA. But to explore this region, one needs high centre of mass energies which are not yet available. These measurements will be possible at the next linear collider.

- *Maas:*

There are considerations to build an ep-collider of type HERA-TESLA or LHC-LEP. What would be the impact for proton structure and DIS physics?

- *Abramowicz:*

There is a whole book on this subject. In the context of the proton structure, and the low x physics in particular, the phase space would be extended by two orders of magnitude. It will constrain the evolution in Q^2 at low x. Also, by having access to lower x values than at HERA, we can expect to see parton saturation phenomena, that might be present already at HERA. The reason why this is expected is that at HERA we observed a very copious production of diffractive events, even at large Q^2. Diffraction, in which the proton remains unscathed after the interaction, can be thought of as elastic scattering of the virtual photon on the proton. From the optical theorem, we expect the elastic cross-section to increase with energy like the square of the total cross-section. If this were to be the case, eventually the elastic cross-section would become bigger than the total!!! The unitarity limit requires that the elastic cross-section be at most half of the total. So obviously something must happen. If this starts happening, in the region where α_s is still small, it might finally shed light on the approach to the unitarity limit at high energy.

- *Wendland:*

Can HERA be run with polarized protons?

- *Abramowicz:*

The answer is not clear. RHIC has installed the so-called Siberian snakes and it was successful in polarizing protons of 100 GeV at a level of 25%. The same technology could in principle be applied at HERA. The question is, of course, the cost. There is a proposal to build ERHIC, a polarized electron-proton machine at RHIC, with 10 GeV electrons colliding with 250 GeV protons. This might be a cheaper option. Although the centre of mass energy would be by about a factor 10 lower than at HERA, it might be sufficient for the polarized structure function programme.

- *Wendland:*

I was wondering more about HERA.

- *Abramowicz:*

But if there is ERHIC, there might be less motivation to do it at HERA.

- *White:*

There will be a polarized proton program at RHIC at $\sqrt{s}$ = 200 and 500 GeV and the major focus of that is to measure Δ g, to see what the missing component of the proton is. In the further distant future, there is a plan for an electron ion collider, which would also address some of these aspects.

- *Bechtle:*

What is the physics that is planned with polarized electrons and positrons?

- *Abramowicz:*

The most beautiful experiment to imagine is to have polarized right-handed electrons and to measure charge currents. The cross-section is zero. This is a typical example of a null experiment that we would like very much to do. Of course, the beam would not be 100% polarized, we would be lucky if we manage to have it polarized to 60%. Especially in the context of knowing now that the masses of neutrinos are not zero, I think that such a measurement is a must. For right-handed polarization of electrons and left-handed polarization of positrons, the charged current cross-section should be zero, for massless neutrinos. The presence of polarization also substantially increases sensitivity to the electroweak parameters. This has been seen at the SLC, where with substantially lower luminosity than LEP but polarized beams, $\sin^2\theta_W$ was measured with similar precision. The same applies to HERA, so, e.g., the measurement of the W mass will be competitive to the one of Fermilab Run II if only polarized lepton beams are available.

- *Bechtle:*

Is there a possibility to use it for testing transversely or longitudinally polarized photons?

- *Abramowicz:*

The fact that you have polarized leptons does not give you the polarization of the photon, so that does not really help.

- *Stamen:*

There are many different data sets which enter the global fits for parton densities. Which data set constrains the fits most and which data sets might be omitted?

- *Abramowicz:*

How should I answer? I do not think that there are actually data sets which are not necessary. In similar regions of phase space, obviously the more precise data will determine the fate of the fit. In the case of HERA data, in principle we try to use only our own data for the DGLAP fits, but in fact we use the knowledge of shapes from previous fits and, therefore, indirectly we rely on them as well. In my opinion, there are no useless data and rejecting particular sets is wrong, unless the culprit is theory.

Selected Highlights from the First Heavy Ion Runs at RHIC

Timothy J. Hallman
Brookhaven National Laboratory, Upton, NY 11973, USA[1]

For the first time, nucleus-nucleus collisions at RHIC occur at sufficiently high $\sqrt{s_{NN}}$ that the yield of jets from hard and semi-hard scattering of partons may be used as an effective probe of the hot, dense matter which is produced. This has led to several intriguing observations in $AuAu$ collisions at $\sqrt{s_{NN}} = 130$ and 200 GeV.

Before discussing high p_t measurements, it is useful to say a word about elliptic flow. Elliptic flow is a phenomena largely unfamiliar in high energy physics. It results from the finite size of the colliding nuclei, and from the fact that for non-central nucleus-nucleus collisions, there is a geometrical asymmetry in the region of overlap swept out when two colliding nuclei pass through each other. The interaction region in this case is elliptical, or "almond shaped". It has been recognized for some time [1-4], that if in the early moments of the collision the matter in this region is strongly interacting, (meaning there is a short mean free path for multiple interactions among the reaction products) this geometrical asymmetry may be converted into a momentum space anisotropy of the emitted particles.

In general, the transverse momentum distribution for particle emission is expected to show both directed flow as well as a possible elliptic flow component:

$$\frac{d^3N}{dp_t^2 d\phi dy} = \frac{dN}{2\pi dp_t^2 dy}[1 + 2\sum_n v_n \cos(n\phi)] \tag{1}$$

where the angle of emission ϕ of each particle is determined with respect to the reaction plane (the plane formed by the z axis and the impact parameter). P_t is the transverse momentum of a given particle and y is its rapidity. The v_n harmonic coefficients in this decomposition are termed anisotropy parameters; v_1 is the first harmonic Fourier coefficient indicative of the magnitude of the directed flow, and the elliptic flow, v_2, is the second harmonic coefficient of the Fourier decomposition of the azimuthal distribution of particles with respect to the reaction plane. Specifically,

$$v_2 = \langle \cos(2\phi) \rangle \tag{2}$$

where the azimuthal angle ϕ of each particle is determined with respect to the reaction plane, and an average is performed over an ensemble of events e.g. of a given centrality class.

Figure 1a shows the first measurement, as a function of centrality, of the elliptic flow at RHIC for $AuAu$ collisions at $\sqrt{s_{NN}} = 130$ GeV/n [5]. Figure 1b, shows the elliptic flow measured for identified particle spectra for momenta up to ~ 1 GeV/c. Of particular

[1]This research was supported by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886.

interest in both Figs. 1a and 1b is the fact that the magnitude of v_2 saturates the prediction of models predicated on hydrodynamics. From Fig. 1a it is seen that the size of v_2 decreases for more central collisions as expected, since the geometric asymmetry of the region swept out by the two nuclei is much smaller (i.e., the overlap region is nearly spherical) in this case. The expansion of the interaction volume which occurs after the initial impact tends to reduce the amount of elliptic flow. Observation of significant v_2 therefore provides *prima facie* evidence that the matter produced in the earliest stage of $AuAu$ collisions at RHIC is strongly interacting.

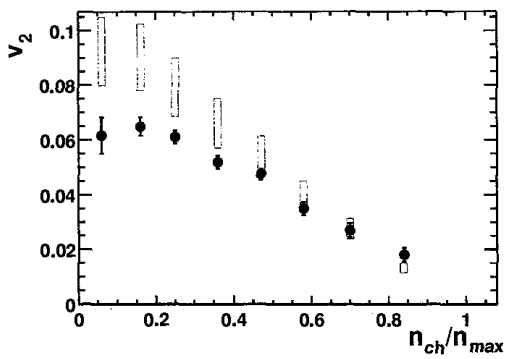


Fig. 1a. Elliptic flow (solid points) as a function of centrality ($n_{ch}/n_{max} = 1$ for the most central collisions). The open rectangles show a range of values expected for v_2 based on a hydrodynamic model; when the initial space eccentricity is multiplied by 0.19 (lower edge), and 0.25 (upper edge, close to the hydrodynamic limit).

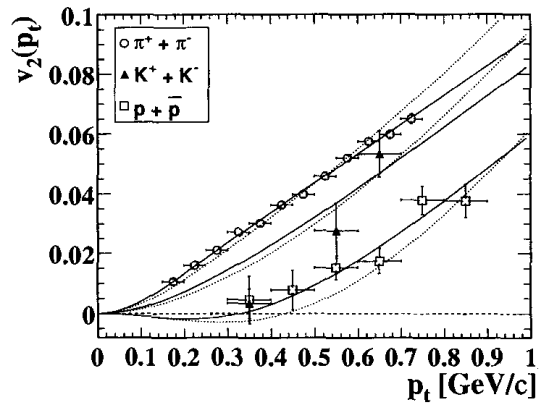


Fig. 1b. Differential elliptic flow for pions, kaons, and protons. The solid and dashed lines are fits to several hydrodynamic models. [6]

The extension of these measurements to high p_t (Fig. 2) shows a rather surprising feature. Up to approximately 2 GeV/c, the elliptic flow agrees surprisingly well with the prediction of models based on hydrodynamics. Above that momentum however, saturation occurs, with the magnitude of v_2 staying approximately constant out to moderately high p_t ($\sim$ 10 GeV/c). Work to identify the contribution to v_2 from non-flow effects such as jet production, resonances, final state interactions, etc., has shown [7] that although there are significant non-flow effects out to transverse momenta $\sim 4 - 5$ GeV/c, there continues to be sizeable elliptic flow at these momenta. (Above this transverse momenta the non-flow component is difficult to determine thus far due to limited statistics). This is a striking observation: even at transverse momenta sufficiently high that hard and semi-hard scattering play a role, the pattern of particle emission exhibits sizeable elliptic flow. It suggests that the matter produced in relativistic heavy ion collisions at RHIC is rather opaque, coupling the asymmetry in the geometry of the overlap region to an anisotropy in the azimuthal distribution of fragments from hard and semi-hard scatters.

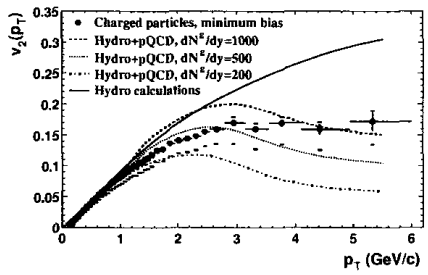


Fig. 2a. The elliptic flow (v_2) measured for minimum bias inclusive charged particles as a function of transverse momenta (data points). The solid line is the prediction based on hydrodynamics. The broken lines are predictions of hydrodynamics + pQCD assuming various gluon densities in the early stage of the collision. [8]

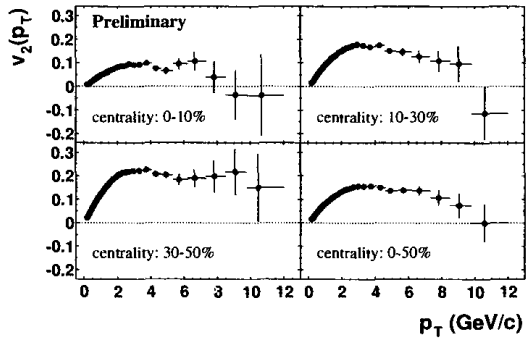


Fig. 2b. Elliptic flow as a function of p_t for inclusive charged particles from events in several centrality class.

The second striking feature related to high p_t particle production at RHIC concerns the inclusive yield of charged particles as a function of p_t. One expects *a priori*, for rare processes, that the cross section will scale in nucleus-nucleus collisions as the number of hard binary scatters. Using a monte carlo Glauber calculation [9] to estimate the number of binary scatters, it is expected that in principle, if $AuAu$ collisions were properly described as an incoherent superposition of individual nucleon-nucleon collisions, the ratio as a function of p_t of the inclusive yield in nucleus-nucleus collisions compared to that in $p+p$ collisions would be

$$R_{AA}(p_t) = \frac{d^2N^{AA}/dp_t d\eta}{T_{AA}\ d^2\sigma^{NN}/dp_t d\eta} \tag{3}$$

where $T_{AA} = \langle N_{binary}\rangle/\sigma^{NN}_{inelastic}$ accounts for the geometric scaling from elementary to nuclear collisions and $\langle N_{binary}\rangle$ is the number of binary collisions calculated using a Glauber model. R_{AA} is sometimes termed the "nuclear modification factor". In the absence of nuclear effects beyond a simple incoherent superposition of nucleon-nucleon scatterings, R_{AA} is expected to be 1. In reality, as observed in proton - nucleus measurements some time ago, in the region of approximately $2-3$ GeV/c, the inclusive particle yield can actually exceed that in $p+p$ collisions due to initial state multiple scattering, or the so-called Cronin effect [10]. This enhancement is seen rather clearly for inclusive π^0 production in $PbPb$ collisions at the SPS [11], as shown in Fig. 3.

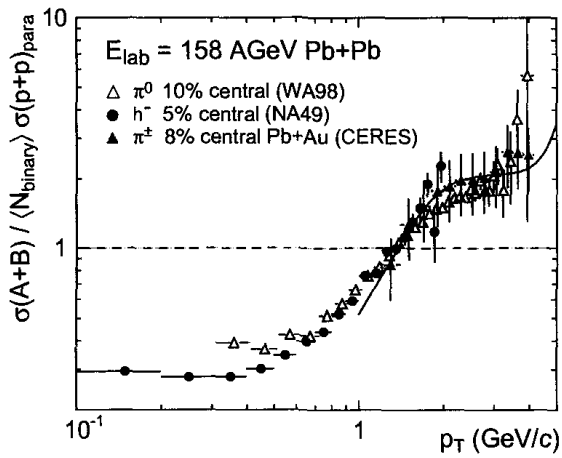


Fig. 3 R_{AA} plotted as a function of p_t for particle production in $PbPb$ collisions at the CERN SPS.

The behavior observed in central $AuAu$ collisions at RHIC is strikingly different. Comparing the inclusive charged hadron spectra observed in STAR and inclusive π^0 spectra from PHENIX (Fig. 4) in $AuAu$ collisions to pp comparison data in the same detectors using eq. 3, the resulting value of $R_{AA}(p_t)$ is plotted as function of p_t and centrality in Fig. 5 for $AuAu$ collisions at $\sqrt{s_{NN}} = 200$ GeV. For inclusive charged hadron spectra in Fig. 5 (left panel), for example it is observed for the most peripheral event class (60-80%, lower right panel) that the ratio $R_{AA}(p_t)$ rises, approaching the binary scaling limit for momenta above approximately 2 GeV/c. For the most central collisions (0-5%, upper left panel) $R_{AA}(p_t)$ shows something strikingly different. There is an initial rise as for peripheral collisions, but at $\sim$ 2 GeV/c the R_{AA} turns over, monotonically decreasing to $\sim$ 6 GeV/c after which the suppression saturates and remains approximately constant out to $\sim$ 10 GeV/c. This is qualitatively new behavior observed for the first time at RHIC, suggesting that the matter produced in central $AuAu$ collisions is strongly interacting, perhaps sufficiently so that it is opaque to fragmentation products from hard and semi-hard parton scattering. Theoretical work surrounding this observation [12] suggests that the detailed behavior may in fact reflect an interplay, over the range of p_t studied, of Cronin enhancement, nuclear shadowing of the parton distribution functions, and enhanced energy loss in the produced matter.

If the matter produced in central $AuAu$ collisions is indeed sufficiently "opaque" to influence the distribution of fragmentation from hard and semi-hard scatters, it is reasonable to expect a modification of the azimuthal correlation of back-to-back jets relative to that observed in $p+p$ collisions. To examine this possibility, it must first be demonstrated that it is even possible to isolate correlations related to jet production in nucleus- nucleus collisions.

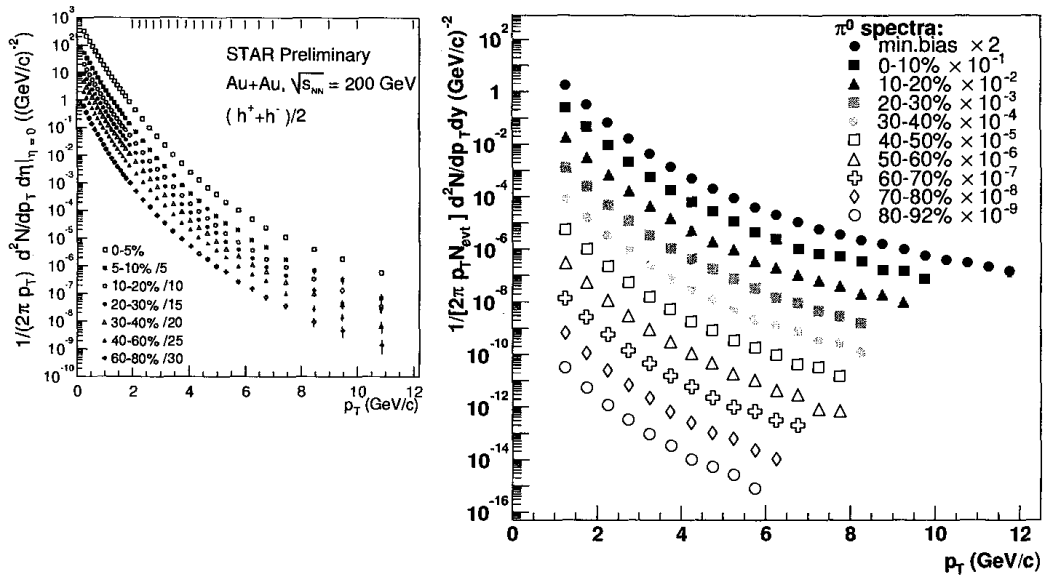


Fig. 4 Inclusive hadron invariant distributions as a function of p_t for various centralities from STAR (charged hadrons; left panel)[13] and PHENIX (π^0 mesons; right panel). [14]

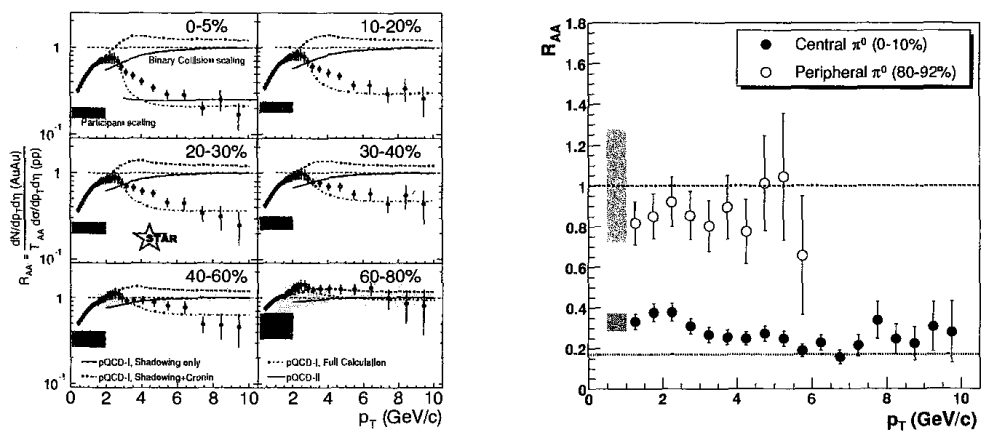


Fig. 5 $R_{AA}(p_t)$ plotted for inclusive charged hadrons (STAR, left panel) and inclusive π^0's (PHENIX, right panel) for events of differing centrality.

It is not possible at RHIC to use traditional high energy jet finding techniques, since even though $\sqrt{s_{NN}}$ is high enough to result in measureable yields of high p_t fragments from jets, fluctuations in soft particle production from the underlying event result in an unacceptably

high rate of "false" jets using traditional cone type jet finding techniques. This should not be a problem for future nucleus-nucleus jets studies at the LHC, where the yield of high p_t ($\sim$ 100 GeV/c) jets should be large. It is expected based on simulations to be a tractable problem for lighter nuclear species at RHIC as well, but for full energy $AuAu$ collisions at RHIC, the p_t range accessible in practice is limited to $35 - 40$ GeV/c, and a different technique must be used.

The approach taken is to look for a correlation among the leading particles of a jet; a "trigger" particle above a given threshold in p_t ($\sim$ 6 GeV/c) is selected offline in software, and the distribution in azimuthal angle of all other particles in the event having p_t above a second threshold ($\sim 2 - 4$ GeV/c) are examined relative to the trigger particle. The distribution from this type of analysis is shown in Fig. 6. To properly extract the correlation resulting from jet fragmentation, it is necessary to account for the effect of elliptic flow, which is significant as discussed above. This is accomplished by noting that the correlations resulting from jets will be relatively small in size ($|\Delta\eta| < 0.5$) whereas the correlation resulting from elliptic flow will be large, affecting particle emission globally in the event. Calculating the correlation separately for particles having $|\Delta\eta| < 0.5$, and for particles having $|\Delta\eta| > 0.5$ and subtracting the two distributions, one observes the correlation shown in Fig. 6 (bottom panel). This correlation is typical of that expected for jet fragmentation and is taken as evidence of jet production in $AuAu$ collisions at RHIC [15]. Further study of the size of the fragmentation cone observed as a function of p_t, as well as the charge balance within the cone show the characteristics of the cone match expectations from PYTHIA [16]. (e.g. h^+h^-/h^+h^+ & $h^-h^- \sim 2.5 - 2.7$)

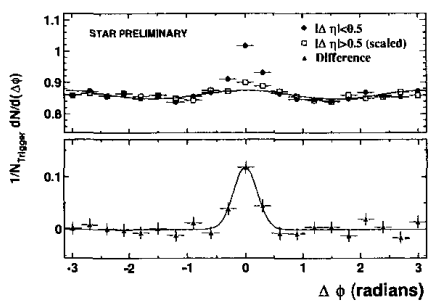


Fig. 6 (Top) The distribution with respect to a "trigger" particle having $p_t \sim 6$ GeV/c of the relative azimuth of all other particles in the even ($p_t \gtrsim 2 - 4$ GeV/c) for $|\Delta\eta| < 0.5$ (closed circles) and $|\Delta\eta| > 0.5$ (open squares). (Bottom) The difference of the above correlations. Clear evidence of the correlation expected for jets is visible at $\Delta\phi = 0$.

Having isolated the "same-side" correlation for leading particles for jets, the next step is to examine the "away side" or back-to-back correlation expected from hard and semi-hard parton scattering. Since back-to-back jets by definition are expected to have a broad correlation in $\Delta\eta$ and $\Delta\phi$, a subtraction like that used above for the same side correlation will not work. The technique used instead is to test the hypothesis that if $AuAu$ collisions are simply a superposition of individual nucleon-nucleon collisions, the back-to-back correlation they exhibit should be the same as that for pp collisions, taking the underlying elliptic flow component into account. Mathematically, the azimuthal distribution in $AuAu$ in this instance is

$$C_2^{AuAu} = C_2^{pp} + B(1 + 2v_2^2 cos(2\Delta\phi)) \tag{4}$$

where C_2^{pp} is the azimuthal back-to-back distribution measured in pp collisions, and v_2 is measured independently; B is an arbitrary normalization constant to account for combinatoric background fit in the non-jet region ($\Delta\phi \approx \frac{\pi}{2}$). The observed back-to-back correlation for peripheral and central collisions are shown in Fig. 7. The open circles indicate the correlation expected from pp collisions when superimposed with the elliptic flow observed in $AuAu$ collisions. The closed points are the observed back-to-back correlation. In peripheral events, the data match the expectation from pp collisions plus flow. For central events however, back-to-back correlation has disappeared.

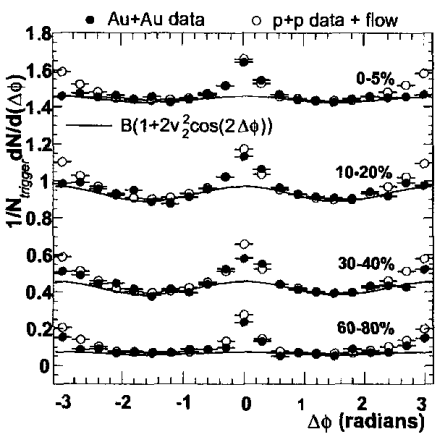


Fig. 7 The correlation as a function of relative azimuthal angle of particles in an event above $\sim 2-4$ GeV/c with a "trigger" particle having $p_t \sim 4-6$ GeV/c for events of different centrality. The correlation expected based on a superposition of pp collisions plus elliptic flow is represented by the open circles. While the correlation observed in peripheral collisions (bottom panel) is consistent with that for pp+ flow for both the same side and away side (back-to-back) correlation, for central $AuAu$ collisions (top panel), the away side correlation disappears.

These data provide tantalizing evidence that the matter produced in central $AuAu$ collisions may be opaque to fragmentation from jets, and that a large amount of radiative energy loss due to gluon bremsstrahlung in this matter may result in "quenching" of the jets which traverse it. This signature of high gluon density matter was postulated by Bjorken [17] more than 20 years ago, and has been discussed as well by many authors since [18-21]. It is consistent with all of the observations discussed in this paper, but can not be definitively concluded as the correct interpretation without additional $d + Au$ measurements to demonstrate that the observed suppression of high p_t particle production is not due to initial state effects of the colliding nuclei [22]. Those measurements will be provided by the forthcoming RHIC run.

As a final remark, it is noted that first measurements of inclusive open charm production scaled by the number of binary collisions based on single electron inclusive measurements [23] in $AuAu$ collisions at RHIC appear to be consistent with predictions PYTHIA and do not show a suppression. It has been discussed [24] that heavy quarks should not experience the same energy loss due to gluon bremsstrahlung and would therefore not be suppressed to the

same degree as light quarks traversing a high gluon density medium. This first measurement of heavy flavor highlights the interest in open charm and charmonium measurements that future $AuAu$ running will provide at RHIC.

References

[1] J.-Y. Ollitrault, Phys. Rev. D**46**, 229 (1992).

[2] P. Kolb, J. Sollfrank, and U. Heinz, Phys. Rev C **62**, 054909 (2000).

[3] P. Kolb, J. Sollfrank, and U. Heinz, Phys. Lett. B **459**, 667 (1999).

[4] D. Teaney and E.V. Shuryak, Phys. Rev. Lett. **83**, 4951 (1999).

[5] K.H. Ackermann *et al.*, Phys. Rev. Lett. **86**, 402 (2001).

[6] C. Adler *et al.*, Phys. Rev. Lett. **87**, 182301-1 (2002).

[7] C. Adler *et al.*, Phys. Rev. C **66**, 034904 (2002).

[8] M. Gyulassy, I. Vitev, and X.N. Wang, Phys. Rev. Lett. **86**, (2001) 2537.

[9] B.B. Back *et al.*, Phys. Rev. C**65**, 031901(R)(2002); K. Adcox *et al.*, Phys. Rev. Lett. **86**, 3500 (2001); I.G. Bearden *et al.*, Phys. Lett. B**523**, 227 (2001).

[10] J.W. Cronin *et al.*, Phys. Rev. D**11**, 3105 (1975); D. Antreasyan *et al.*, Phys. Rev. D**19**, 764 (1979).

[11] E. Wang and X.N. Wang, Phys. Rev. C**64**, 034901 (2001).

[12] I. Vitev and M. Gyulassy, hep-ph/0209161.

[13] C. Adler *et al.*, Phys. Rev. Lett. **89**, 202301 (2003).

[14] S.S. Adler *et al.*, nucl-ex/0304022, April 2003, Phys. Rev. Lett. (submitted).

[15] C. Adler *et al.*, Phys. Rev. Lett. **90**, 082302 (2003).

[16] T. Sjostrand, Comput. Phys. Commun. **82**, 74 (1994).

[17] J.D. Bjorken, FERMILAB-Pub-82/59-THY.

[18] M. Gyulassy and M. Plumer, Phys. Lett. B**243**, 432 (1990); R. Baier *et al.*, Phys. Lett. B**345**, 277 (1995).

[19] R. Baier, D. Schiff, and B.G. Zakharov, Ann. Rev. Nucl. Part. Sci. B**50**, 37 (2000).

[20] X.N. Wang and M. Gyulassy, Phys. Rev. Lett. **68**, 1480 (1992); X.N. Wang, Phys. Rev. C**58**, 2321 (1998).

[21] E. Wang and X.N. Wang, Phys. Rev. Lett. **89**, 162301 (2002); F. Arleo, Phys. Lett. B**532**, 231 (2002).

[22] D.E. Kharzeev and J. Raufeisen, nuc-th/0206073.

[23] K. Adcox *et al.*, Phys. Rev. Lett. **88**, 192303-3 (2002).

[24] Y. Dokshitzer and D. Kharzeev, Phys. Lett. B**519**, 199-206 (2001).

Highlights from Gran Sasso

A. Bettini
Dipartimento di Fisica G. Galilei dell'Università di Padova
INFN Gran Sasso National Laboratory and Sezione di Padova

Abstract

Experiments in underground laboratories searching for rare phenomena in the low background environment they provide can give information on energy scales that cannot be reached by present or future accelerators. Already and for the first time experiments on neutrino produced by natural sources have given evidence for physics beyond the Standard Model. We are now planning the next phase of the new neutrino physics both with natural and artificial neutrino sources. The search for cold dark matter is another avenue to new physics that must be followed.

After an introduction on the new neutrino physics I'll report on the results obtained at the INFN Gran Sasso National Laboratory and on the perspectives of its future program.

1. Introduction

Underground laboratories are complementary to those with accelerators in the basic research of the elementary constituents of matter, of their interactions and symmetries, providing the low radioactive background environment necessary to the search for these extremely rare, nuclear and subnuclear phenomena. Already and for the first time experiments on neutrinos produced by natural sources have given evidence for physics beyond the Standard Model.

The Standard Model of subnuclear physics is the greatest synthesis of experimentally known phenomena from the smallest structures of matter to the Universe itself, as the result of experiments performed at particle accelerators and colliders and to the related theoretical investigations and breakthroughs. The coupling constants of three fundamental forces, the strong, the electromagnetic and the weak, become closer to one another when energy increases, converging to a common value at 10^{16} GeV, the unification scale. As for the last force, gravity, we do not even have a microscopic theory, but only its macroscopic approximation. To discover it we need experimental or at least observational input. Presumably the quantum features of gravity show up around the Planck scale, 10^{19} GeV, which is extremely large just because the gravitational charge is so small. Experiments at colliders have tested the Standard Model with extreme precision, but only at very low energies compared to these scales. The colliders of the next generation will reach 1000 GeV, which may be enough to discover supersymmetry, but still thirteen and sixteen orders or magnitude lower than the unification and the Planck scales respectively. It is clearly impossible to build an accelerator to reach those energies.

Astro Particle Physics is, albeit indirectly, the only way to have experimental, or at least observational, information of those scales. We try to understand elementary particles and their

fundamental interactions from the observation of natural, but very rare, phenomena like proton decay, neutrino-less double beta decays, phenomena connected with the Majorana mass, etc. Or we can detect and study radiations from the Universe such as the cold dark matter, antimatter, topological defects, etc. In both cases the phenomena are extremely rare and we must reduce all the interfering natural backgrounds, both in the environment and in the detector itself. The struggle against background is the way to push back the high-energy frontiers. Underground laboratories provide the low radioactivity environment necessary for a large class of experiments[1].

As already mentioned, underground experiments, mainly at Kamioka in Japan, at Gran Sasso in Italy and, more recently, at SNO in Canada, have provided evidence for physics beyond the standard model. On the basis of this evidence we know now that some of the assumptions of the Standard Model are not correct. Neutrinos have non zero masses, electron neutrinos, muon neutrinos and tau neutrinos - the particles produced by weak interactions and detected by our apparatuses - are not the mass eigenstates and their flavour quantum numbers are not conserved. We have entered a new neutrino physics[2].

There is also an increasing overlap with positive cross fertilisation between cosmology and elementary particle physics. In particular, cosmology is already giving important information on neutrinos properties and hints on supersymmetric particles. For a review see [3]

There has been tremendous progress in cosmology in the last years, both in measuring and in modelling the Universe. Large amounts of data are produced with unprecedented precision with different types of telescopes on the surface, on balloons and on satellites and by new laboratory experiments. As the Universe is unique, the only possibility to compare a theory with Nature is through consistency checks. The data becoming available allow rather stringent tests overconstraining the models, and, in particular, a Standard Cosmological Model has emerged. Even if its theoretical bases are not so robust as those of the elementary particles Standard Model, it gives a unified and self-consistent description of the available experimental and observational data. The model is based on a flat, accelerating universe, whose structures are seeded by initial quantum fluctuations and which has gone through an early period of inflation. The main parameters of the model are rather well determined, often from completely independent observations: the expansion rate (Hubble constant), the total mass–energy density, the matter density and its components (baryons visible and invisible, cold dark matter and neutrinos) and the cosmological constant or dark energy density.

In summary, we know that the curvature of the Universe is compatible with zero, hence its matter-energy density has the critical value, with a contribution from matter of Ωm =33±4% and one from an unknown property of the vacuum (the cosmological constant, "energy", "tension"?) of Ω_Λ = 67±7%, where the densities are expressed as fractions of the critical densities. The baryonic component of matter is small, Ω_b = 4±0.8%. The contribution of the stars, the part we see, to this baryonic matter is even smaller, Ω_{stars} = 0.5±0.2%, so that about 90% of baryonic

matter is dark. The largest fraction of matter, again almost 90%, is non-baryonic and, once more, "dark". It does not emit light or, more generally, detectable electromagnetic radiations. In principle, it might be made of neutrinos, but, again, the evidence is that its largest fraction is made of particles that were non-relativistic at the time of decoupling, not of neutrinos. This, cold dark matter, component has a density $\Omega_{cdm} = 30\pm4\%$. Weak interacting particles, with no electric charge because they do not radiate, are around us, fill the space, undetected, unknown to us in their nature.

In conclusion, the Standard Cosmological Model is now internally consistent and over constrained and its parameters are known at 10 - 20 % level. But it is based on a dark matter we do not see, a dark energy we do not understand, and a large fraction of baryons we can't find. On the other hand the model gives, both to astrophysics and to particle physics, clear tasks to be pursued in the next years. A contribution to the search has and will be given by experiments at Gran Sasso.

2. The Gran Sasso Laboratory (LNGS)

The INFN Gran Sasso Laboratory is located beside the freeway tunnel (10.4 km long) connecting L'Aquila (West) and Teramo (East), at about 6 km from the west entrance, 120 km from Rome.

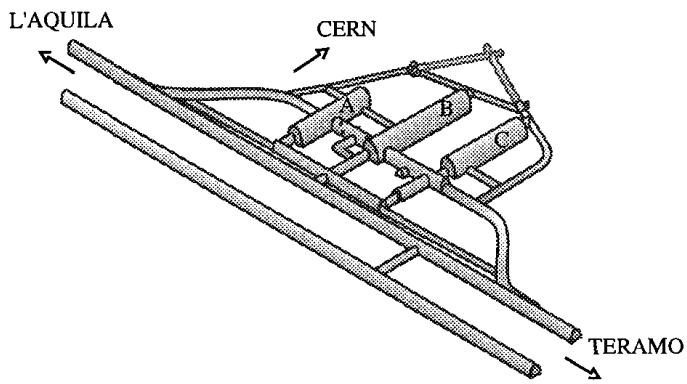


Fig. 1. Artist's view of the underground facilities of the Gran Sasso Laboratory

Fig. 1 shows a view of the facility. The access is through the West heading freeway gallery. Two internal tunnels parallel to the freeway can be seen in Fig. 1, the narrower is for cars, while the wider U-shaped one is for lorries, allowing the easy transportation and installation of large pieces of apparatus. The underground facilities consist of three experimental halls, called hall A, B and C, and a set of connecting tunnels and service areas, for a total surface of 18 000 m^2. The three halls are approximately 100 m long, 18 m wide, 18 m high and are oriented toward CERN, to host experiments on a neutrino beam from that laboratory. The infrastructures of the laboratory are completed with facilities to support the experimental activities (assembly halls, mechanical and electronic shops, chemistry department, computing and networking, safety, security and

environmental services, library and canteen) located on the surface, near the West entrance of the freeway tunnel.

The laboratory is located at 963 m above sea level. The rock overburden is 1400 m, with a rather flat shape, providing uniform coverage at all angles, with cosmic rays muon flux attenuation by 10^6. The neutron fluence underground does not depend substantially on the depth beyond a few hundreds metres, but it does on the radioactive components of the surrounding rocks. The dolomite surrounding the Gran Sasso Laboratory is particularly radio-clean and the neutron fluence is one thousand times less than on the surface.

The mission of the laboratory is to host and run experiments in fundamental physics requesting very low levels of radioactive background and researches in other disciplines (notably geophysics and biology) that can profit of the unique environmental characteristics of the site.

Neutrino physics will be the principal, but not the only issue of the research program for the next years. Experiments both with naturally produced neutrinos (from the Sun, from the atmosphere and from Supernova explosion) and artificially produced ones (mainly from CERN, but possibly by other sources too) are running, being built or planned. Other experiments try to understand the nature of the electron neutrino and search for the Majorana mass, still others will continue with increased sensitivity the search for non-baryonic dark matter. At the underground accelerator facilities the measurements of thermonuclear cross-sections at energies relevant for the stars and Sun are continuing. For a review see ref.[4].

3. The Search for Cold Dark Matter

A hint on the nature of cold dark matter comes from elementary particle physics. Supersymmetry, the principal candidate extension of the Standard Model, predicts a large number of new elementary particles. The lightest, presumably the neutralino, is stable, provided *R*-parity is conserved. As such, neutralinos produced very early in the history of the Universe should be still present around us. We do not observe them because, like neutrinos, they interact only very weakly with matter, but neutralinos might account for a large fraction of the mass of our Universe. More generally these hypothetic particles are called WIMPs (weak interacting massive particles).

There are two basic approaches to WIMPs detection, indirect - searching for high-energy neutrinos from the Sun and the Earth or antiprotons and positrons in the Galaxy - and direct, performed in underground laboratories. Reviews have been given by A. Morales at TAUP 2001 and the III International Meeting on Fundamental Physics, IMFP2002[5]. Only a few hints will be given here. We assume that our Galaxy contains cold dark matter in the form of WIMPs distributed smoothly as an enormous cloud with a density, near the Sun of 0.3 GeV/cm^3. The Sun moves in the Galaxy at 230 km/s crossing the WIMPs cloud. The Earth moves around the Sun at 30 km/s.

In the direct detection of WIMPs one measures the energy of the nuclear recoil produced by an elastic scattering. For masses from several tens of GeV (masses below 30 GeV have been excluded by LEP) to a few TeV, the recoil energies are very low, E_k = 1 - 100 keV. At these low energies radioactive background is severe, being due not only to the environment (working underground is mandatory) but also to the detector itself and its surroundings. The signal spectrum is a decreasing function of the energy, very similar to that of the background.

The signals from WIMPs are very rare too. The use of heavy nuclei as target is convenient, as neutralinos, in many coupling schemes, interact coherently with nuclei, with a probability proportional to A^2. But even for coherent interactions the rates are small. For example supersymmetric models predict rates between 10 and 10^{-6} events per day per kilogram of detecting mass. A low background environment is clearly mandatory. Detectors must have a very low energy threshold, large sensitive mass, good energy resolution, ultra high radio-purity and efficient background discrimination. In an extreme synthesis there are two main techniques to fight the background.

The first technique is used at Gran Sasso by DAMA[6], which looks for an annual modulation of the rate due to the motion of the Earth around the Sun, and indeed observes it with the expected characteristics (period, phase and amplitude).

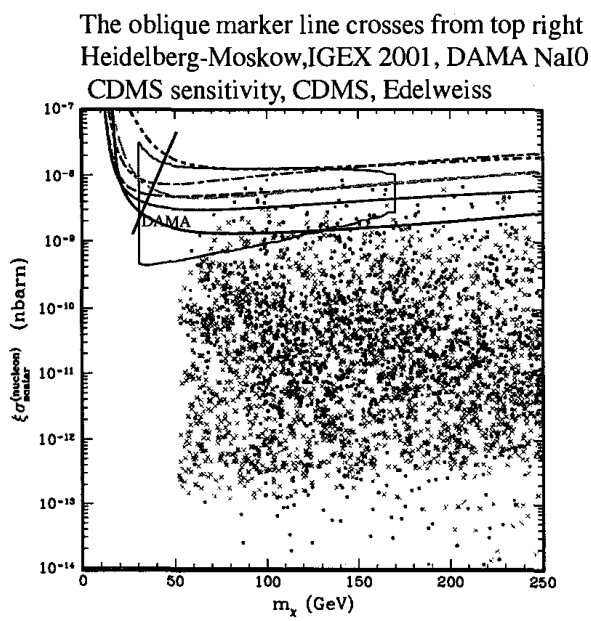


Fig. 2. Exclusion curves are from a compilation by S. Cebrian, I. Irastorza and A. Morales[5]; theoretical expectations and contour of the DAMA region are from A. Bottino, N. Fornengo and S. Scopel. (Somewhat surprisingly CDMS limit is better than CDMS sensitivity)

The second technique, used at LNGS by CRESST, takes into account that there are two main backgrounds: gamma rays and electrons on one side, nuclear recoils due to elastic scattering by neutrons on the other. The second is practically indistinguishable from the signal; the first can be separated by measuring two quantities, energy deposit and ionisation (or light). The several experiments using these techniques have provided only limits till now.

A compilation by the Zaragoza group of the limits together with the DAMA region and some theoretical expectations is shown in Fig. 2. The figure reports the available results in the parameter space of WIMP-Nucleus cross-section (assuming a standard WIMP density) vs. WIMP mass. Notice that comparison amongst different experiments is not trivial because the annual modulation signal, once all artefacts have been excluded, gives model-independent positive evidence, while the other approaches give lower limits, which depend both on the model assumed to correlate the experimental negative evidence with the parameter space and on the experimental technique. The "DAMA region" in the figure is model-dependent too. Finally the comparison between different target nuclei is model-dependent.

On the other hand, one can see from Fig. 2 that much remains to be done to explore the parameter space. At LNGS two dedicated experimental collaborations follow complementary lines.

The DAMA collaboration has already collected data for three years that are being analysed and will be published early in 2003. In parallel the LIBRA set-up is being installed, using the same technique as DAMA, with several improvements and with an increased sensitive mass of 250 kg.

The CRESST collaboration searches for WIMPS with cryogenic detectors working at several milliKelvin temperatures. More traditional methods rely on the detection of the ionisation and of the scintillation produced by the recoiling nucleus. These methods are limited by the relatively high energy involved in the ionisation process and by the sharp decrease with decreasing energy of the ionisation probability by nuclei. On the contrary, cryogenic detectors use much lower energy excitations, such as phonons. The cryogenic calorimeters developed by CRESST are both targets and detectors. The basic unit is a crystal on a surface of which a superconducting film is evaporated. When the film is held at a temperature in the middle of the superconducting to normal transition, it becomes a very sensitive thermometer. Crystals can be made with different nuclear composition, an important possibility for the verification of an eventual positive signal.

Background discrimination will be achieved by CRESST2[7], the experiment now under construction, simultaneously detecting phonons and light. As shown in Fig. 3, obtained with an exposure of a prototype detector to γ, electron and neutron sources, high discrimination efficiency against the dominant gamma background can be achieved even at the low (15 - 20 keV) relevant energies. For the presently used $CaWO_3$ crystals, the measured rejection is 99.7% for E>15 keV.

During 2002 the experimental hut, used by the completed experiment CRESST1, has been moved from Hall B (where the space was needed for experiments on CNGS) to Hall A and a new apparatus was built. The sensitive mass will be 10 kg, based on 33 modules of 300 g each. The

expected sensitivity with an exposure of 10 kg years, an energy threshold of 15 keV and a background at threshold of 1c/(kg keV d), is below 10^{-10} nbarn, allowing exploring a relevant fraction of the parameter space shown in Fig. 2.

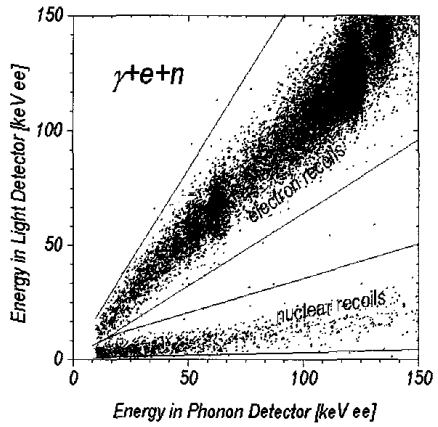


Fig. 3. Light yield vs. energy deposit as measured in a $CaWO_3$ CRESST2 crystal

A third experimental option has been proposed by GENIUS, which I'll describe later, as a by-product of its search for double beta decay. The idea is to use Ge crystals naked in a bath of liquid nitrogen, which can be produced extremely radio-pure. Even in this case the crystals are both targets and detectors. Notice that, while for double beta decay Ge must be enriched in the ^{76}Ge isotope, natural germanium can be used for dark matter search. For example a sensitive mass of 100 kg with a background rate at threshold of 3×10^{-3} c/kg keV d will be sensitive to 10^{-11} nbarn.
I'll notice here that experience shows that many unforeseen difficulties are encountered in reducing the background rates to the proposed values. Both dark matter and double beta decay searches are extremely difficult; the achieved background level largely dominates the sensitivities of both. The background sources are completely different at the relatively high energy (2 MeV) of double beta decay and at the very low energies (several keV) for dark matter. As a consequence I believe it worthwhile to have dedicated experiments, and to be very risky to try to have a frontier detector for both searches.

4. The New Neutrino Physics

In the last years, experiments on electron neutrinos from the Sun and electron and muon neutrinos indirectly produced by cosmic rays in the atmosphere have shown that, very likely, neutrinos oscillate amongst states of definite flavour. This phenomenon, forbidden in the Standard Model, implies that, contrary to its assumptions: 1. electron, muon, and tau neutrinos are not the mass eigenstates, which we will call ν_1, ν_2, and ν_3, with masses m_1, m_2 and m_3 respectively; 2. at least two of the masses are not zero; 3. flavour lepton numbers are not conserved, so that, for example, an originally pure muon neutrino beam can produce tau leptons. It is just a matter to let it fly for a

long enough time. While in the case of quarks, each mass eigenstate is dominated by a single flavour component, neutrino eigenstates are radically different from flavour states. Neutrino eigenstates do not have a dominant flavour component[2].

To introduce the concept of neutrino oscillations, consider the case of two neutrino flavours, ν_α and ν_β, linear superpositions (mixing) of the mass eigenstates. The mixing matrix elements can be expressed in terms of one real parameter only (Dirac neutrinos), the mixing angle θ.

In initially pure ν_α beam of energy E, the other, ν_β, can be detected with a probability that is an oscillating function of the flight distance L, $P_{\nu_\alpha \to \nu_\beta} = \sin^2 2\theta \sin^2\left(1.27\Delta m^2\left(\mathrm{eV}^2\right)\frac{L(\mathrm{km})}{E(\mathrm{GeV})}\right)$.

The oscillation period is inversely proportional to $\Delta m^2 = m_2^{\,2} - m_1^{\,2}$, the difference between the squares of the masses of the two eigenstates (absolute value). The oscillation amplitude is $\sin^2 2\theta$, a factor that is largest (100%) for $\theta = \pi/4$. This case is called maximum mixing.

If an electron neutrino (or antineutrino) beam propagates into matter (in the Sun, in the Earth or in a Supernova) another mechanism may cause flavour conversion, the so-called MSW effect[8]. Electron neutrinos (and antineutrinos) interact with electrons via charged current weak interactions giving, in particular, coherent forward scattering, a phenomenon exactly similar to the forward scattering of light in matter at the origin of the refractive index. Similarly to light, the electron neutrino wave acquires a refractive index different than in vacuum, or, equivalently a different effective mass. Notice that the effect is proportional to the forward scattering amplitude, hence to the Fermi constant, not to the cross section, which is proportional to its much smaller square. This explains its importance. The phenomenon is limited to electron neutrinos because the other neutrino flavours interact via neutral current only and the effects of negative electrons and positive nuclei compensate.

The effect depends on the electron density (N_e) times the neutrino energy (E). In appropriate conditions a level crossing phenomenon takes place: at a critical N_eE value the effective electron neutrino "mass" becomes equal to that of a different flavour. Electron neutrinos will convert into neutrinos of that flavour, or vice versa, while crossing a variable density medium when they reach this critical value.

In Nature, neutrino flavours are three. The flavour eigenstates are linear combinations of the mass eigenstates, $\nu_l = \sum_{i=1}^{3} U_{l,i}\nu_i$, where $l = e, \mu, \tau$ and U is the mixing (unitary) matrix.

In the Standard Model neutrinos and antineutrinos are distinguished because they have opposite chirality and because they have opposite lepton number. But if a neutrino has non-zero mass it moves with a speed less than c and we can find a reference frame in which the neutrino velocity is in the opposite direction: chirality is not a good quantum number. If not even lepton number is conserved, neutrino and its charge conjugate antineutrino are identical. Neutrinos are then Majorana particles.

The elements of the unitary mixing matrix can be expressed in terms of six independent real parameters: three mixing angles, θ_{12}, θ_{13} and θ_{23}, and three phase factors, which can give CP violation. Two phase factors, α and β, can be eliminated if neutrino and antineutrinos are different (Dirac) particles; they are relevant in double beta decay but are irrelevant for oscillations.

The situation is much more complex than for two neutrinos. Two different oscillations take place with different frequencies (or at different flight times), with two considerably different periods; as a consequence, the resulting probability for appearance of a new flavour and for disappearance of the original one is not in general a simple oscillation between two states. It can be approximated as such in two cases: at flight times comparable with the first period but much shorter than the second and at times comparable with the second and much longer than the first; in this last case the non monochromaticity of the beam in practice averages the first oscillation, which becomes non observable. Even when this two-flavour approximation holds, the oscillation amplitude is not, in general, $\sin^2 2\theta$, as often but wrongly assumed.

Neutrino oscillations have been searched for since decades with artificial neutrino beams from accelerators (mainly muon neutrinos) and reactors (anti electron neutrinos) but no reliable experiment has reported a positive and independently confirmed result (before K2K). This can be understood if the square mass differences are so small that the oscillation times, or the flight paths, requested to observe oscillations, are very large, much larger than the characteristic lengths, order of 1 km, of most neutrino beams.

On the other hand, deficits in the fluxes of electron neutrinos from the Sun and of muon neutrinos from the atmosphere have been observed with increasing reliability and precision. The simplest interpretation of both anomalies is that they are due to two different oscillation phenomena, both at L/E values much larger than those available at accelerators up to recently with the K2K experiment.

The anomaly in atmospheric neutrinos has been observed by several experiments, as reviewed in ref. [9]. Its interpretation in terms of muon neutrino disappearance in an oscillation phenomenon became gradually established in the last years, due mainly to the Super-Kamiokande[10] and confirmed with a tracking technique by MACRO[11].

At the L/E values involved in atmospheric oscillations, the "solar" oscillation has not yet started and does not affect the data. The two-neutrino oscillation formalism is a good approximation with $\theta = \theta_{23}$. The best fit of the Super-Kamiokande data to the oscillation parameters is for $\Delta m^2 = 2500$ meV^2 and $\theta_{23} = \pi/4$. The confidence level interval for the square mass difference is 1800 meV$^2 < \Delta m^2 < 4000$ meV^2 at 90% confidence level. MACRO gives consistent results within a larger confidence interval.

The CHOOZ[12] experiment, a disappearance experiment on anti electron neutrinos from reactors (1 km baseline) provides the best upper limit on $\theta_{13}^2 \approx |U_{e3}|^2 < 0.025$.

The solar neutrino puzzle has been known for three decades from the Homestake experiment. Gradually, a series of experiments, Kamiokande and Superkamiokande, GALLEX and SAGE (but also helioseismology and the measurement of the most relevant nuclear cross section by LUNA) not only confirmed the anomaly, but also excluded both nuclear and astrophysical interpretations, clearly pointing to neutrino flavour conversion. The field has been reviewed by T. Kirsten[13]. More recently, SNO showed definite evidence for appearance of flavour, ν_μ and/or ν_τ, through their neutral current interactions.

To summarise the present status, I'll use the fit of G. Fogli and collaborators[14]. These authors have made a global fit of all the solar data (81 observables), including the total event rate from the chlorine experiment, the total gallium event rate and its winter-summer difference, the Superkamiokande electrum spectra for different zenith angle bins and the day-night energy spectrum from SNO. The relevant flight times are much longer than the "atmospheric" oscillation period, so that the effects of this oscillation average out and do not affect the data and the two-neutrino approximation can be used.

The results are shown in Fig. 4. The best-fit solution is the Large Mixing Angle (LMA), but others are statistically acceptable, in the low mass-difference regime (LOW) and in the quasi-vacuum regime (mainly vacuum oscillations). For LMA and LOW, MSW flavour conversion is dominant. Central values for the LOW are δm^2 = 55 meV^2, $\tan^2\theta_{12}$ = 0.42. Notice, in particular, that θ_{12} appears to be close, but not equal to maximal mixing.

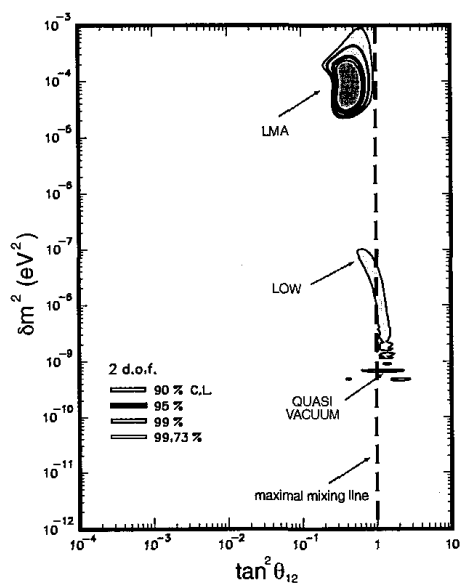


Fig. 4. Solutions of the solar anomaly, for $\theta_{13} = 0$

From this evidence we know that the neutrino mass spectrum consists of a doublet plus a singlet. The singlet has a very small (if any) electron neutrino component (U_{e3} is small). We call m_1, m_2

with $m_2 > m_1$ the masses of the doublet and $\delta m^2 = m_2^2 - m_1^2$. We call the mass of the singlet m_3 and $\Delta m^2 = m_3^2 - m_2^2$. The smaller mass difference δm^2 is responsible for the solar anomaly, the larger one Δm^2 for the atmospheric one. As we do not know either the sign Δm^2 or the absolute value of the masses, there are three basic alternatives for the mass spectrum: almost degenerate, when the masses are much larger than the mass differences, "normal" ($\Delta m^2>0$) if the singlet is higher, or "inverted" if it is lower ($\Delta m^2<0$) than the doublet.

The absolute scale of neutrino masses is smallest in the hierarchic case. For example for normal spectrum and $m_3 \geq \sqrt{\Delta m^2} \approx 50$ meV and, for the favourite (but not at all guaranteed) LMA solution, $m_2 \geq \sqrt{\delta m^2} \approx 7.5$ meV.

We can conclude that, very likely, the unit for neutrino masses is the millielectronvolt.

Notice that in the seesaw mechanism $m_i = v^2/M$, where M is a large mass, the energy scale of lepton number violation. With $v=M_{top}$ and m_i a few meV, we have $M \approx 10^{15}$ - 10^{16} GeV. The smallness of the neutrino masses, compared with quarks (or to the electroweak scale), gives a lepton number (and baryon number, perhaps) violation scale close to the GUT scale. We see, in this example, how the search for rare phenomena can give information on the largest energies.

5. Neutrino masses from cosmology

The non baryonic dark matter in the Universe has two components, distinguishable due to their different effects on the large scale structures: "hot" and "cold" depending on whether its particles were relativistic or not at the time of decoupling. Neutrinos belong to the hot dark matter. On the other hand, the number densities of the three neutrino states, ν_1, ν_2, and ν_3, are independent on their masses. As a consequence a limit on (or a value of) neutrino mass density gives a limit on (a value of) the sum of the masses of the three neutrinos.

Important progress has been achieved and will continue in understanding the non-homogeneity of the Universe at large scales. The largest observed structures are up to 200 Mpc across and only the recent galaxy surveys (LCRS, SDSS and 2dFGRS) extend to large enough distances (order of 1000 Mpc, $z \approx 0.3$). At these scales Galaxies can be considered point-like. The surveys determine the power spectrum $P(k)$, which is, in simple words, the Fourier-transform of the distribution of the density fluctuations at large scales. The "wave number" k is the conjugate of the size of the structures.

Neutrinos, moving at relativistic speeds tend to suppress smaller scale structures. The observed power spectrum at small and intermediate scales (large k) follows a power law, turning over at larger scales. Its shape depends on neutrino density, which, in its turn, is linked to total neutrinos mass by

$$h^2\Omega_\nu = \frac{\sum m_i}{94\text{eV}}$$

where h is the reduced Hubble constant. Presently, the largest data set available is that of 2dFGRS[15]. Using these data, and the values of the other cosmological parameters, Ø. Elgaroy et al.[16] have recently published an upper limit on neutrino masses, $\sum_{i=1}^{3} m_i < 1.8$ eV.

This limit translates immediately in a limit on each mass, observing that close to the limit neutrinos are almost degenerate. Then m_1, m_2, $m_3 <$ 600 meV, implying that $m_{\nu e}$, $m_{\nu\mu}$,$m_{\nu\tau}$ < 600 meV, the best limits on neutrino masses! Other authors, Lewis and Bridle[17], claim a limit twice as small. Notice that this is the only, albeit indirect, means to have a limit on, or a value of, the sum of the masses of the stationary states, independently on the mixing parameters and CP violating phases. Further progress is expected from the data of the other surveys, for example by the Sloan Digital Sky Survey (SDSS), which is expected to measure the spectrum with 1% accuracy and by increased precision on the other cosmological parameters. We can expect to be in a few years sensitive to $\sum m_i$ = 300 meV. Optimistically, if systematic uncertainties will be also reduced, Cosmology may become close to the lower limit of neutrino masses, from atmospheric and solar data (at least in case of LMA), $\Sigma m_i > \sqrt{\Delta m^2} + \sqrt{\delta m^2}$ = (50+7.5) meV.

6. An interlude on neutrino masses

Mass is a property of the eigenstates in the vacuum i.e. of the stationary states, the states that have well defined mass. Historically, since a few years ago, neutrino flavour states have been assumed (in the Standard Model) to be coincident with mass eigenstates with "masses" and terms like electron neutrino mass, muon neutrino mass are widely used. We know now that this is improper and in some cases misleading. What is meant depends in fact on what one measures (or limits). To clarify these concepts I'll try to start with a gedankenexperiment.

Suppose neutrinos were two, ν_1 and ν_2, with masses m_1 and m_2 and that ν_μ and ν_τ were maximum mixings of ν_1 and ν_2: $\nu_\mu = \frac{1}{\sqrt{2}}(\nu_1 + \nu_2)$ and $\nu_\tau = \frac{1}{\sqrt{2}}(\nu_1 - \nu_2)$.

Consider the decay at rest of a pseudoscalar meson of mass M into a μ and a ν_μ. To determine the neutrino masses, we measure the neutrino energy E_ν (in practice is the muon energy, obviously). We should find a dichromatic energy spectrum with two lines at, say $E_{\nu 1}$ and $E_{\nu 2}$ corresponding to m_1 and m_2. The lines have equal height, because the mixing is maximum.

This conclusion appears to be absurd at first sight for the following reason. Having a sample of mesons at rest we can produce a tagged beam pure in ν_1's. Now $\nu_1 = \frac{1}{\sqrt{2}}(\nu_\mu + \nu_\tau)$ and when these tagged neutrinos interact they will produce μ's and τ's with equal probabilities. τ's are produced by an originally pure ν_μ beam. But, is it really absurd?

How much is the resolution needed to separate the two lines? It is easy to calculate that the energy difference is $E_{\nu 2} - E_{\nu 1} = \frac{m_2^2 - m_1^2}{2M} = \frac{\Delta m^2}{2M}$. To simplify the argument, assume that M is large, then $E_{\nu 2} - E_{\nu 1} \approx \frac{\Delta m^2}{2E_\nu}$.

Considering that the two components are two monochromatic waves with periods $T_1=1/E_1$ and $T_2=1/E_2$, we might try to measure the two periods by counting the numbers of crests in a certain time interval τ: $N_1 = \tau/T_1 = \tau E_{\nu 1}$ and $N_2 = \tau/T_2 = \tau E_{\nu 2}$. The condition to be able to resolve is that the difference between these two numbers be of the order of one, $\tau(E_{\nu 2} - E_{\nu 1}) > O(1)$. We recognise the uncertainty relation. Notice - in passing - that τ is not the duration of the measurement but the time interval between the preparation of the system and its observation, as indeed in the time-energy uncertainty relation.

In conclusion to resolve the doublet we need a measurement time $\tau > \frac{1}{E_{\nu 2} - E_{\nu 1}} \approx \frac{2E_\nu}{\Delta m^2}$.

Numerically: $\tau = 20$ ms for $\Delta m^2 = 2.5 \times 10^{-3}$ eV^2 and $E = 10$ GeV, giving N_1 and $N_2 \approx 10^{22}$. We need an energy resolution of one part in 10^{22}!!

In practice we cannot count the crests of the wave functions, but we do have a way to count their difference $N_2 - N_1$ by beating the two waves. This happens naturally: the pure monoenergetic ν_μ beam is not a monochromatic (single frequency) but a dichromatic wave: a coherent superposition of two monochromatic waves. The two waves are initially in phase and add constructively, but after a definite time they are in phase opposition and the resulting amplitude is zero. This time is such that the difference $N_2-N_1=1/2$.

It is now clear that we are dealing just with the oscillation phenomenon. $L=E_\nu/\Delta m^2$ is the half period for flavour oscillations. After this time both μ's and τ's are produced by an originally pure ν_μ beam.

Let's go back to real experiments to measure the "flavoured neutrino masses". The best experimental (measured) limits, if neutrinos are Dirac particles, are those on the "electron neutrino mass" $<m_{\nu e}>$. They have been obtained by measuring the electron energy spectrum in the Tritium beta decay. Near the end point the slope is zero, if neutrinos have no mass. If neutrinos are massive, the maximum electron energy decreases (because the smallest mass, say m_1, is non zero) and the spectrum ends with a step of vertical slope. In principle, two further steps are present near the end point, in correspondence with the non zero values of m_2 and m_3. An argument similar to that we have just made shows that the steps cannot be resolved and one measures a weighted average

$$< m_{\nu e}^2 > = |U_{e1}|^2 m_1^2 + |U_{e2}|^2 m_2^2 + |U_{e3}|^2 m_3^2.$$

Presently two experiments[18] give upper limits of a few electronvolts, $<m_{\nu e}> < 2.9$ eV for the Mainz experiment. For the future, a new spectrometer, KATRIN, by the joint Mainz and Troitzk groups aims to be sensitive down to 350 meV.

7. Neutrino masses from double beta decay

If neutrinos are massive Majorana particles, a very rare process, the neutrino-less double beta decay ($0\nu 2\beta$) takes place in some nuclides, a process that violates the lepton number by two units. The double beta active nuclides are stable against normal beta decay - i. e. $Z \rightarrow (Z+1) + e^- + \overline{\nu_e}$ is forbidden - but have the two-neutrino double beta decay ($2\nu 2\beta$) channel open: $Z \rightarrow (Z+2) + 2e^{\pm} + 2\overline{\nu_e}$. This last is a very rare, but standard, second order weak process and happens if the ground level of the Z isotope is lower than that of Z+1 but higher than that of Z+2. Fig. 5a shows the relevant graph for the ($0\nu 2\beta$) decay, forbidden in the Standard Model.

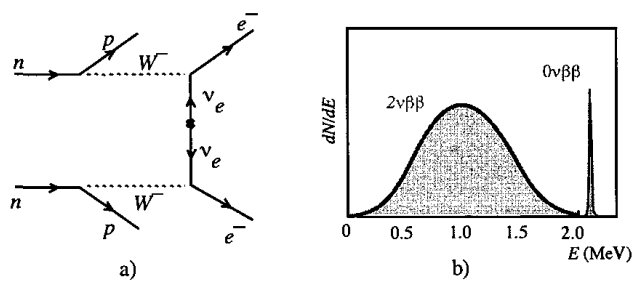


Fig. 5. a) Neutrino-less double beta decay. Arrows indicate the flux of flavour. b) Sum energy spectrum

The most common experimental approach is to measure the total energy released in the decay by the two electrons. Ideally, a spectrum as shown in Fig. 5b is expected: continuous for ($2\nu 2\beta$) decay, where some energy is taken by neutrinos, a single line at the transition energy for ($0\nu 2\beta$), where all the energy goes to the electrons. In practice the signal spectrum is superimposed on the possibly large background spectrum. To fully exploit the advantage given by the mono-chromaticity of the signal, detectors must attain very good energy resolution, which must be coupled to extremely low background conditions.

After measuring or limiting the ($0\nu 2\beta$) lifetime of a nuclide, we can extract the "effective mass"

$$| M_{ee}^{M} | = | [| U_{e1} |^2 m_1 + | U_{e2} |^2 e^{2i\alpha} m_2 + | U_{e3} |^2 e^{2i\beta} m_3 |] |$$

As double beta decays happen in nuclei, the relevant nuclear matrix elements must be known. The uncertainty on the (calculated) matrix elements induces uncertainties on M^M_{ee} typically of a factor two-three (that we'll call h). As a consequence, for a complete research program it is mandatory to include different double-beta active isotopes in the search. Much more theoretical effort, joined to

experiments aimed to measure critical quantities, is clearly also needed to reduce the theoretical uncertainties.

Presently the best limit, $M^M_{ee} < 340\ h$ meV (90% c.l.), based on the half-life limit of 2.1×10^{25} years, is given by the Heidelberg-Moscow[19] experiment at Gran Sasso, obtained with a 37.2 kg yr exposure of an enriched ^{76}Ge detector.

The above expression shows that cancellations can take place in M^M_{ee}, due to the phase factors. This fact forbids, with the present knowledge, to calculate a completely safe lower limit for this quantity. Nonetheless in the vast majority of the possible scenarios $M^M_{ee} \gtrsim$ 10-20 meV. Measuring the absolute value of neutrino masses is clearly of paramount importance and adequate effort should be dedicated in the next years to improve the present sensitivity by an order of magnitude.

Notice that the effective mass M^M_{ee} is inversely proportional to the square root of the half-life $T_{1/2}$ and that the sensitivity of an experiment or, in case of negative result, the limit on the half-life that it can obtain is directly proportional to the square root of the exposure (the detector sensitive mass M times the exposure time t) and inversely to the square root of the energy resolution Δ and of the background rate per unit mass b. In a formula, limit on $\frac{1}{M^M_{ee}} \propto \sqrt[2]{T_{1/2}} \propto \sqrt[4]{\frac{t \times M}{\Delta \times b}}$.

We see that to increase substantially the limit, increasing the sensitive mass is not sufficient. The background must be contemporarily reduced, without compromising the energy resolution. To gain one order of magnitude in neutrino mass sensitivity, one needs, for example, to increase by two orders of magnitude the sensitive mass and decrease by two orders of magnitude the background rate per unit mass. The task is very challenging indeed. I'll focus now on the LNGS programme.

The Heidelberg group has proposed in 1997 at LNGS the GENIUS[20] experiment aiming for a forward jump in the sensitivity with a large increase in the enriched Ge mass (1000 kg) and a drastic reduction of the background. Naked enriched Ge crystals would be used in a liquid N_2 bath, 10 m across used both for cooling the crystals and to screen the external radioactivity. Monte Carlo calculations show that the technique should allow to reduce the background, at the relevant energy, to reach $b = 3 \times 10^{-4}$ events/(kg keV yr), two orders of magnitude better than in the present. This would allow the experiment to reach the 10 meV neutrino mass range. To prove that such a large reduction in the background is possible in practice, the collaboration is building the GENIUS-TF[21] test facility, based on 40 kg of natural Germanium.

The most sensitive experiment on a different isotope (^{130}Te) is MIBETA, again at Gran Sasso, with 20 TeO_2 crystals operated as bolometers at cryogenic temperatures. The total detector mass is almost 7 kg of natural Te corresponding to 2.3 kg of the double-beta active ^{130}Te isotope (34% natural abundance). MIBETA has reached an exposure of 3.26 kg yr with a background level b = 0.6 ev/(kg keV yr), giving the limit $M^M_{ee} < 2$ eV [22].

The next experiment with the same technique is CUORICINO[23] with 56 TeO_2 crystals, 0.76 kg each, corresponding to a total ^{130}Te mass of 14.3 kg. The first crystals are in the test phase. If the background level will be reduced at b = 0.1 ev/(kg keV yr), as it appears to be feasible from the results of the tests, sensitivity approximately of 0.5 – 1 eV will be reached in M^M_{ee}.
Further increase in the mass, by an order of magnitude, and drastic reduction of the background are being studied in view of the CUORE project. It will consist of 1000 natural Te crystals equal to those of CUORICINO with a sensitive ^{130}Te mass of 250 kg. To exploit the larger mass a further reduction of the background is needed. With $b = 10^{-2}$ events/(kg keV yr) and improving the resolution, the limit $M^M_{ee} < 100$ meV may be reached.

8. Neutrinos from the Sun and their spectrum

The light and the other electromagnetic radiation emitted by the Sun are surface phenomena, because the solar matter is not transparent to photons. On the contrary, the nuclear reactions that are at their origin take place near the centre of the Sun. The largest fraction, 98%, of the energy emitted by the Sun surface is produced by the overall reaction $4p \rightarrow {}^4He + 2e^+ + 2\nu_e + 26.73$ MeV, the "*pp* fusion" and by the following annihilation of the two positrons. As a consequence, the electron neutrino flux at the sun core due to this reaction is well, within 2%, known from the observed Sun luminosity. Their energy is relatively low, as shown in Fig. 6b) and their observation became possible only with Gallium experiment, GALLEX at LNGS (and later SAGE at Baksan).
Electron neutrinos absorption is completely negligible in the Sun, but their flux may be altered by flavour conversion into other neutrino flavours. This mechanism, the MSW effect, changes the electron neutrino flux but not the flux of all neutrino flavours.
The electron neutrino spectrum at the Sun core is more complex. To understand it we must consider the hydrogen fusion mechanism more closely: the "*pp* cycle" shown in Fig. 6a. There are two other branches, which contribute only marginally to the electromagnetic energy output, but do produce electron neutrinos: the neutrinos produced by the electron capture by ^7Be and by the ^8B decay, called "Beryllium neutrinos" and "Boron neutrinos" respectively. Their fluxes are low but their energies are much larger than those from the *pp* fusion. For the last reason they were observed historically first by the Chlorine (Homestake) and Cerenkov (Kamiokande) experiments. Fig. 6b shows the calculated electron neutrino flux at the Sun core. Its accurate and reliable knowledge is of fundamental importance not only for solar physics, but also for neutrino physics: flavour conversion phenomena, both in matter and in vacuum, are in fact detected as modifications of the spectrum. Fig. 6b includes the contribution of a second cycle of nuclear reactions, called CNO because it involves Carbon, Nitrogen and Oxygen. This is not very important for the Sun, but is relevant in other stars.

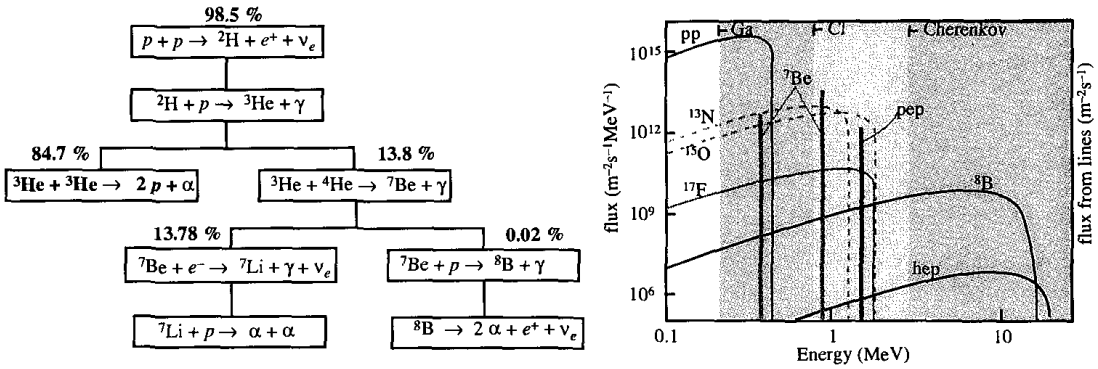


Fig. 6 a) The *pp* cycle. b) The calculated solar neutrino spectrum and its components

The results of the solar model calculations for all the components different from *pp* fusion are very sensitive both to the assumed solar parameters (notably the core temperature) and to the values of the cross sections of the nuclear reactions involved in the different branches of the *pp* cycle. Notice that the proton average kinetic energy in the Sun core, corresponding to their Maxwell-Boltzmann distribution at $T = 1.6 \times 10^7$ K, is 1.3 keV. In order to fuse, two protons - or two other nuclei - must tunnel through the Coulomb barrier. Its height is, depending on the reaction, between 1500 and 2500 keV, much larger than the average energy. The cross sections are proportional to the tunnel probability, which decreases exponentially with decreasing energy (the Boltzmann factor). The interaction probability, obtained by convolution of the Boltzmann energy distribution with the Coulomb penetration probability, is a bell-shaped curve, called Gamow peak. The maximum of the peak is typically at E_0=20 - 25 keV.

The cross sections fall dramatically with decreasing energy, by an order of magnitude in a few tens of keV or even faster, becoming extremely small at E_0, as shown on an example in Fig. 7. To give two other examples, $\sigma(E_0)$=0.7 pb for ${}^3He+{}^3He \rightarrow 2p+{}^4He$ and $\sigma(E_0)$=9 fb for ${}^4He+{}^3He \rightarrow \gamma+{}^7Be$. Taking into account that other cross sections are even smaller, one sees that femtobarn to sub-femtobarn sensitivities are needed to reach and cross the Gamow peak. This was not possible until recently and the cross sections at the relevant energies were obtained by extrapolation from the measured values at higher energies.

To formally, but not substantially, avoid in the extrapolation the rapidly changing exponential factor, the "astrophysical factor" $S(E)$ is defined as

$$S(E) = E\sigma(E)\exp\left(31.3 Z_1 Z_2 \sqrt{\frac{\mu}{E}}\right)$$

where μ is the reduced mass in units of amu, Z_1 and Z_2 are the charges of the nuclei and E is the centre of mass energy in keV. The extrapolation to E_0 is then done assuming a smooth behaviour of S. This procedure is clearly very risky in all cases and completely unreliable in presence of one or more resonances above or below threshold.

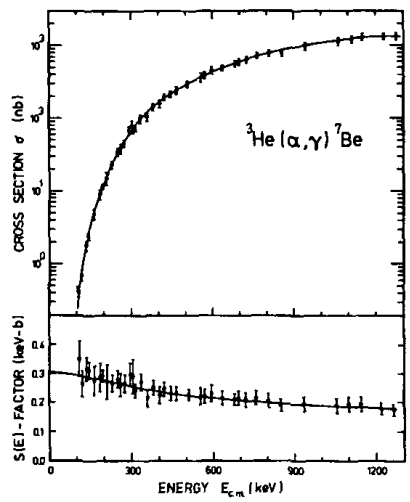


Fig. 7. A cross section and its astrophysical S factor

The experimental breakthrough came with the LUNA project at LNGS. Taking into account that at the sea level 2000 cosmic rays cross a 10 cm^2 detector in a day and that the counting rates at the Gamow peak are of the order of a few per month, it becomes clear that these measurements must be done in an underground laboratory. This is why the LUNA project was started at LNGS, the first worldwide underground accelerator facility, with the mission to provide a direct measurement of the most relevant fusion reactions of astrophysical interest.

^{3}He+^{3}He→2p+^{4}He was first reaction measured by LUNA with a 50 kV, 500 μA ion accelerator facility. Fig. 8a shows the LUNA results down to 17 keV[24] (where the cross section is only 20 fb and the rate 2 events/month!) below the Gamow peak. Notice that before these measurements the hypothesis of a resonance close to the Gamow peak would have explained, increasing the rate of this process, the reduced Beryllium and Boron neutrinos fluxes observed by Homestake and Kamiokande. The non observation of a resonance by LUNA substantially eliminated the nuclear solution of the solar neutrino puzzle.

LUNA2 is the second-generation accelerator facility at LNGS: a 400 kV, 650 μA accelerator with beam energy resolution better than 70 eV and long-term stability of 10 eV. A BGO-4π-summing detector, consisting of six optically separated segments, each observed by two photomultipliers at either side is another important component of the facility. The gas target is located inside a bore-hole of the BGO detector. The high, 85%, efficiency allows to detect nearly all reaction products.

^{2}H+p→γ+^{3}He was the first reaction measured by LUNA2 in 2001 and the second one below the Gamow peak, with a cross section of 10 fb at the lowest energy[25]. The astrophysical factor is shown in Fig. 8b. The relatively easy measurement was used as a test for the new apparatus, but it is very interesting on its own, particularly for the evolution of proto stars.

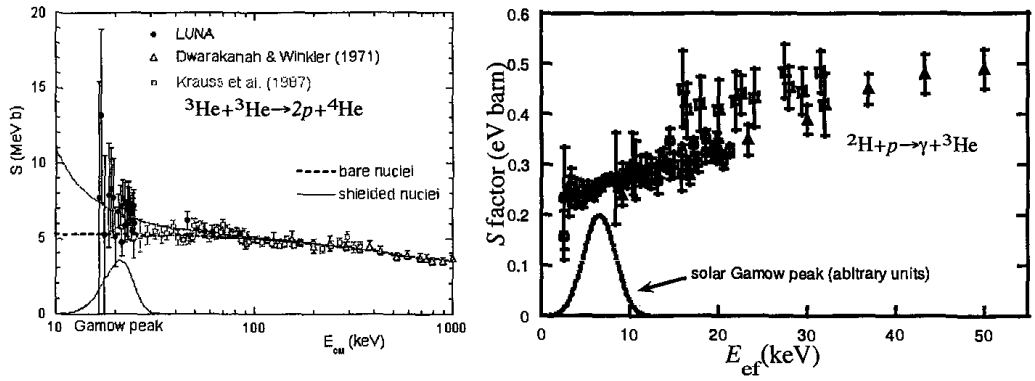


Fig. 8 Astrophysical factors measured by LUNA and LUNA2 at Gran Sasso for the reaction a) $^3He+^3He \rightarrow 2p+^4He$ b) $^2H+p \rightarrow \gamma+^3He$. Higher energy data from other experiments are also reported.

Presently the reaction $^{14}N+p \rightarrow \gamma+^{15}O$ is being measured. This reaction is the slowest in the CNO cycle, determining its speed with its low cross section and, as a consequence, its importance relative to the *pp* cycle in the solar neutrino spectrum. The extrapolation to the Gamow peak region is strongly affected by the presence of a sub-threshold resonance, giving a factor 10 uncertainty in the extrapolated values. Notice that the value of this cross section is of relevant astrophysical interest well beyond the solar physics, in view of its importance in the stellar evolution and on the evaluation of the age of the Universe. This is an example of the growing importance of underground nuclear astrophysics.

The next step will be the measurement in 2003 and 2004 of the $^3He+^4He \rightarrow \gamma+^7Be$ reaction, whose importance may be seen looking at Fig. 6a.

Complementary studies on the screening by the electron clouds are going on in parallel. They are very interesting, but will not described here.

9. Solar neutrinos experiments

We have already sketched the status of solar neutrino physics in §3. We stress here that the fit reported in Fig. 4 as the similar ones performed by others are based on the assumption that all the experiments (seven) are right and that all the uncertainties are correctly evaluated and correctly treated by the fitting procedures. Notice also that the solutions areas are usually defined at three standard deviations, not a large distance from the best fit for an assembly of many different experiments. These considerations point to the need in the future to have redundant information and, possibly, to prove oscillations and find the solution within a single experiment.

At LNGS GALLEX was completed in 1997. Its final result is 77.5±6.2(stat.)±4.5(syst.) SNU (one SNU is 10^{-36} captures per target nucleus per second). GNO, started in 1998, uses the

GALLEX structures with several improvements in the counters and in the electronics. It will continuously take data for several years, gradually reducing the overall uncertainty in the flux measurement. For some solutions seasonal variations might be observable. The results of the first two runs have been published[26]. The measured yield in 43 solar runs is 65.2±6.4(stat)±3.0(syst). Notice that systematic uncertainty is down to 4.6%. The weighted average of GALLEX and GNO is 70.8±4.5±3.8. For comparison, the solar model predictions range between 115 and 135 SNU.

Beryllium neutrinos flux appears to be particularly sensitive to neutrino oscillations parameters. The measurement of this mono-energetic, 0.86 MeV, flux in real time is the principal aim of the BOREXINO[27] experiment. Electrons resulting from a neutrino (any flavour, but ν_μ and ν_τ with smaller cross-sections than ν_e) scattering in the liquid scintillator detector medium will produce a light flash that will be detected by photomultipliers. 300 t of ultra-pure pseudocumene will be contained in a nylon sphere, the 100 t innermost mass being the sensitive volume. A larger volume of pseudocumene inside a 13.7 m diameter stainless steel sphere hosting the optical modules surrounds the nylon sphere. This sphere is immersed in a 2500 t purified water tank.

Taking into account the continuum energy spectrum of the recoiling electrons, the experiment is designed with a threshold of 0.25 MeV. The main problem at such low energies is the control of the background due to the always-present radioactive isotopes. An intense R&D program has been carried out in the last ten years to select materials and to purify them at unprecedented limits of radio-purity. The construction of the apparatuses is almost completed.

The neutrino yield from Be is extremely sensitive to the oscillation parameters. The yield is 40 events/day for the standard solar model. If δm^2 is sufficiently low, < 0.01 meV^2, strong seasonal variations are expected.

To find the solution in a single experiment, we need to measure in real time the low energy neutrino spectrum in order to separate the contributions of the different branches, *pp*, ^{7}Be and ^{8}B. The experiment should provide flavour sensitivity too, at least in combination with others. An intense R&D effort is going on in the last years on the LENS[28] proposal. Several targets have been investigated for ν_e capture (inverse beta decay), with a sufficiently low threshold to be sensitive to the *pp* spectrum. Initially, stable targets were chosen to avoid the huge background present for β decaying target nuclei, Se, Gd and, mainly, ^{176}Yb, whose ground states are stable against β decay.

The idea is to dope a liquid scintillator with metallorganic molecules containing the target nucleus to detect the electron resulting from the ν_e capture and, as a tag, the gamma (or gammas) resulting from the decay of the final nucleus from the excited state which is produced (Fig. 9a). The intense R&D led to the development of a high performance liquid scintillator containing a substantial concentration of Yb. On the other hand, the analysis cuts necessary due to the presence of small contaminations of radioactive nuclides resulted in severe losses of signal.

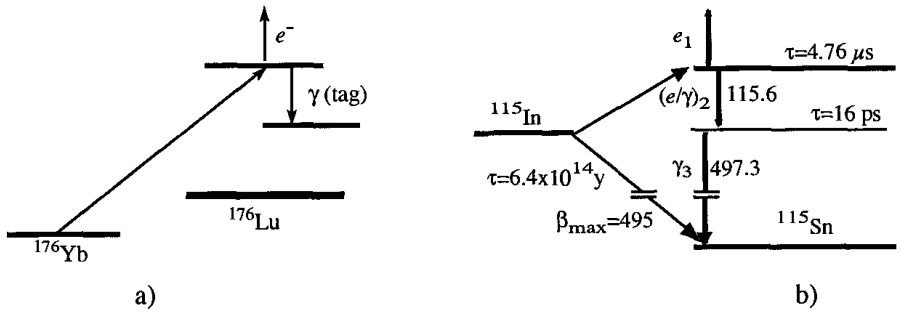


Fig. 9. Schematics of the energy levels involved in the electron neutrino capture by ^{176}Yb (a) and by ^{115}In (b)

The chemical technology developed for Yb liquid scintillator was found to be applicable to In. This target had been a candidate for *pp* ν_e detection since many decades, but it was abandoned in the eighties because of the light yield insufficient to cope with the strong background due to In beta decay. The new In liquid scintillator provides the superior energy resolution necessary to make In a viable solution for real time solar *pp* neutrino spectroscopy.

Fig. 9b shows the (simplified) level and tag structure of the ν_e capture in the In-Sn system. The ν_e energy threshold is as low as 118 keV. In liquid scintillator the prompt electron e_1 (616 keV) is detected. The transition is tagged by the delayed coincidence with the cascade $[(e/\gamma)_2 + \gamma_3]$ (116 + 497) keV with a tag time of 4.76 μs. Calculations show that about 1000 events per year can be detected in a 10 t In target. Backgrounds are on the other hand formidable, Indium β decay background being 11 orders of magnitude larger than the solar neutrinos signal. It can be reduced below the signal rate with a detector spatial granularity of the order of 10^5 cells, a not impossible number.

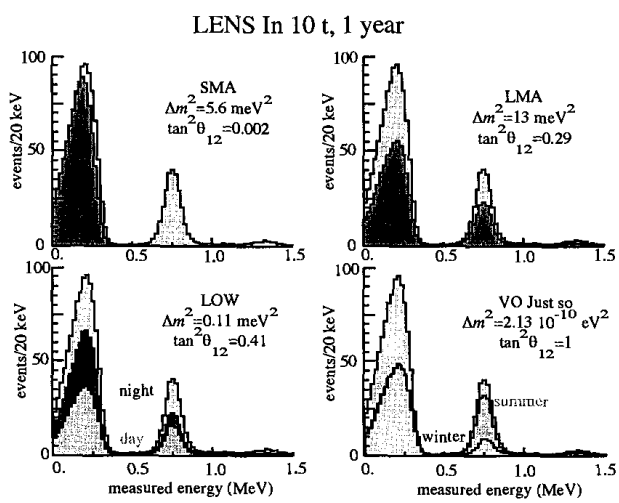


Fig. 10. Response of LENS, with 10 t In target to different solar oscillations scenarios

The expected yield has strong dependence on the mixing parameters allowing determining them accurately as shown in Fig. 10.

10. Neutrinos from Supernovae

The detection of the neutrinos from a core collapse Supernova and the measurement of their spectra and time evolution will provide important information on neutrino physics and on the implosion dynamics. A neutrino burst can be observed, with sub-megaton mass detectors, only if the explosion is in the Galaxy or in the Magellan Clouds.

It is very important to give an early warning to astronomers by neutrino detection. The Supernova watch network, SNEW, including Superkamiokande, LVD and SNO, has been set-up on purpose. KAMland, BOREXINO and Amanda, will soon join in. The gravitational wave detectors network should be integrated.

The explosion produces neutrinos and antineutrinos of the three flavours in a burst lasting 20-50 s. Non-zero masses ("m_ν") can lead to longer burst duration and one can think to have information on neutrino masses. But, what does "m_ν" mean? Neutrinos of different flavours produced in the Supernova core cross a medium of very high density in which important matter induced flavour conversions take place. Having left the star, the mass eigenstates propagate, independently one from the other, to our detector, possibly with slightly different velocities. Clearly the flux of a flavour we measure may be extremely different from that originally produced. As a consequence, the measurement of a delay between the arrival times, say, of a tau neutrino and of an electron neutrino cannot, even in principle, measure or limit, as is frequently but wrongly claimed, the mass of the tau neutrino.

On the other hand we can extract information on neutrino mixing and mass hierarchy. Two are the mechanisms that produce neutrinos in a Supernova: neutronisation and thermal emission. In the neutronisation process, electrons are captured by protons and nuclei and a ν_e flux results. This is the dominant flux in the first few milliseconds. Thermal emission follows, due to $e^+ e^-$ annihilation into neutrino-antineutrino pairs of all the flavours and ν_e and scattering, which follows. Clearly for each flavour the neutrino and antineutrino fluxes are equal. Notice also that muon and tau neutrino fluxes are identical because both are due only to neutral currents.

In the first phases of the Supernova expansion the density is so high that not even neutrinos can escape. Later, when the star becomes transparent, muon and tau neutrinos (and antineutrinos) that interact only via neutral currents, exit at higher temperatures than electron neutrinos and antineutrinos, which have charged current interactions too. As a final result the electron-neutrino (and anti-neutrino) spectrum is softer (average energy approximately 12 MeV) than those of the other flavours (average energy approximately 20 MeV).

The detected fluxes additionally depend on the values of the mixing parameters, mainly $|U_{e3}|^2$ and on the sign of Δm^2. I'll take as an example the case of $|U_{e3}|^2 >$ a few 10^{-4}, to guarantee adiabaticity

at the first level crossing. It can be shown that in this case, if $\Delta m^2>0$, the electron neutrino energy spectrum at the detector is equal to the original, harder, muon (and tau) neutrino spectrum. On the contrary it is the electron antineutrino spectrum that becomes equally harder if $\Delta m^2<0$. The measurement of the ratio of neutrino and antineutrino spectra, an almost model-independent distribution, might allow determining the sign of Δm^2.

At Gran Sasso the LVD is a dedicated experiment with 1080 t sensitive mass of organic liquid scintillator in a modular structure consisting of 912 tanks, each seen by three photomoltipliers. The tanks are read out independently, allowing a very high up time (99.7% during the year 2001). LVD is mainly sensitive to electron-antineutrinos through the process $\overline{\nu_e} + p \rightarrow n + e^+$ followed by the neutron capture $n + p \rightarrow d + \gamma + 2.2\,\mathrm{MeV}$, used as a tag with 60% efficiency. A few hundreds of events are expected for a SN explosion in the centre of the Galaxy (8.5 kpc). Their number and their energy spectrum can help in distinguishing amongst the different neutrino mass spectra.

LVD is sensitive to electron neutrinos through the charged current process $\nu_e + {}^{12}\mathrm{C} \rightarrow \mathrm{e}^- + {}^{12}\mathrm{N}$, followed by ${}^{12}N \rightarrow {}^{12}C + e^+ + \nu_e$. With an energy threshold of 17.3 MeV, the yield would be only a few events if the electron neutrino spectrum would be the cold original one. The process is particularly sensitive to flavour conversions because the induced hardening of the spectrum greatly changes the fraction of neutrinos above the energy threshold[29].

The situation is very similar for electron antineutrinos detected through the charged current process $\overline{\nu_e} + {}^{12}C \rightarrow e^+ + {}^{12}B$ followed by ${}^{12}B \rightarrow {}^{12}C + e^- + \overline{\nu_e}$, with an energy threshold of 14.4 MeV.

On the other hand the rate of the neutral current process $\nu_x + {}^{12}C \rightarrow \nu_x + {}^{12}C^*$ followed by ${}^{12}C^* \rightarrow {}^{12}C + \gamma$ with a similar threshold (15.1 MeV), as due to all neutrino flavours, is insensitive to oscillations.

In conclusion the ratio between charge currents and neutral currents yields may contribute further information to distinguish amongst the different neutrinos mass spectra. Very important will be the comparison of data obtained with different detectors, with their different sensitivity to energy and flavours.

11. Laboratory neutrinos. CNGS, OPERA and ICARUS

Experiments on atmospheric neutrinos have given evidence for muon neutrino disappearance that can be interpreted as an oscillation phenomenon at $\Delta m^2 \approx 2500\ \mathrm{meV}^2$ and maximum mixing, as recalled in §4. The non-observation of relevant effects in atmospheric electron neutrinos at the same flight times proves that the oscillation is not dominantly between muon and electron neutrinos. This conclusion is strengthened by the CHOOZ experiment. The most obvious conclusion is that muon neutrinos oscillate mainly into tau neutrinos. This fundamental conclusion must be directly confirmed observing the tau neutrino appearance in an originally pure, or almost pure, muon neutrino beam.

An artificially produced ν_μ beam of controlled characteristics is needed for these experiments. The small value of Δm^2 imposes a baseline of several hundred kilometres between beam and detector, which, for the first time must be located in different laboratories. Two types of experiments are possible on such a ν_μ beam, i.e. in disappearance and in appearance modes. A complementary program has been planned worldwide: K2K in Japan, already running, and MINOS in the US, expected to run in 2005, are disappearance experiments, while the CNGS project (CERN Neutrinos to Gran Sasso) is on ν_τ appearance and will run in 2006.

The pions and kaons produced by the 400 GeV SPS extracted proton beam will be focussed by a two-horn system followed by a 1 km long decay tunnel, a hadron stop and muon detectors for beam characteristics determination[30]. The resulting beam is almost pure in ν_μ's, with a small contamination of the other flavours, the most important being ν_e's, that is about 0.8%.

The optimum (first maximum) L/E value for observing ν_μ oscillations for Δm^2 = 2500 meV^2 is $(L/E)_{max} \approx$ 400 km/GeV. At a distance of 730 km (that from CERN to Gran Sasso and from Fermilab to Soudan), the energy for maximum oscillation is $E_{max} \approx$ 2 GeV. This is indeed the optimum energy for disappearance (and ν_e appearance) experiments, but not for ν_τ appearance ones, because in this last case tau leptons must be produced through the reaction

$$\nu_\tau + \mathrm{N} \rightarrow \tau^\pm + \mathrm{N}'.$$

Neutrino energy must be well above threshold ($\approx$ 3.5 GeV) for tau production, in practice larger than $\approx$ 10 GeV. Fig. 11 shows the neutrino fluence at Gran Sasso for the CERN beam compared with the product of the oscillation probability (decreasing with increasing energy) and the τ production cross section (increasing with energy). The beam spectrum is optimised for appearance at the Gran Sasso site.

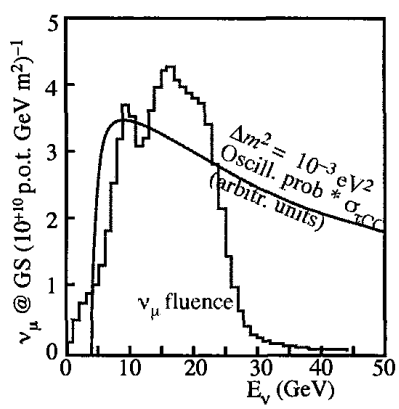


Fig. 11. Neutrino fluence at Gran Sasso for the CERN beam compared with the product of the oscillation probability and the τ production cross section for Δm^2 = 1000 meV^2.

The ν_τ appearance probability in an initially pure ν_μ beam of energy E is

$$P_{\nu_\mu \to \nu_\tau} = \sin^2(2\theta_{23})\cos^4(\theta_{13})\sin^2\left(1.27\Delta m^2(\mathrm{eV}^2)\frac{L(\mathrm{km})}{E(\mathrm{GeV})}\right)$$

As $\theta_{13} \approx 0$, we have approximately

$$P_{\nu_\mu \to \nu_\tau} = \sin^2(2\theta_{23})\sin^2\left(1.27\Delta m^2(\mathrm{eV}^2)\frac{L(\mathrm{km})}{E(\mathrm{GeV})}\right)$$

In this case the commonly used two-neutrino approximation is justified, but only as long as we forget the minority muon neutrino to electron neutrino oscillation.

Running in the "shared" mode, the beam will give 2630 CC ν_μ interactions per year in a kiloton fiducial mass detector at LNGS corresponding to 15 ν_τ interactions for Δm^2 = 2500 meV^2 and maximum mixing (multiply by 1.7 to have the yields in a "dedicated" running). A 50% gain in these yields is achievable with further optimisation at modest cost.

The charged daughters of τ's will be detected, in one or more decay channels: $\tau^- \to \mu^- \nu_\mu \nu_\tau$ (18%); $e^- \nu_e \nu_\tau$ (18%); $h^- \nu_\tau n\pi^0$ (50%); $2\pi^- \pi^+ \nu_\tau n\pi^0$ (14%). Two main background rejection tools are available: 1. the direct observation of τ decays requiring micrometer scale granularity and sub-micron resolution, which are possible only by the emulsion technique (OPERA); 2. the use of cinematic selection, which requires good particle identification and good resolution in momentum unbalance (ICARUS).

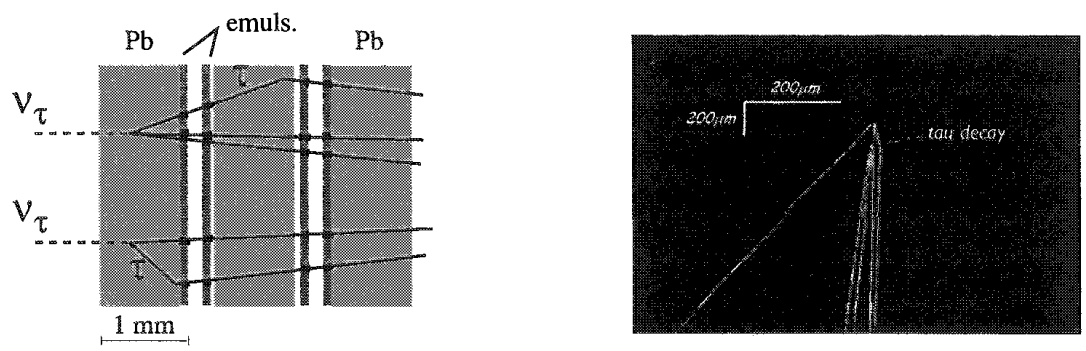


Fig. 12. a) The OPERA basic cell structure. Two types of events are shown and the principles for τ detection. b) A tau neutrino event from the DONUT experiment

OPERA[31] design is based on the ECC concept, which combines in one *cell* the high precision tracking capability of nuclear emulsion and the large target mass given by lead plates. Fig. 12a shows the cell structure (1 mm thick Pb plate followed by a film made by two emulsion layers 50

μm thick on either side of a 200 μm plastic base). Two topologies of τ events are shown: "long", where the τ production and decay are separated by at least one film (detected via the decay angle), and "short", where they are not (detected via impact parameter). The same technique has been used by the DONUT experiment, which discovered the tau neutrino at Fermi Lab. A tau neutrino interaction from this experiment is shown in Fig. 12b.

Several cells are sandwiched in a structure called a *brick*, the basic element used to build vertical planar structures called *walls*. A wall followed by a set of tracker planes makes up a *module*. A *supermodule* is made of a sequence of modules, and a downstream *muon spectrometer*. Overall two supermodules are foreseen, providing a sensitive mass of about 2 kt.

The target trackers, made of scintillator strips, are used to identify the fired brick, for general tracking and for shower energy. A muon spectrometer is made of a dipole magnet (1.5 T) preceded and followed by precision trackers (drift tubes) and low resolution tracking planes (RPC) embedded in the magnet. It provides measurement of muon momentum and charge (to reduce charm background) and beam characteristics. Fired bricks will be removed and processed on a day by day basis.

In a five year run, 16 identified τ decays with 0.6 background events are expected for $\Delta m^2 = 2500$ meV^2, 40 for 4000 meV^2 and 6.6 for 1300 meV^2.

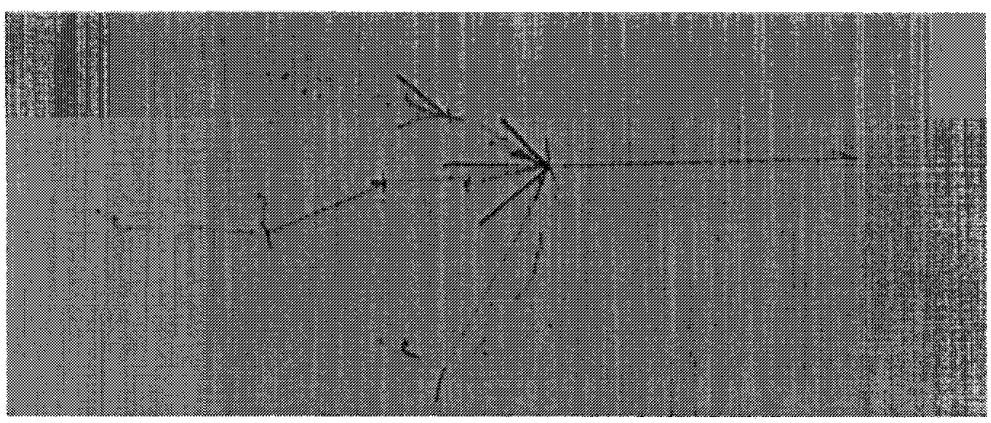


Fig. 13 A cosmic ray event in the first 300t ICARUS module

ICARUS[32] is a liquid argon time projection chamber providing bubble chamber quality 3D images of the events, continuous sensitivity, self-triggering capability, high granularity calorimetry and *dE/dx* measurement. The R&D program performed between 1991 and '95 on a 3 t detector solved the major technical problems with the detector continuously running for several years, showing the reliability of the technique. The technique was then developed for the industrial production of a kiloton size detector. Its structure will be modular. A module has a mass of 600 t (T600) and is composed of two 300 t units, transportable on highways. The units will be

completely assembled and tested before being separately transported to Gran Sasso. The first 300 t unit has been completed and successfully operated in summer 2001. Fig. 13 shows the superior quality and the richness of information provided by the technique. The safety issues connected with the installation of a large cryogenic volume underground are now under study. In its last version, T3000, the ICARUS project foresees the construction of a series of five T600's to cover a broad physics program, including ν_τ and ν_e appearance in the CNGS program.

The electron channel will be the "golden" one for the ICARUS search for τ appearance, mainly looking for an excess at low electron energies. There are two principal sources of background in this channel. The first is due to the π^0's produced by muon neutrinos of the beam via neutral currents, producing a shower similar to those of the electrons from tau decay. The superior e/π^0 separation capability of the detector will be exploited to cut out the largest fraction of these events. The second background is due to electrons produced via charged currents by the small (0.8%), but important, ν_e component of the beam. These electrons have a harder energy spectrum and different kinematics in the plane transverse to the beam compared to those of the electrons from tau decays. The definition of a suitable likelihood function, including electron energy and transverse plane kinematical variables, allows the separation on statistical basis of the tau neutrino appearance signal from the background. The proposed 3000 t configuration in five years running will collect 9 τ events for $\Delta m^2 = 2500$ meV^2 with a residual background of 0.6 events. The experiment will give also 3 τ events in the hadronic channel with a background of less than 0.1 events.

The excellent electron identification of the detector allows searching for ν_e appearance. Its oscillation amplitude is $P_{\max}(\nu_\mu \to \nu_e) = \sin^2(2\theta_{13})\sin^2(\theta_{23}) \approx 4\theta_{13}^2 \frac{1}{2} = 2\,|U_{e3}|^2$. It is proportional to $|U_{e3}|^2 \approx \theta_{13}^2$, hence very small. Contrary to CHOOZ, presently providing the best limit, the *L/E* values of CNGS are far from the oscillation maximum. Optimum neutrino energy for ν_e appearance is 2-3 GeV. These energies are too low for ν_τ appearance at which the CNGS beam is tuned. Moreover, as can be seen from the above expression, for $\theta_{23} \approx 45°$ an appearance experiment, in the same conditions, is twice less sensitive than a disappearance one. This can be understood from the fact that in disappearance the non-identified final state is half ν_μ half ν_τ, in appearance one starts from ν_μ only. In a eight-year run, at the CHOOZ limit (θ_{13}=7°), for θ_{23}=45° and Δm^2=3500 meV^2 the signal can be discovered at 4σ; _fno effect is observed, for_Δm^2=2500 meV^2 the CHOOZ limit can be improved by a factor five.

12. Conclusions

Neutrino physics has entered a new age in the last years. The discovery of neutrino oscillations has shown that neutrinos have non-zero masses and that the leptonic flavour numbers are not conserved. New neutrino physics appears to be a route for the exploration of the extremely high energies. Neutrino masses are much lower than the quarks masses and neutrino mixing pattern is very different from that of quarks. The mass generation mechanism might well be different too.

The search for neutrino-less double beta decays has already reached sensitivity that might be close to allow discovery of Majorana mass. These results, which point to physics beyond the standard theory, have been obtained in underground low background laboratories. Gran Sasso has contributed, as I have discussed.

An extremely interesting future appears to be in front of us where revolutionary discoveries might become possible. We are planning experiments to measure the mass eigenstate mixing in the lepton sector, to measure the neutrino masses and to search for dark matter. In particular the space that is now available at Gran Sasso and the quality of its infrastructures have stimulated many interesting ideas and proposals. These are in different stages of research and development, of test and of preparation. Presumably not all of them will become a running experiment, but we have good chances that at least a few will, with a bit of fortune, produce in the next years outstanding results.

References

[1] A. Bettini. Physics beyond the Standard Model at the Gran Sasso Laboratory, *Uspekhi* **171** (2001) 977-1003 [English version *Physics-Uspekhi* **44** (2001) 931-954]

[2] A. Bettini. New Neutrino Physics, *Rivista del Nuovo Cimento*, **24**, n.11, 2002

[3] A. Bettini. Perspectives of Astroparticle Physics.; Proceedings of the III International Meeting on Fundamental Physics, IMFP2002 to be published in *Nucl. Phys. (Proc. Suppl.)*

[4] A. Bettini. *The Gran Sasso Laboratory 1979-1999.* INFN 1999

[5] A. Morales *Nucl. Phys.* (Proc. Suppl.) **B 110** (2002) 39; Proceedings of the III International Meeting on Fundamental Physics, IMFP2002 to be published in *Nucl. Phys. (Proc. Suppl.)*

[6] R. Bernabei et al. Phys. Lett. **B480** (2000) 23; EPJ **C18** (2000) 283; Phys. Lett. **B509** (2001) 197, Eeur. Phys. J. **C23** (2002) 61; hep-ph/0203242

[7] C. Bucci et al. "Update of the Proposal to the LNGS for a Second Phase of the CRESST Dark Matter Search". LNGS Proposal P24/2001, approved as CRESST2; MPI preprint MPI-PhE/2001-02

[8] L. Wolfenstein, *Phys. Rev.* **D17**, (1978) 2369; ibid **D20**, (1979) 2634; S. P. Mikheyev and A. Yu. Smirnov, *Yad. Fiz.* **42**, (1985) 1441 [*Sov. J. of Nucl.* Phys. **42**, (1985) 913]; *Nuovo Cimento* **C9**, (1986) 17

[9] T. Kajita and Y. Totzuka, *Rev. Mod. Phys.* **73** (2001) 85

[10] Super-Kamiokande Collaboration; Y. Fukuda et al. *Phys. Rev. Lett.* **81** (1998) 1562; NEUTRINO 2000. *Nucl. Phys.* **B 91** (2001) 127; Y. Totzuka "Atmospheric Neutrinos" at TAUP 2001, Gran Sasso National Laboratory, Sept. 10 2001

[11] MACRO Collaboration at NEUTRINO 2000. *Nucl. Phys.* **B 91** (2001) 141

[12] CHOOZ Collaboration. M. Apollonio et al. *Phys. Lett.* **B466**, (1999) 415

[13] T. A. Kirsten. *Rev. Mod. Phys.* **71** (1999) 1213

[14] G. L. Fogli, E. Lisi, A. Marrone, D. Montanino and A. Palazzo hep-ph/0206162, 2002

[15] M. Colles et al. MNRAS **328** (2001) 1039; V. J. Percival et al. MNRAS **327** (2001) 1297

[16] Ø- Elgaroy et al. astro-ph/0204152 v3, 2002

[17] A. Lewis and S. Bridle astro-ph/0205436 v2, 2002

[18] C. Weinheimer et al. *Phys. Lett.* (1999) **B460** 219; NEUTRINO 2000, *Nucl. Phys.* B (Proc. Suppl.) **91** (2001) 273

V. M. Lobashev et al. *Phys. Lett.* **B460** (1999) 219; NEUTRINO 2000, *Nucl. Phys.* B (Proc. Suppl.) **91** (2001) 280

[19] H.V. Klapdor-Kleingrothaus, "Sixty Years of Double Beta Decay – From Nuclear Physics to Beyond Standard Model Particle Physics", World Scientific, 2000

[20] LNGS-LOI 9/97; LNGS-LOI 9/97 add. 1; LNGS P23/2000 GENIUS MPI-Report MPI-H-V26-1999

[21] MPI-Report MPI-H-V4-2001; LNGS P27/2001

[22]] LNGS-LOI 16/99; LNGS-LOI 12/98

[23] A. Alessandrello et al: *Physics Letters* **B486**, (2000) 13

[24] R. Bonetti et al.; *Phys. Rev. Lett.* **82**, 26, (1999), 5205-5208

[25] C. Casella et al. *Nucl. Phys.* **A706** (2002) 203-216

[26] M. Altmann et al., *Phys. Lett.* **B490** (2000) 16; C. Cattadori, for the GNO Collaboration. Presented at TAUP 2001, September 2001

[27] C. Arpesella et al. BOREXINO proposal (Univ. of Milan) 1991; F. v. Feilitzsch et al. *Astro Phys.* **8** (1998) 141; G. Alimonti et al. (BOREXINO Collaboration) hep-ex/0012030.

[28] LNGS-LOI 18/99 and add.1/00, add. 2/01, add. 3/01, add. 4/02

[29] M. Aglietta et al., astro-ph/0112312 and presented at TAUP 2001

[30] CERN 98-02,INFN/AE-98/05 and CERN-SL/99-034(DI),INFN/AE-99/05

[31] OPERA proposal. CERN/SPSC 2000-028; SPSC/P318; LNGS P25/2000. July 10, 2000

[32] ICARUS T3000 proposal LNGS-EXP 13/89 add. 2/01; ICARUS –TM/2001-09; 2001

CHAIRMAN: A. BETTINI

Scientific Secretaries: P. Bechtle, A. Polouektov

DISCUSSION

- *Levell:*

As far as I understand, another experiment has searched in a region of the parameter space similar to that of DAMA, but has not confirmed the DAMA result. I think you said in your talk a bit earlier, that comparison with other experiments is model-dependent. Could you elaborate?

- *Bettini:*

DAMA, and other similar experiments, look for WIMPs (Weakly Interacting Massive Particles). A WIMP may, very rarely, elastically collide a nucleus of the detector; the recoiling nucleus will then release its energy to the detector. This energy can be detected as light, ionisation or increase in temperature. The detecting material of DAMA are NaI crystals, about 100 kg in mass, where the signal is a tiny flash of light. The energy deposit is indeed small, a few keV and the energy spectrum is a decreasing function of the energy. The signal is very similar to background. To distinguish the two, DAMA looks for the time variation of the rate. Think of a WIMP cloud filling our Galaxy. The solar system sails through this cloud at about 270 km/s and the Earth revolves about the Sun at 30 km/s. In June the two speeds are about in the same direction, in December the opposite. As a consequence, the signal rate should be modulated with a 12 months period and the above phase; the amplitude of the modulation should be several percent. The background, or at least many types of background, should not have these characteristics. DAMA observes exactly the expected modulation. This observation is model-independent evidence for WIMP. It might be falsified by another experiment, but only in a model-independent way. The interpretation of the signal as a region in the parameter space is instead model-dependent. No other detector has enough sensitivity to search for the modulation. How do they cope with background? There are two main backgrounds. The dominant one is due to gamma rays, the second to neutrons. The two can be distinguished by measuring two quantities, heat and light (CRESST and Edelweiss) or ionisation and energy deposit (CDMS). The gamma background can be suppressed in this way. The neutrons on the other hand produce nuclear recoils similarly to WIMPs and cannot be distinguished. These experiments in practice produce upper limits on the WIMP flux. These limits depend on the assumed WIMP distribution on one side and on the detector characteristics, in particular on its isotopes, on the other.

- *Bechtle:*

You mentioned the Heidelberg-Moscow-Experiment, but you said nothing about this maybe statistically questionable report on this neutrinoless double-beta-decay. What is the status of that? Has this been withdrawn or is there still discussion about it and is there the possibility to check this result with other experiments?

- *Haidt:*

I have looked at that and I disbelieve it. There are very simple reasons for this: He [Klapdor-Kleingrothaus et al] has made a classification in bins, where the average signal per bin is about 8 events. There are of course fluctuations upwards and downwards. A fit consisting of the hypothesis 'a flat distribution+a gaussian with fixed width' was applied to the data. This fit treats fluctuations upwards and downwards rather asymmetrically. The fluctuations give rise to random peaks. The task of interpreting these peaks has not been performed convincingly.

- *Bettini:*

The energy spectrum is almost flat in first approximation, but at closer sight you can see up and downs not very significant statistically. If you insist these are physical lines and search for them, the algorithm finds the four Bi lines which, indeed, should be there. The algorithm finds a line at the position of the neutrinoless double beta decay too. But there are two more lines not corresponding to known isotopes. The results depend on the width used by the searching algorithm too.

- *Haidt:*

The assumed beta-beta signal is isolated by cutting out the two peaks at the left and the right of it without justification. The fit procedure then reduces the contribution to the flat part and consequently enhances the signal, thus pretending a large effect.

- *Wendland:*

Could I ask a follow-up there? Has someone applied Bayesian statistics to Klapdors neutrinoless double-beta-decay data? Did he release the data for the people to fit? Is any other experiment capable to check the result?

- *Bettini:*

The HM data are published and available to everybody. They have been critically reanalysed, to my knowledge, by Feruglio, Strumia and Vissani (hep-ph/0201291), which conclude that there is no really significant evidence for a $0\nu 2\beta$ signal. On the other hand, HM is the most sensitive experiment existing. They have 10 kg of enriched Ge, very pure, run for several years. Another experiment, IGEX, has almost the same sensitivity again based on Ge, but with a little bit less exposure. In order to compare with experiments using different isotopes, as those with tellurium, the closer

competitor, one must take into account the uncertainty in the nuclear matrix elements too.

- *Haidt:*
  By the way, what is the claim of the Edelweiss collaboration?

- *Bettini:*
  Edelweiss is a mainly French collaboration, working in the Frejus underground Laboratory, searching for dark matter. Their exclusion limit is the most stringent one in a model-dependent, comparison amongst experiments

- *Sellitto:*
  If ν_1, ν_2 and ν_3 are not the mass eigenstates in the MSW effect, which are then the mass eigenstates?

- *Bettini:*
  ν_1, ν_2 and ν_3 are the mass eigenstates in a vacuum, not in matter. To find the eigenstates you must add a matter dependent term (the potential) in the mass-matrix and then diagonalise it. The potential depends on the electron density. As a consequence, if neutrinos cross a variable density medium, a level crossing may happen, at a critical density, inducing a flavour changing transition.

- *Maas:*
  Are any of the Gran Sasso experiments able to test the LSND measurements?

- *Bettini:*
  Substantially not, but a small fraction of the area of the LSND claim might be in principle tested in the CNGS programme. But MiniBoone will already have given the answer by that time.

- *Haidt:*
  You have mentioned that you have also an experiment, which measures the sign of the squared mass difference.

- *Bettini:*
  The sign of the mass difference might be, with some luck, determined with supernova neutrinos by LVD. We are deeply studying this issue.

  Another possibility for underground laboratories is a next generation experiment on atmospheric neutrinos. The main idea comes from MONOLITH, an experimental proposal for Gran Sasso which, unfortunately, did not gather a strong enough collaboration. The idea is to measure the muon neutrino flux as a function of L/E, the variable of the oscillation function. L and E are the flight length and the energy of the neutrino. They cannot be measured, but can be evaluated from the direction and

energy, or momentum, of the muon for high enough (>2 GeV or so) energy. MONOLITH is an iron tracking calorimeter made of magnetised Fe slabs interleaved by tracking (and timing) elements. To observe the oscillation pattern, as originally proposed, 30-40 kt mass is needed, to be sensitive, through the regeneration effect in the Earth, to the sign of Δm^2 at least 100 kt.

- *Cata:*

You mentioned the possibility of detecting proton decay. How do you expect to get signals of proton decay?

- *Bettini:*

This should really be a question to Professor Totsuka. Let me recall that the order of magnitude of neutrinos masses indicates that the p-decay lifetime might be not too far from the SuperK sensitivity. The way appears to be to build a water Cherenkov an order of magnitude larger (HyperK). The dominating background is atmospheric neutrinos. Studies made by Professor Totsuka's group show that the Cherenkov technique allows to well separate signal from background for megaton-size detectors and even larger. Notice also that, contrary to a common belief, the detector does not need to be very deep; atmospheric neutrinos flux is depth-independent. So we can hope to go up an order of magnitude or more in the lifetime and discover the effect.

- *Golovnev:*

There was a project of sending a neutrino beam from CERN to Gran Sasso. What is the status, does it depends on the LHC schedule?

- *Bettini:*

Indeed there is, not was, such a project as I described this morning. The whole CERN program has been reviewed, due to the known financial problems, but the CNGS program has survived the review and will continue. The program has progressed according to its schedule and, as a consequence, can be ready in mid 2005 as originally foreseen. But in 2005 the SPS machine will not run for financial reasons and CNGS is being rescheduled to start in 2006. This is independent of the LHC schedule.

- *Stamen:*

You mentioned the KATRIN experiment. Are there any ideas to measure directly the muon and tau neutrino masses?

- *Bettini:*

Electron, muon and tau neutrinos are linear combinations of ν_1, ν_2 and ν_3. A limit on the "electron neutrino mass" is effectively a limit on the other ones, of the same order of magnitude. No "direct" measurement of muon or tau neutrino mass can reach that sensitivity in any foreseeable future.

- *Stamen:*

There are no current direct measurements?

- *Bettini:*

Since the measurement of the electron neutrino mass is an indirect measurement of the real masses, those of ν_1, ν_2 and ν_3, in my view there is no reason to go through something which is far less sensitive, like muon and tau neutrino masses.

- *Bechtle:*

You mentioned the LENS proposal. Could you be a bit more explicit about the experimental techniques which are going to be used there?

- *Bettini:*

The LENS Collaboration has initially considered three target options, three isotopes of Gd, Se and Yb. All of them are stable against beta decay and suitable to detect low energy (*pp*) neutrinos. The basic challenge was the development of a stable liquid scintillator, doped with one of these elements with substantial concentration, with a high enough light yield and long attenuation length for the light. It soon became clear that the only manageable candidate was Yb. It became later clear that non-eliminable background sources in Yb would require strong cuts, so reducing considerably the effective yield. On the other hand, the chemical technology developed for Yb liquid scintillator was found to be fully applicable to another metal suitable for low energy electron neutrinos detection, namely In. This target had been studied for many years due to its very attractive features, but abandoned for the impossibility to cope with the enormous background due to the beta decay of In itself. The superior light luminosity obtainable with the new LENS technology appears to make the suppression of this severe background possible. Indium is now the main target option of LENS. The background rate is eleven orders of magnitudes higher than the solar neutrinos signal. Background signals are due to small chains of events and, as a consequence, are not localised in a single point. This feature gives a second handle, namely the granularity of the detector. It appears that of the order of 10^5 cells are needed, a manageable number.

In order to test prototypes of the detector an advanced test laboratory - the LENS low background facility (LLBF) - has been built at Gran Sasso.

- *Totsuka:*

Suppose that KamLAND finds neutrino oscillations consistent with the LMA solution, is it still valuable to carry out solar neutrino experiments that measure *pp* neutrinos?

- *Bettini:*

We are developing R&D now for LENS, and we are progressing. Before making a final decision, I think the results from KamLAND, Borexino and SNO must be

considered. On the other hand, *pp* neutrinos come from the basic reaction responsible for energy generation in the Sun and, as such, are interesting by themselves.

- *Haidt:*
  May I ask what the status of Borexino is?

- *Bettini:*
  The last important piece of Borexino to be delivered was the Nylon vessel, a very delicate component. It reached the laboratory a few weeks ago. The schedule was to start filling in a couple of months from now. Unfortunately, manoeuvres not foreseen by the procedures led to a spill of pseudocumene in the environment. This accident did not lead to relevant environmental consequences, but to very severe political ones. We are in the process of reviewing all the procedures to see how exactly the system failed, to avoid that a similar accident happens again.

- *Bechtle:*
  You told us about the emulsion experiment OPERA. Could you tell us about how you scan the incredible amount of emulsions? Could one also include the scan mechanism in the experiment?

- *Bettini:*
  The scanning is certainly a very important point, which I skipped due to the time. It is an enormous amount. The experiment has two phases. The data taking on the beam is the first phase. The emulsion-lead sandwiches, the basic cells of the experiment, are assembled in units, called "bricks". Bricks are used to build vertical planes, called walls. The walls are interleaved with electronic trackers. When a neutrino, normally muon neutrino, rarely tau neutrino, interacts in a wall, the trackers are fired. The detected tracks are followed back in real time to identify the fired brick. Fired bricks are extracted on a day-to-day basis (20-30 bricks) and processed. The second data taking phase is the extraction of the track elements from the emulsion. The scanning and the tracking from emulsion sheet to emulsion sheet are performed automatically using highly sophisticated electronic and mechanical instruments, developed by the collaboration (mainly the Nagoya group). Similar instruments have already been used in the CHORUS and DONUT experiments; the performance, scanning speed, must be improved by an order of magnitude. The collaboration is working on this problem.

Experimental Highlights from Super-Kamiokande

Erice-2002
Yoji Totsuka
Kamioka, ICRR

- Introduction
- Neutrino Masses and Mixings
- Neutrino Oscillation
- Observations
  - Solar Neutrinos
  - Atmospheric Neutrinos
  - Accelerator Neutrinos (K2K experiment)
- Summary of Results

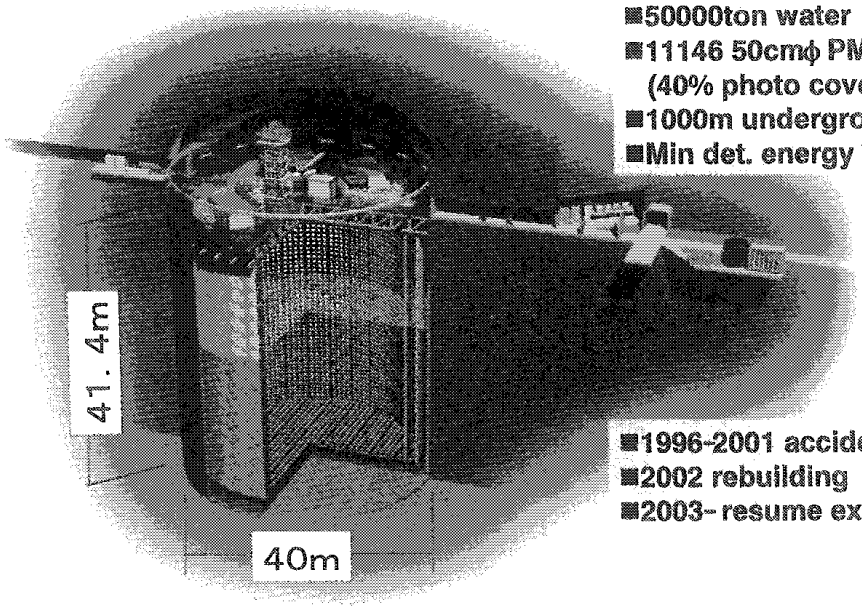
1

Introduction

- Super-K detector
- SNO detector
- Neutrino masses and mixings
- Neutrino oscillations
- Relation to cosmology

2

Super-Kamiokande (1996)

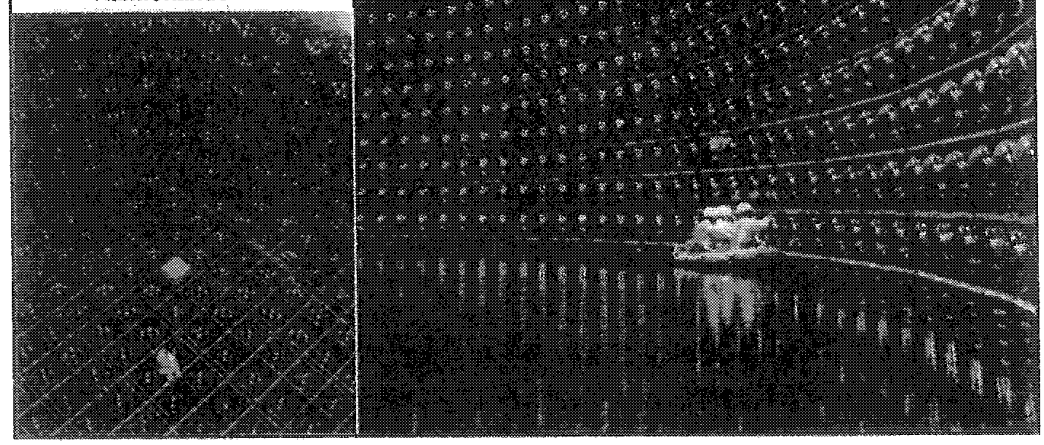


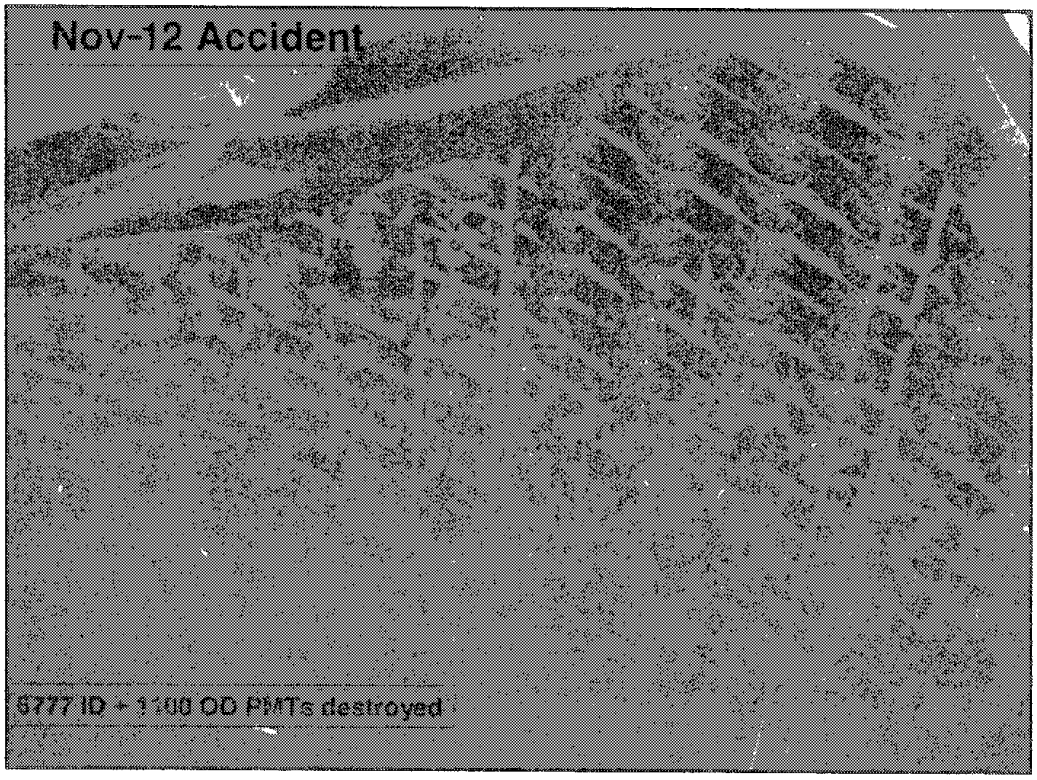
- 1996-
- 50000ton water
- 11146 50cmϕ PMT (40% photo coverage)
- 1000m underground
- Min det. energy ~ 5 MeV

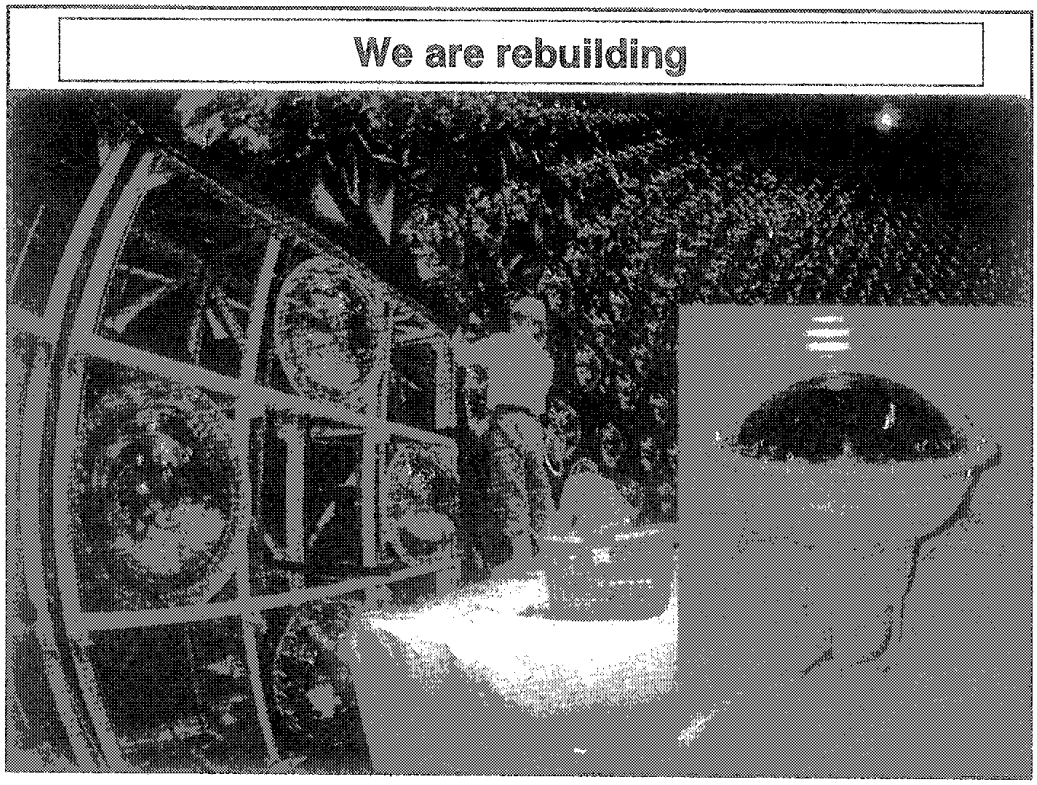
- 1996-2001 accident
- 2002 rebuilding
- 2003- resume experiment

3

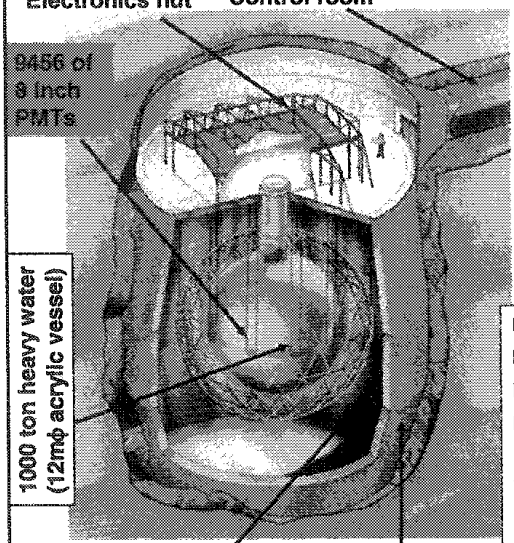
Inside Super-K



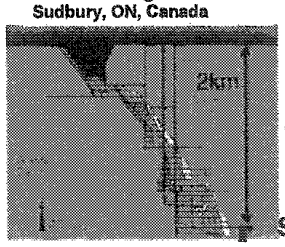

Nov-12 Accident
5777 ID + 1100 OD PMTs destroyed


We are rebuilding

Sudbury Neutrino Observatory

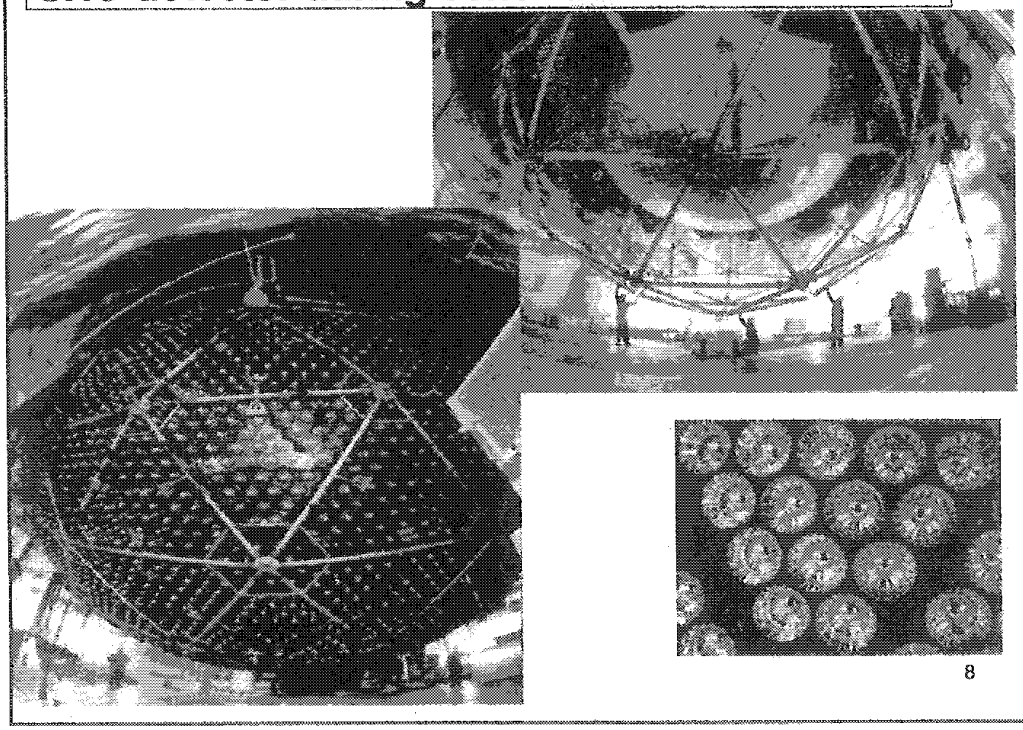


INCO's Creighnton mine
Sudbury, ON, Canada



- photo coverage 55% (R<7m)
- cosmic ray muons ~70 events/day
- fiducial volume 0.7kt (R,5.5m)
- Trigger rate (data) 6~8Hz (~2MeV threshold)
- Trigger efficiency 100%@~3MeV
- From compton e- from ^{16}N source (~5 MeV)
  - vertex resolution 16cm
  - energy resolution 16%
  - angular resolution 27°C

SNO detector during construction



Neutrino mixing for $m_\nu \neq 0$

Weak eigenstates

Mass eigenstates

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

U (Maki-Nakagawa-Sakata Matrix) =

$$\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23}-c_{12}s_{13}s_{23}e^{-i\delta} & c_{12}c_{23}-s_{12}s_{13}s_{23}e^{-i\delta} & c_{13}s_{23} \\ s_{12}s_{23}-c_{12}s_{13}c_{23}e^{-i\delta} & -c_{12}s_{23}-s_{12}s_{13}c_{23}e^{-i\delta} & c_{13}s_{23} \end{pmatrix}$$

$s_{ij}=\sin\theta_{ij}$, $c_{ij}=\cos\theta_{ij}$

Non-zero $m_{\nu l}$ and θ_{ij} will induce neutrino oscillations

9

Neutrino oscillation

- Weak eigenstates ν_e, ν_μ, ν_τ and mass eigenstates ν_1, ν_2, ν_3 are in general not identical.
- Connection between the two:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

- The Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = (p^2+(m_1^2+m_2^2+m_3^2)/2)^{1/2}\, U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

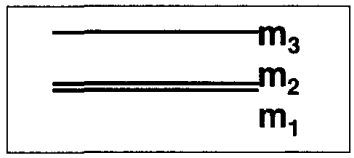
$M = \mathrm{diag}(m_1, m_2, m_3)$

Non-diagonal term appears in the right hand side.

10

cont-1 (ν_e beam)

ν_e - - - - - - - - - - - - ▸ $\nu_{\mu,\tau}$
L



$m_3 - m_2 >> m_2 - m_1$

- $P(\nu_e \rightarrow \nu_{\mu,\tau}) = 0.5\sin^2(2\theta_{13}) + \cos^4\theta_{13}\sin^2(2\theta_{12})\sin^2(\Delta m_{12}^2 L/4E_\nu)$
  $\approx \sin^2(2\theta_{12})\sin^2(\Delta m_{12}^2 L/4E_\nu)$

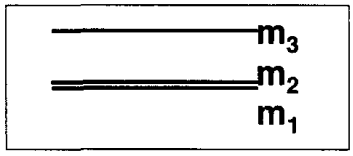
$\nu_{\mu,\tau}$ appearance

- $\Delta m_{12}^2 = m_2^2 - m_1^2$
- L: flight distance ($L >> 4E_\nu/\Delta m_{23}^2$)
- Matter effect must be taken into account
- $\Delta m_{12}^2 L/4E_\nu = 1.27\Delta m_{12}^2(eV^2)L(km)/4E_\nu(GeV)$
- E: neutrino energy
- $\sin^2(2\theta_{12})$, θ_{12} = mixing angle
- $\sin^2(2\theta_{13}) < 0.1$ (from the Chooz experiment)

11

cont-2 (ν_μ beam)

ν_μ - - - - - - - - - - - - ▸ $\nu_{\tau,e}$
L



$m_3 - m_2 >> m_2 - m_1$

$$P(\nu_\mu \rightarrow \nu_{\tau,e}) = \cos^4\theta_{13}\sin^2 2\theta_{23}\sin^2(\Delta m_{23}^2 L/4E_\nu) - \sin^2 2\theta_{13}\sin^2\theta_{23}\sin^2(\Delta m_{13}^2 L/4E_\nu)$$
$$\approx \sin^2 2\theta_{23}\sin^2(\Delta m^2_{23} L/4E_\nu)$$

ν_e appearance

ν_τ appearance

$L << 4E_\nu/\Delta m_{12}^2$
$\sin^2\theta_{13} < 0.026$, $\cos^4\theta_{13} \sim 1$ (from the CHOOZ experiment)
$\sin^2 2\theta_{\mu\tau} \sim \sin^2 2\theta_{23} \sim 1$,
$\sin^2 2\theta_{\mu e} = \sin^2 2\theta_{13}\sin^2\theta_{23} \sim 0.5\sin^2 2\theta_{13} \sim 2|U_{e3}|^2$

12

Matter effect (MSW)

- $E = p + \frac{m_\nu^2}{2p} + V$

$$p = \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} \qquad m = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \qquad V = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}_{flavor}$$

- Consider $\nu_e \rightarrow \nu_\mu$ oscillation.
- The matrix of the mass squared M^2 in the ν_e, ν_μ representation $(E = p + M^2/2p)_{flavor}$

$$M^2 = \frac{1}{2}(m_1^2 + m_2^2 + A)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} A - \Delta m^2\cos 2\theta & \Delta m^2 \sin 2\theta \\ \Delta m^2 \sin 2\theta & -A + \Delta m^2 \cos 2\theta \end{pmatrix}$$

- $A = 2E\Delta V$
- $V_{\nu e} - V_{\nu\mu} = \Delta V = 2^{1/2} G_F N Z = 2^{1/2} G_F N_e$ where N_e is the electron number density.

13

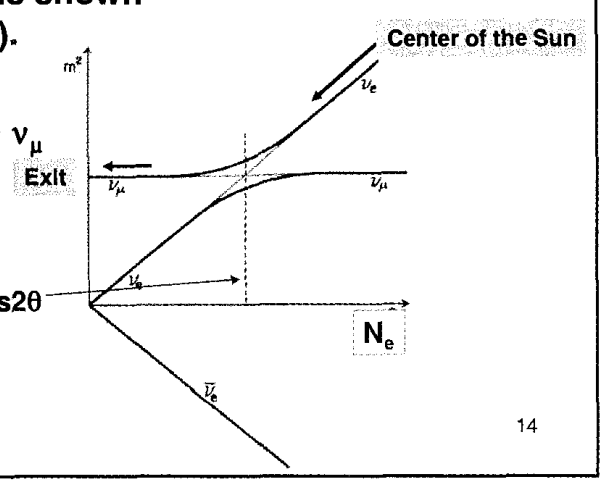
cont1

- The mass eigenvalues are

$$m_\nu^2 = \frac{1}{2}(m_1^2 + m_2^2 + A) \pm \frac{1}{2}[(\Delta m^2 \cos 2\theta - A)^2 + (\Delta m^2)^2 \sin^2 2\theta]^{1/2}$$

- For positive ΔV, m_ν^2 is shown in the figure (small θ).
- Full conversion $\nu_e \rightarrow \nu_\mu$ can take place.
- Resonant point

$A/\Delta m^2 = 2E_\nu \Delta V/\Delta m^2$
$= 2\sqrt{2} E_\nu G_F N_e/\Delta m^2 = \cos 2\theta$



14

cont2

- The Schrödinger equation:

$$i\frac{d}{dt}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} = [U(p^2+m_\nu^2)^{1/2}U^{-1} + \Delta V]\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix}$$

$$=2\pi\begin{bmatrix}\frac{1}{L_e} - \frac{\cos 2\theta}{L_\nu} & \frac{\sin 2\theta}{2L_\nu}\\ \frac{\sin 2\theta}{2L_\nu} & 0\end{bmatrix}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix}$$

- where $L_\nu = 4\pi E_\nu/\Delta m^2$, $L_e = 2\pi/\Delta V = 2^{1/2}/[G_F N_e(t)]$, $\Delta m^2 = m_2^2 - m_1^2$.
- If the density N(t) varies in time, the diagonal term could vanish where the resonant oscillation takes place.

15

Relation to cosmology

- Big Bang theory predicts
  - $n_\nu = 110\ /\mathrm{cm}^3\cdot\mathrm{flavor}$
  - $\Omega_\nu = \Sigma m_\nu n_\nu / \rho_c = 0.0105\ h_0^{-2}\ \Sigma m_{\nu,eV} = 0.021(1\pm0.2)\ \Sigma m_{\nu,eV}$
    - $H_0 = 72\pm7\ \mathrm{kms^{-1}Mpc^{-1}}$
    - $\rho_c = (3/8\pi)H_0^2/G_N$
- How large are $m_{\nu i}$ and hence Ω_ν? Are neutrinos hot dark matter?
- CP violation ($\delta \neq 0$) in the neutrino sector may induce matter anti-matter asymmetry in the Universe

16

Matter anti-matter asymmetry in the Universe

- Sakharov's 3 principles
  - An epoch of thermal non-equilibrium
  - Baryon number non-conservation
  - CP non-invariance
- ~~CP~~ in the quark sector is too small
- ~~CP~~ in the neutrino sector is the key?
- Look for the difference of oscillations between $\nu_\mu \rightarrow \nu_e$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$
  - Answer by 2015?

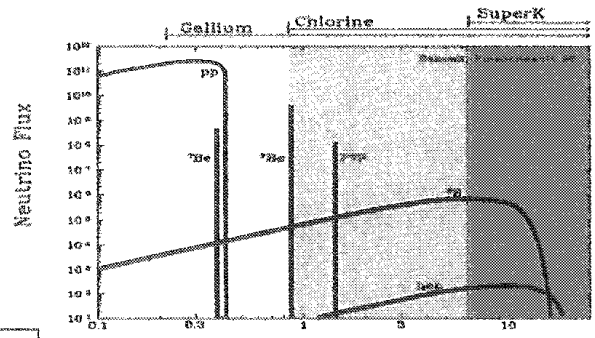
17

Observations

- Solar neutrinos
- Accelerator neutrinos (K2K experiment)
- Atmospheric neutrinos (brief)

18

Solar neutrinos



ν_e

pp-chain

://www.sns.ias.edu/~jnb/

$p + p \rightarrow {}^2H + e^+ + \nu_e$ (99.75%) $\quad p + e^- + p \rightarrow {}^2H + \nu_e$ (0.25%)

${}^2H + p \rightarrow {}^3He + \gamma$

${}^3He + {}^3He \rightarrow {}^4He + 2p$ (86%) $\quad {}^3He + {}^4He \rightarrow {}^7Be + \gamma$ (14%) $\quad {}^3He + p \rightarrow {}^4He + e^+ + \nu_e$

${}^7Be + e^- \rightarrow {}^7Li + \nu_e$ (99.85%) $\quad {}^7Be + p \rightarrow {}^8B + \gamma$ (0.15%)

${}^7Li + p \rightarrow 2\,{}^4He$ $\quad {}^8B \rightarrow {}^8Be^* + e^+ + \nu_e$

${}^8Be^* \rightarrow 2\,{}^4He$

19

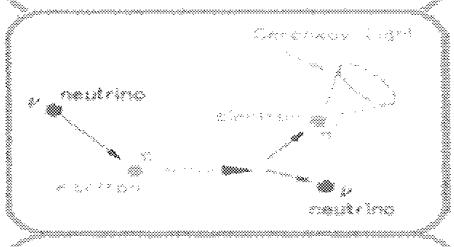
Detection of neutrinos by SNO and Super-K

$\nu_e + d \rightarrow e + p + p$ (CC)

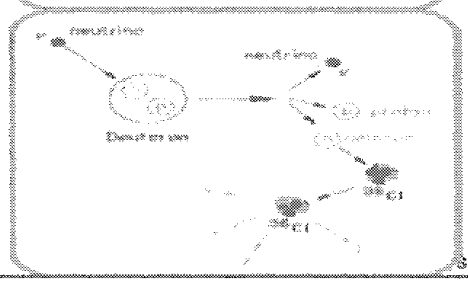
- measure $\phi(\nu_e)$
- $1 - \cos(\theta)/3$
- good ν_e energy resolution

$\nu_x + e \rightarrow \nu_x + e$ (ES)

- measure $\phi(\nu_e) + 0.154\phi(\nu_{\mu,\tau})$
- electrons emitted forward
- small cross section

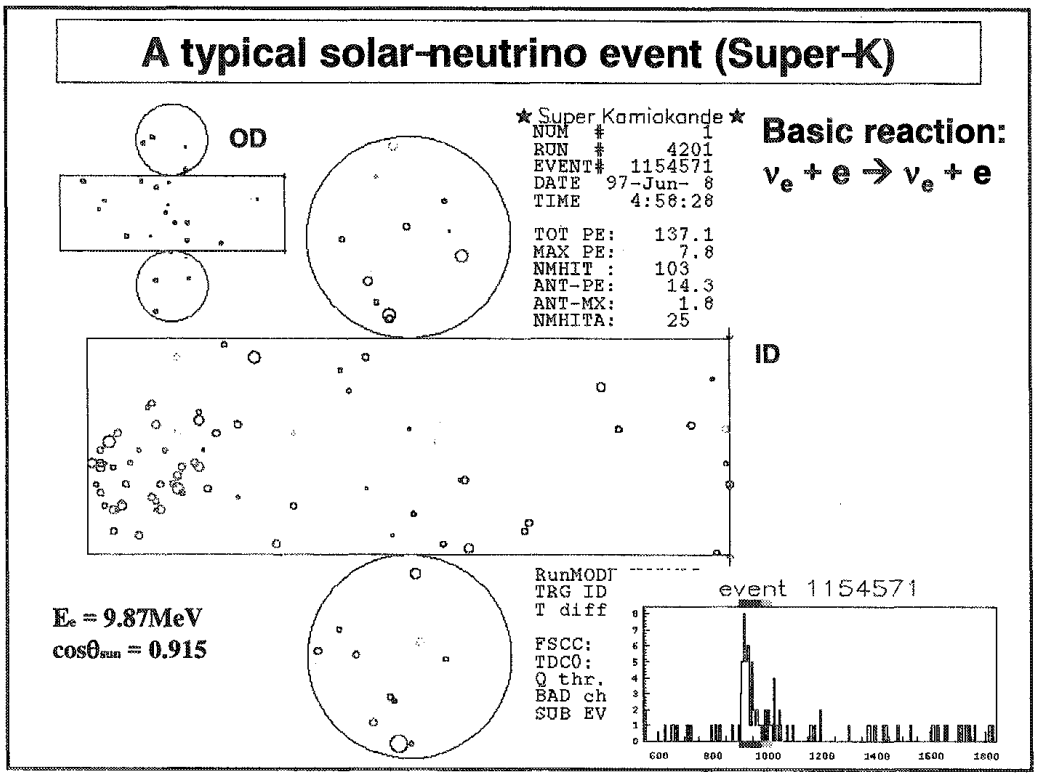


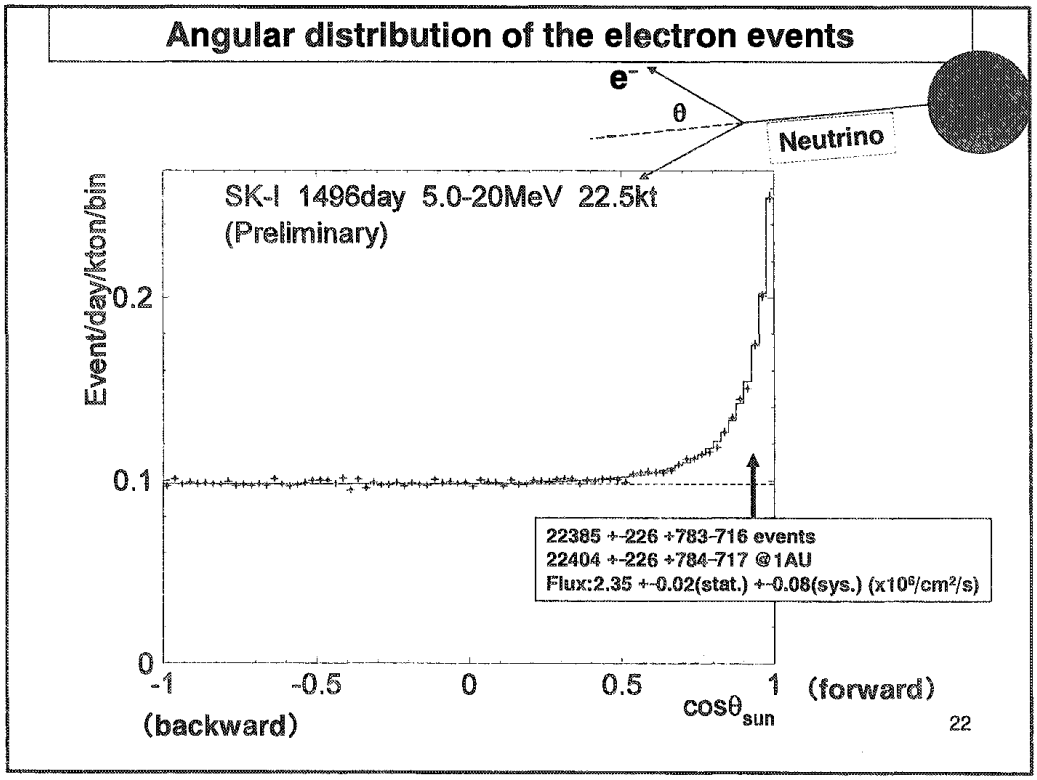
- measure $\phi(\nu_e) + \phi(\nu_{\mu,\tau})$



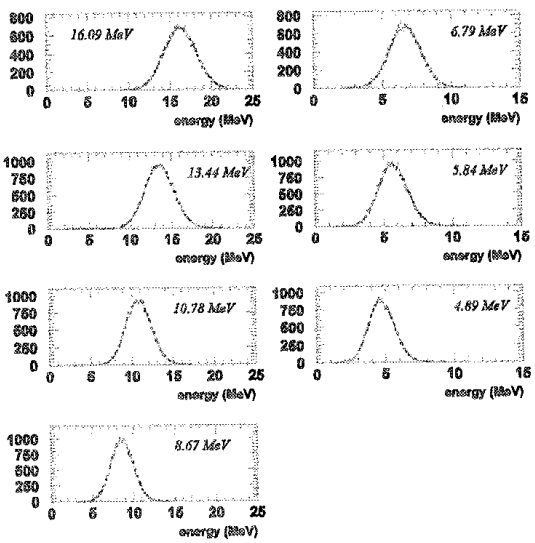
20

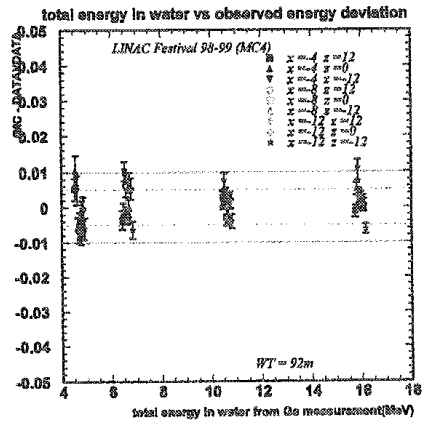
sno/UW_Talk.ppt


A typical solar-neutrino event (Super-K)
OD
Super Kamiokande
NUM # 1
RUN # 4201
EVENT# 1154571
DATE 97-Jun- 8
TIME 4:58:28
TOT PE: 137.1
MAX PE: 7.8
NMHIT : 103
ANT-PE: 14.3
ANT-MX: 1.8
NMHITA: 25
Basic reaction:
$\nu_e + e \rightarrow \nu_e + e$
ID
$E_e = 9.87$MeV
$\cos\theta_{sun} = 0.915$
RunMODE
TRG ID
T diff
FSCC:
TDC0:
Q thr.
BAD ch
SUB EV
event 1154571


Angular distribution of the electron events
e^-
θ
Neutrino
SK-I 1496day 5.0-20MeV 22.5kt
(Preliminary)
Event/day/kton/bin
0.2
0.1
0
22385 +-226 +783-716 events
22404 +-226 +784-717 @1AU
Flux:2.35 +-0.02(stat.) +-0.08(sys.) (x10⁶/cm²/s)
-1
-0.5
0
0.5
1
(forward)
(backward)
$\cos\theta_{sun}$
22

Calibration for solar neutrinos

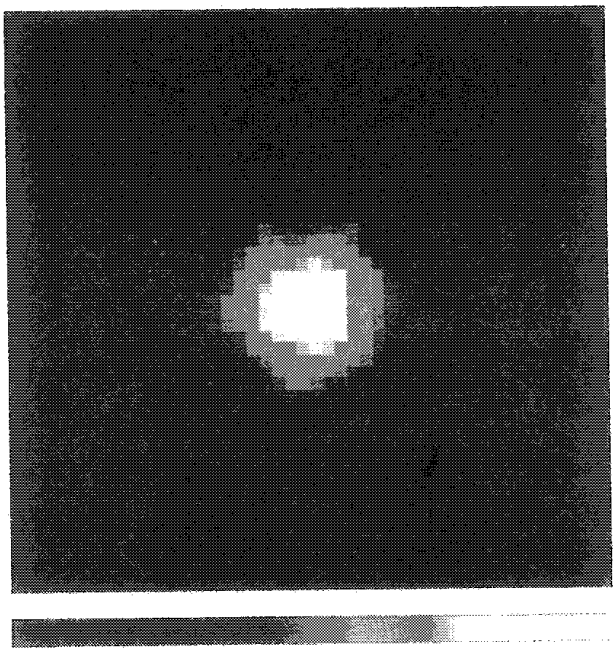




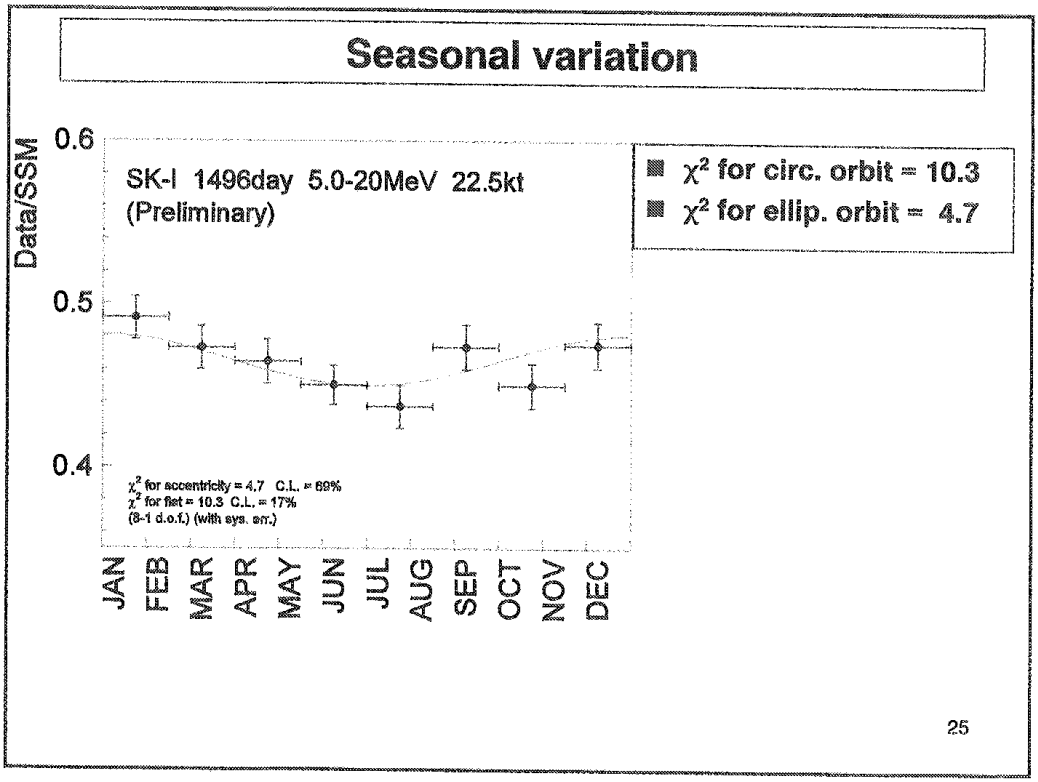
Absolute energy scale known with ± 0.6% over the entire fiducial volume

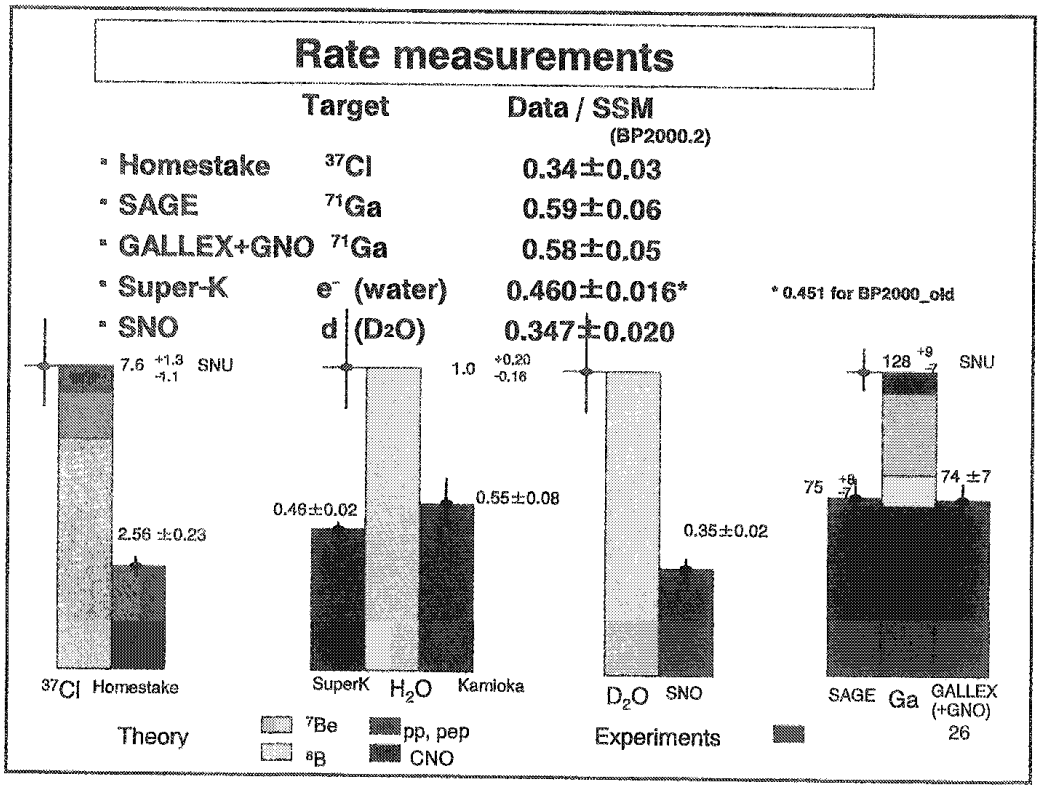
23

Solar image

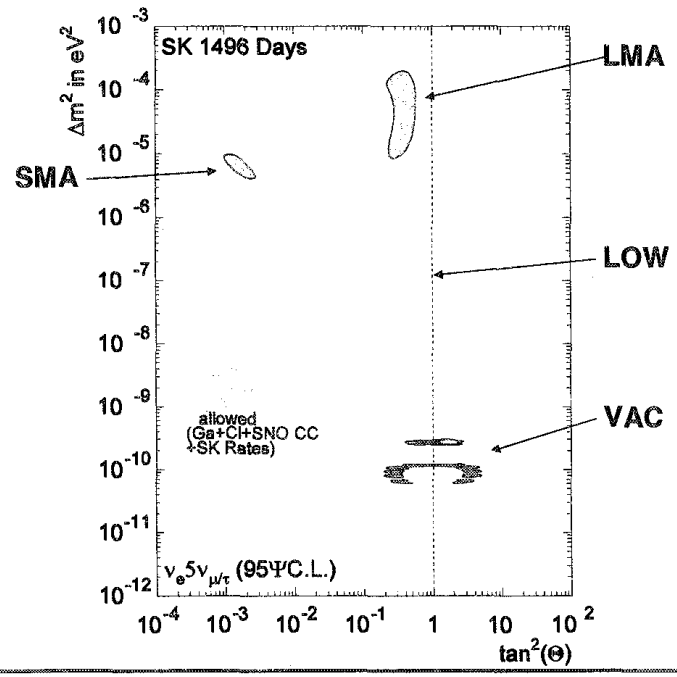


24


Seasonal variation
Data/SSM
0.6
0.5
0.4
SK-I 1496day 5.0-20MeV 22.5kt
(Preliminary)
χ² for circ. orbit = 10.3
χ² for ellip. orbit = 4.7
χ² for accentricity = 4.7 C.L. = 69%
χ² for flat = 10.3 C.L. = 17%
(8-1 d.o.f.) (with sys. err.)
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
25

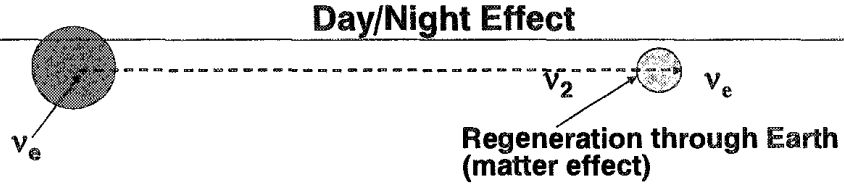

Rate measurements
Target
Data / SSM
(BP2000.2)
Homestake ³⁷Cl 0.34±0.03
SAGE ⁷¹Ga 0.59±0.06
GALLEX+GNO ⁷¹Ga 0.58±0.05
Super-K e⁻ (water) 0.460±0.016*
SNO d (D2O) 0.347±0.020
* 0.451 for BP2000_old
7.6 +1.3 -1.1 SNU
2.56 ±0.23
1.0 +0.20 -0.16
0.46±0.02
0.55±0.08
0.35±0.02
128 +9 -7 SNU
75 +8 -7
74 ±7
37Cl Homestake
SuperK H2O Kamioka
D2O SNO
SAGE Ga GALLEX (+GNO)
Theory
7Be
8B
pp, pep
CNO
Experiments
26

Allowed region from rate measurements

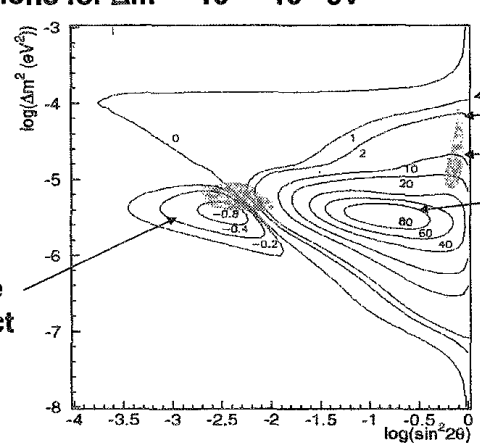


27

Day/Night Effect

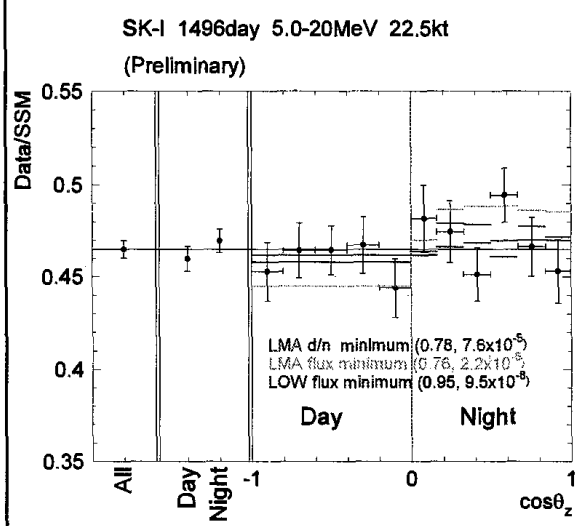


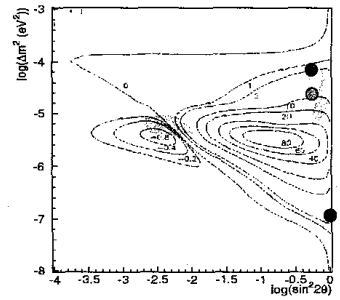
Earth density: ρ=5g/cm^3 (average), 13 (at core)
affects oscillations for $\Delta m^2 = 10^{-6} - 10^{-4}$ eV^2



28

Day/night data

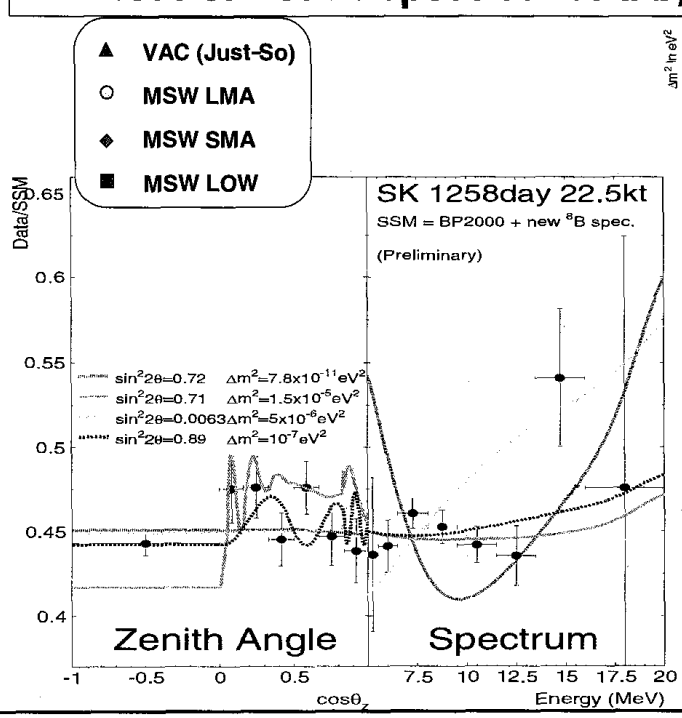


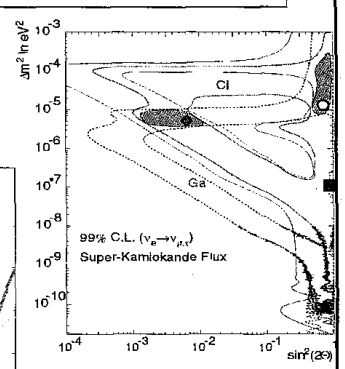


- (D-N)/((D+N)/2) = -0.021 ± 0.020(stat.) +0.013-0.012(sys.)
  - eccentricity corrected
  - 1.0 sigma stat. only
  - 0.9 sigma including sys. err.

29

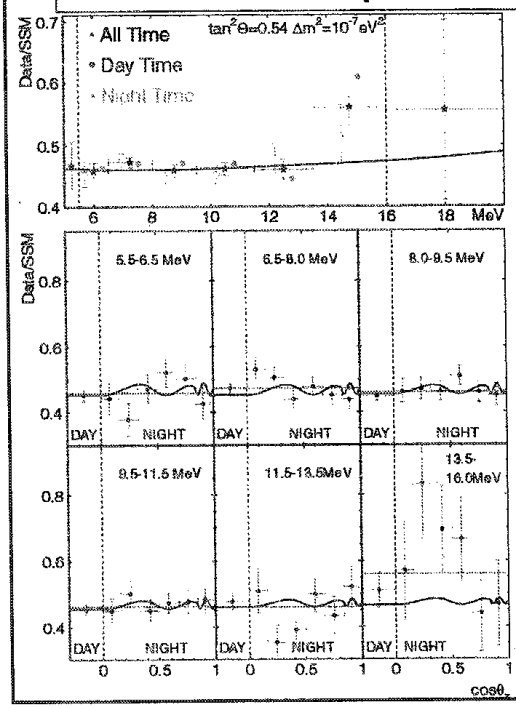
Electron recoil spectrum and D/N (1258 days)





30

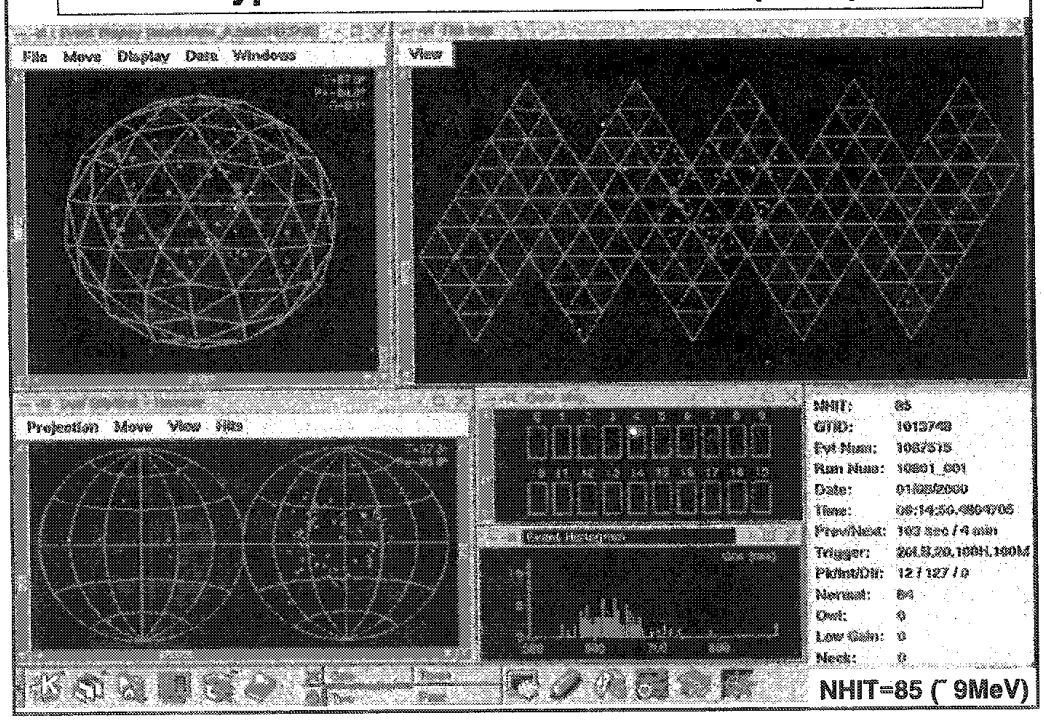
Data on (zenith angle and energy)



- **Zenith spectra**
  - Divide data into day + 6 night bins
  - Make the energy distribution (6 bins) at each bin
- **Line: LOW ($\tan^2\theta$=0.54, Δm^2=1.0E-7) → excluded (95%CL)**

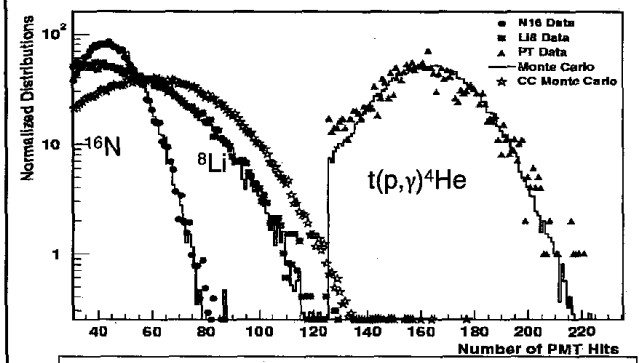
[redacted]

A typical solar-neutrino event (SNO)

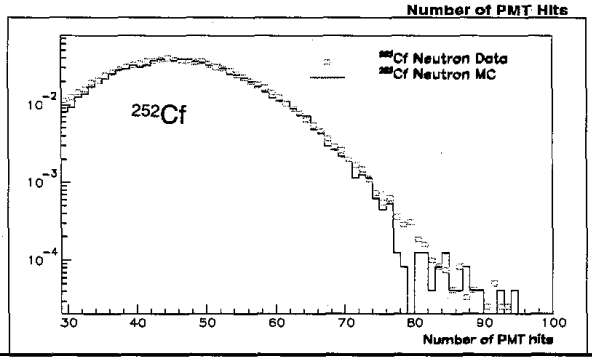


NHIT=85 (~9MeV)

Energy calibration for solar neutrinos

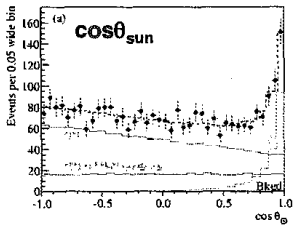


^{16}N	6.13MeV γ
^{8}Li	β (<~ 14MeV)
$t(p,\gamma)^4He$	19.8MeV γ
^{252}Cf	n (6.25MeV γ)



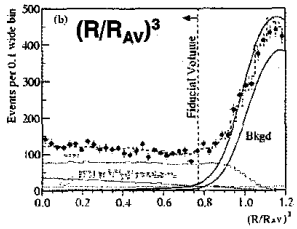
33

SNO spectra

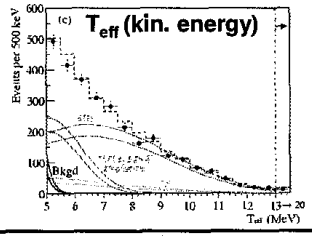


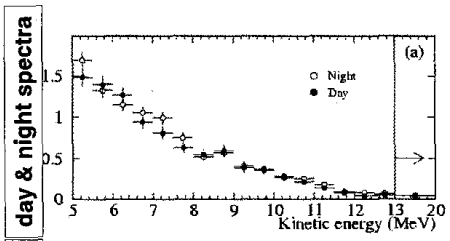
EC

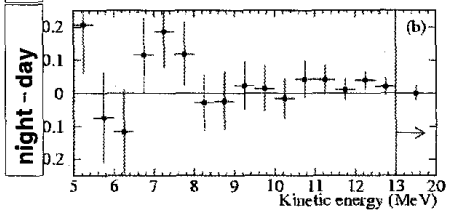
CC
NC



R_{AV}=6m
R_{fid}=5.5m





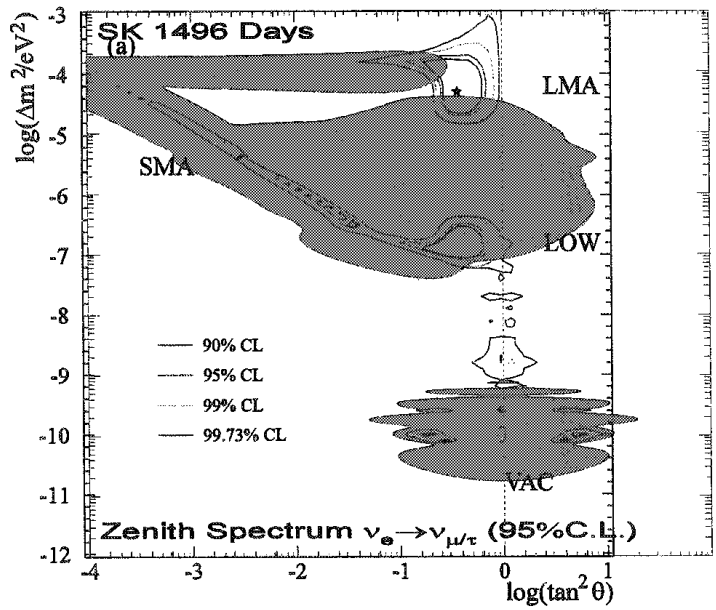


- CC: 1967.7 +61.9-60.9 events
- ES: 263.6 +26.4-25.6 events
- NC: 576.5 +49.5-48.9 events

34

SNO rates and day/night spectra and Super-K zenith spectra

No flux constraints



35

Super-K + SNO combined

SNO $\phi_{CC} = 1.7 \pm 0.05 \pm 0.09$ [$\times 10^6/cm^2/s$]

$\phi_{ES} = 2.39 \,^{+0.24}_{-0.23} \pm 0.12$

$\phi_{NC} = 5.09 \,^{+0.44}_{-0.43} \,^{+0.46}_{-0.43}$

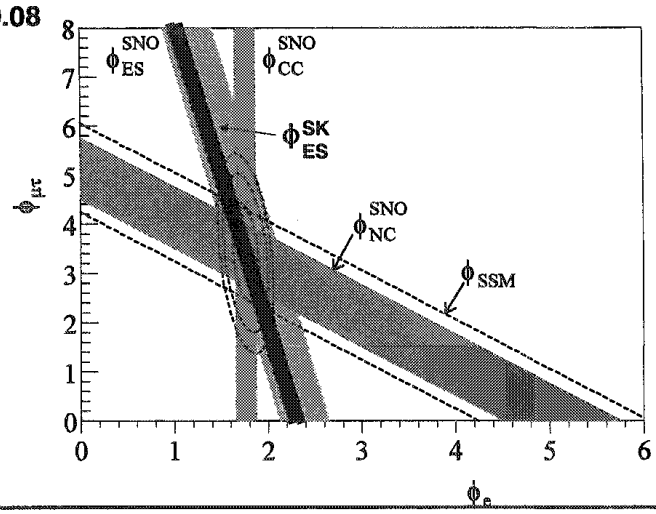
$\phi_{CC} = \phi_e$

$\phi_{ES} = \phi_e + 0.154\,\phi_{\mu,\tau}$

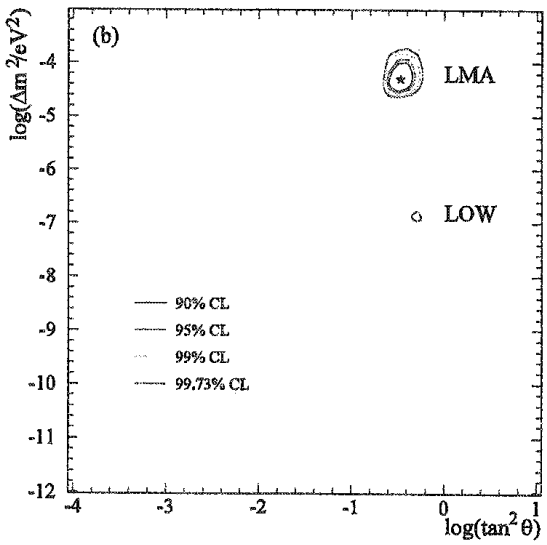
$\phi_{NC} = \phi_e + \phi_{\mu,\tau}$

SK $\phi_{ES} = 2.35 \pm 0.02 \pm 0.08$

Evidence for an active non-ν_e component !



Global analysis

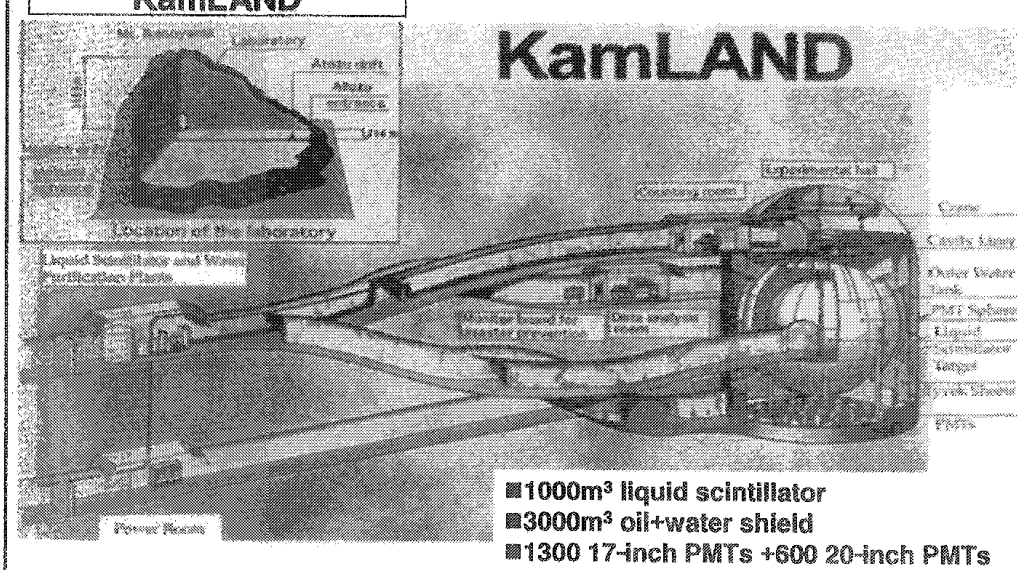


- Rates: Homestake (Cl), GALLEX (Ga), SAGE (Ga), SK (H2O), SNO (D2O)
- Super-K day/night, SNO day/night spectra
- Best fit:
  - $\tan^2\theta = 0.34 \quad (\neq 1)$
  - $\Delta m^2 = 5.0 \times 10^{-5}\ eV^2$
  - ^8B flux $= 5.86 \times 10^6\ cm^{-2}s^{-1}$ (SSM $= 5.05^{+1.01}_{-0.81}$)
  - $\chi^2_{min} = 57.0/72dof$

37

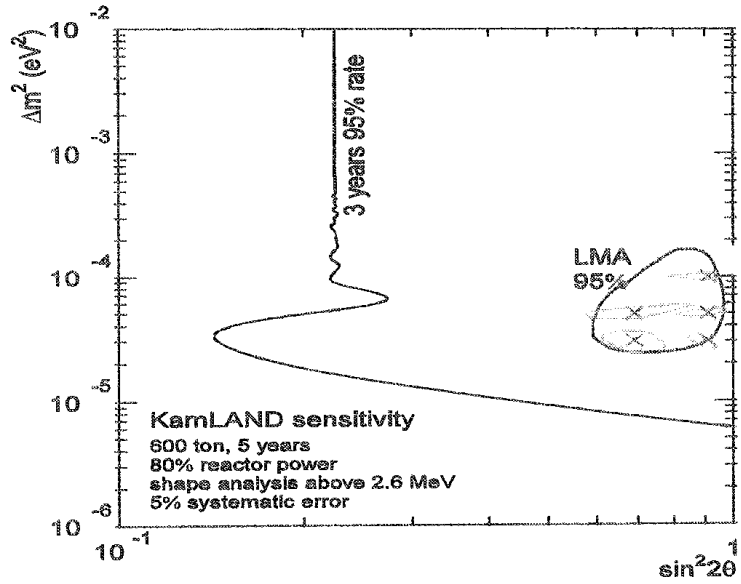
Ref: SNO collab. nucl-ex/0204009

KamLAND



- 1000m³ liquid scintillator
- 3000m³ oil+water shield
- 1300 17-inch PMTs +600 20-inch PMTs
- Anti-ν_e from reactors (L~ 170km)
- Detect e^+ from $\nu_e + p \rightarrow e^+ + n$ (Eth = 1.8 MeV)
- Observation started on 22 Jan 2002
- Threshold energy ~ 1MeV

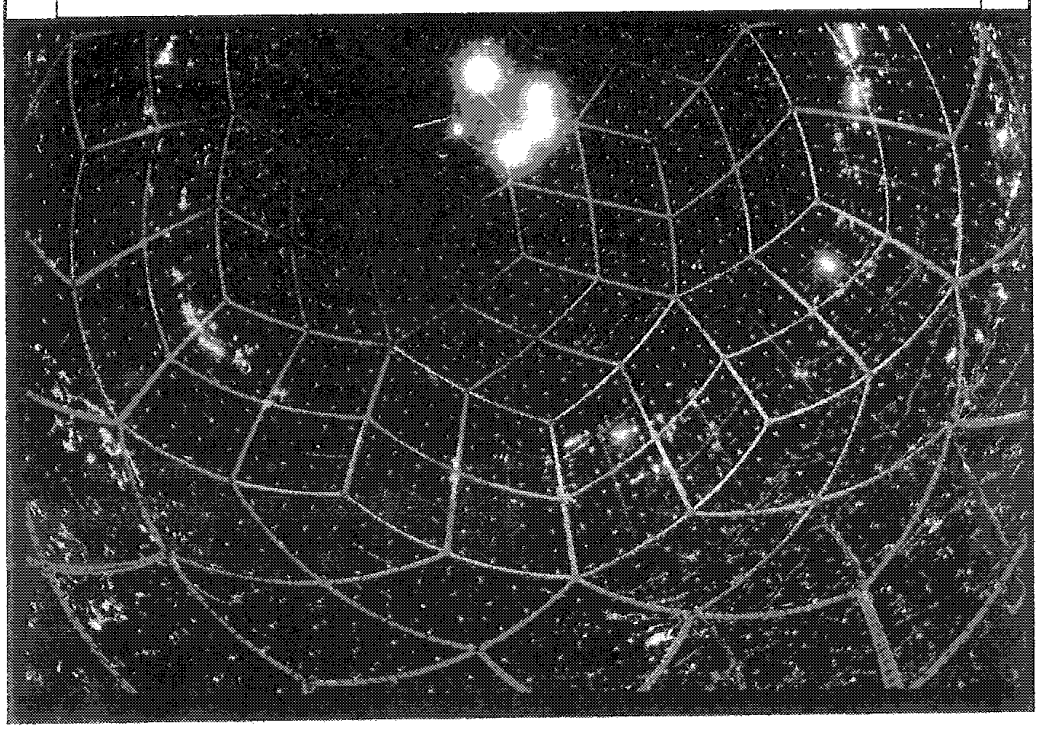
KamLAND: sensitivity



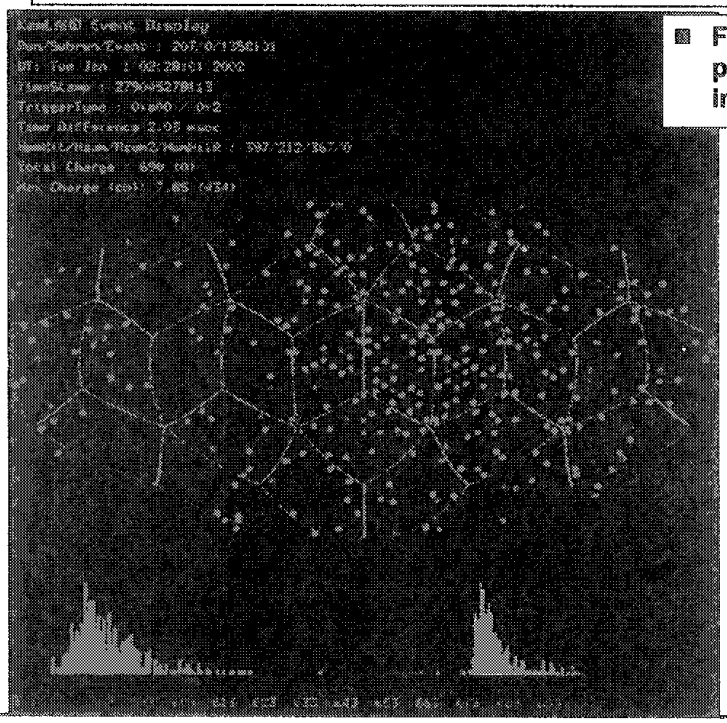
39

From K.Inoue (Tohoku Univ.)

Inside view



A signal from the reactor

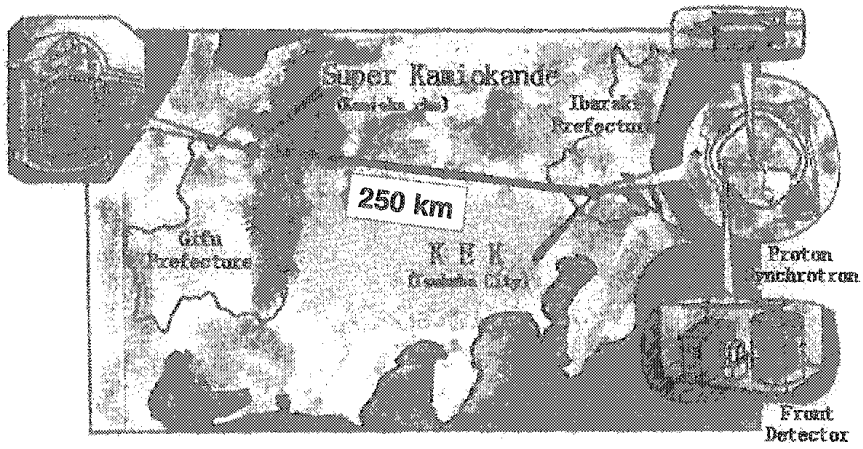


- First results will be presented at PANIC02 in September this year.

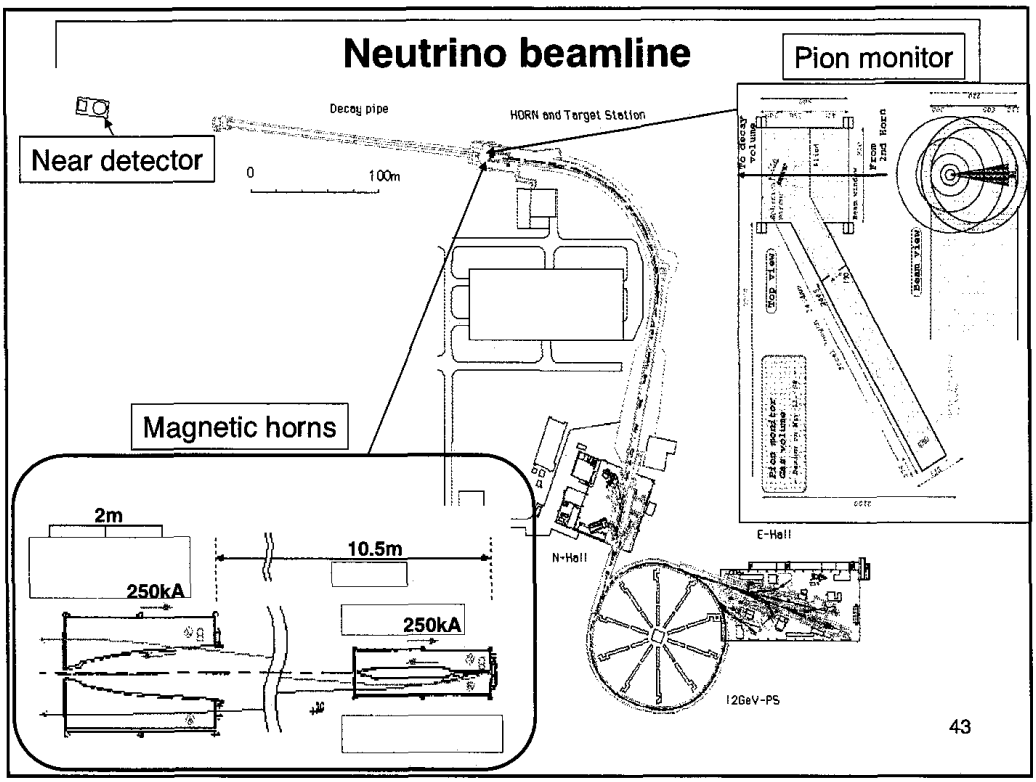
41

Accelerator neutrinos (K2K experiment)

5.6×10^{19} pot (June 1999 – July 2001)

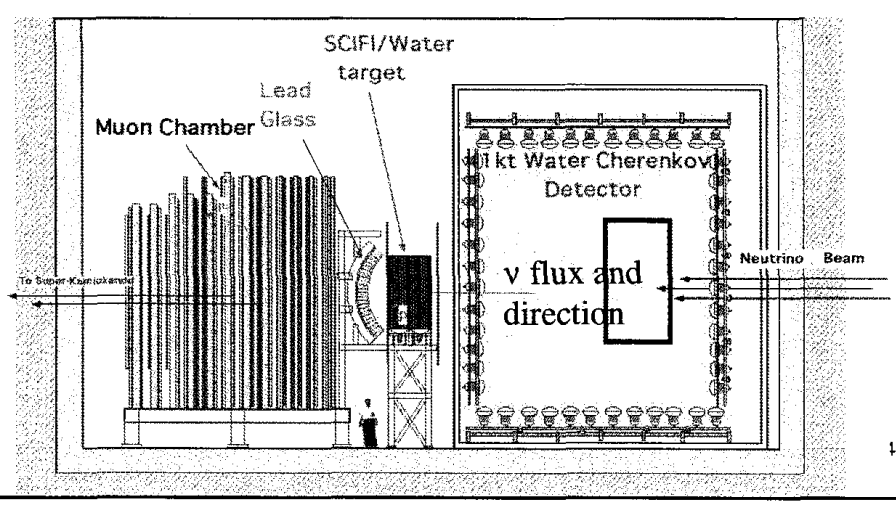


42

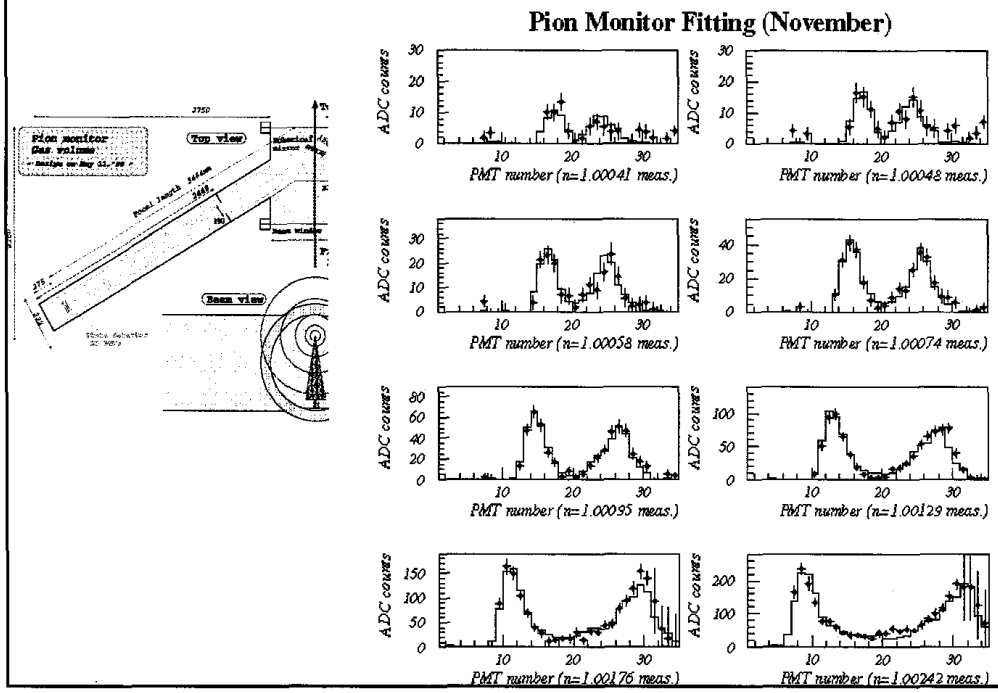


Near Detectors at KEK

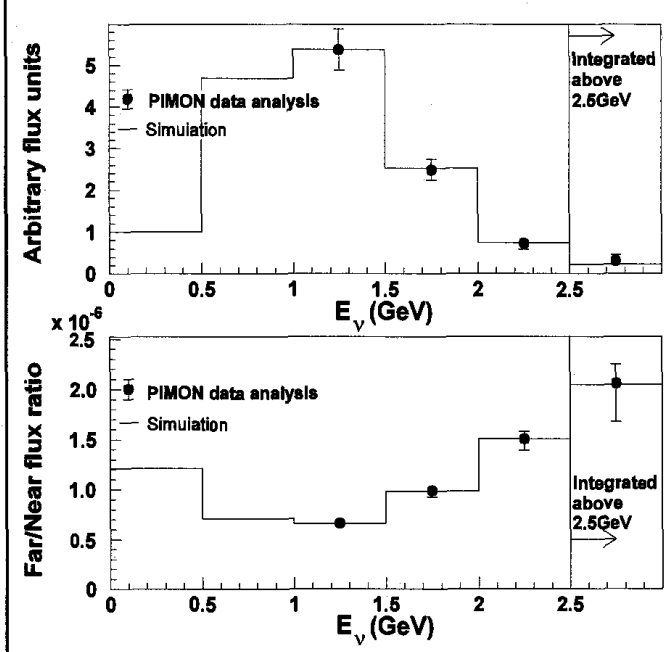
1kt Water Cherenkov detector (KT): 25 ton fiducial (4ϕ×2*l*)
+ Water tube sandwitched Scintillation fiber detector (SciFi): 6 ton
+ Muon range detector(MRD) 700 ton
+ Lead glass detector (LG)



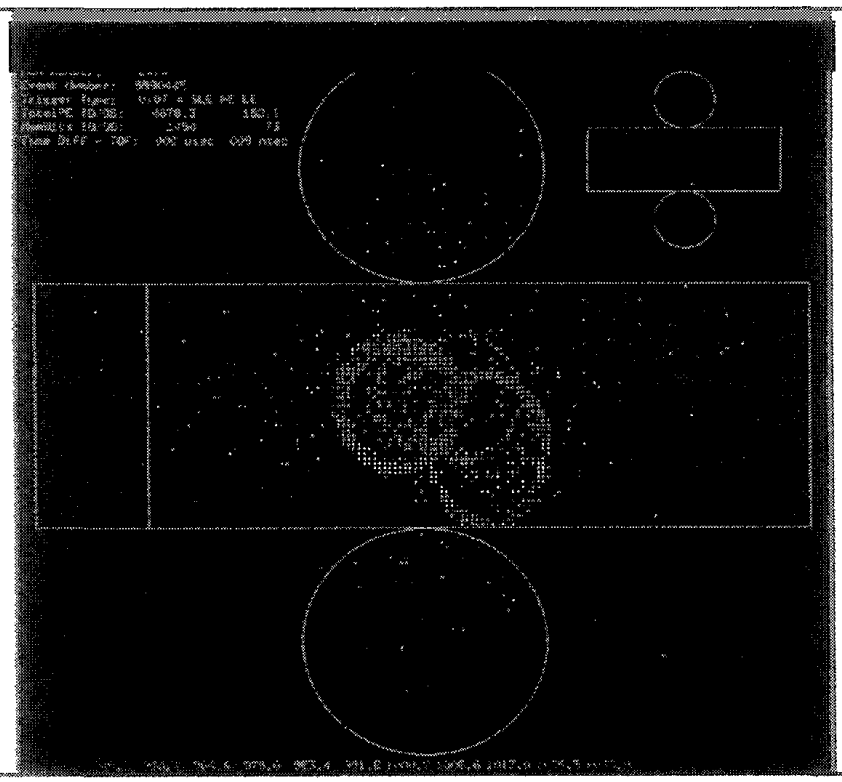
Pion monitor



Eν and Far/Near ratio



Sys. error = +6 –7%



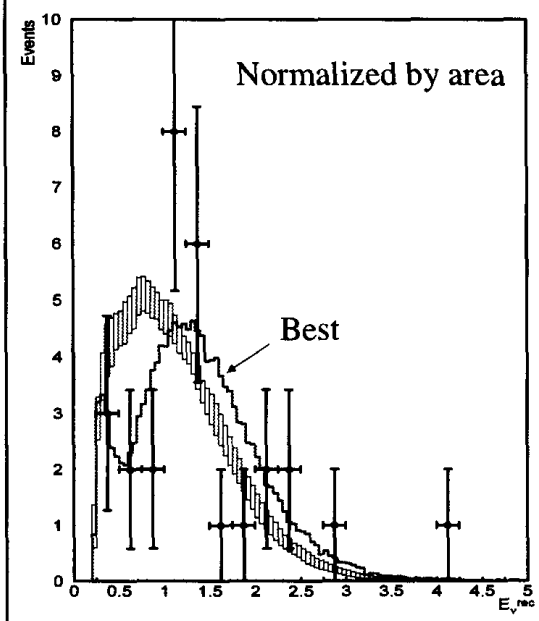
47

Summary of FC events (Jun99-Jul01)

	N_{SK}^{obs}	$N_{SK}^{expected}$ null oscillation	$\Delta m^2 = 3 \times 10^{-3}$	5×10^{-3}	7×10^{-3} (eV^2)
FC (22.5 kt)	56	$80.6^{+7.3}_{-8.0}$	52.4	34.6	29.2
1-ring	32	48.4±6.7	28.1	17.8	16.6
μ-like	30	44.0±6.8	24.4	14.6	13.5
e-like	2	4.4±1.7	3.7	3.2	3.0
multi-ring	24	32.2±5.3	24.3	16.8	12.6

48

Best fit 1Rμ spectrum & Nsk



- Best fit point ($\sin^2 2\theta$, Δm^2) = (1.0, $2.8\times10^{-3}eV^2$)
- KS test prob.(shape): 79%
- N_{SK}=54 (Obs.=56)
- Very good agreement on Shape & N_{SK}

49

Null Oscillation Probability

Null Oscillation Probability

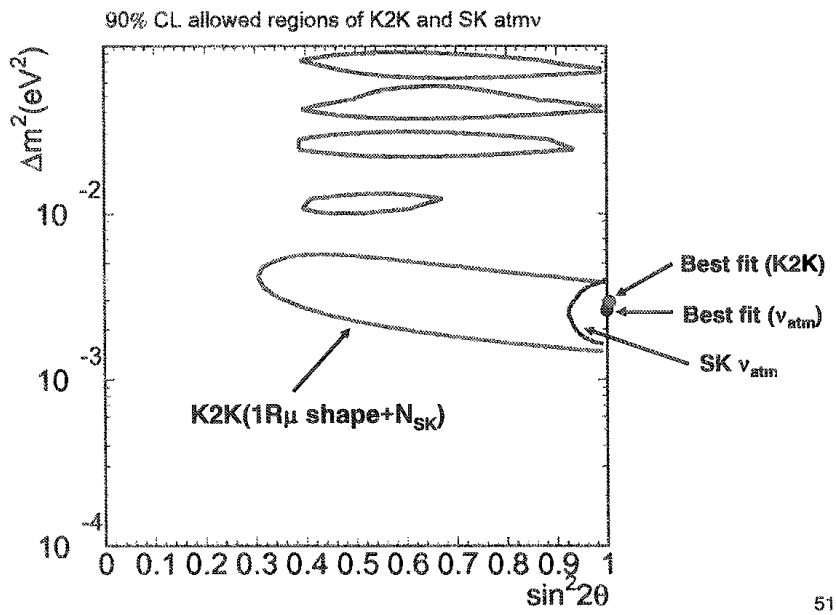
	analysis-1	analysis-2
N_{SK}only	1.3%	0.7%
Shape only	15.7%	14.3%
N_{SK}+Shape	0.7%	0.4%

Best fit ($\sin^2 2\theta$, Δm^2)

Shape only	(1.0, $3.0\times10^{-3}eV^2$)	(1.0, $3.2\times10^{-3}eV^2$)
(Allowing unphys.)	(1.09, $3.0\times10^{-3}eV^2$)	(1.05, $3.2\times10^{-3}eV^2$)
N_{SK} +Shape	(1.0, $2.8\times10^{-3}eV^2$)	(1.0, $2.7\times10^{-3}eV^2$)
(Allowing unphys.)	(1.03, $2.8\times10^{-3}eV^2$)	(1.05, $2.7\times10^{-3}eV^2$)

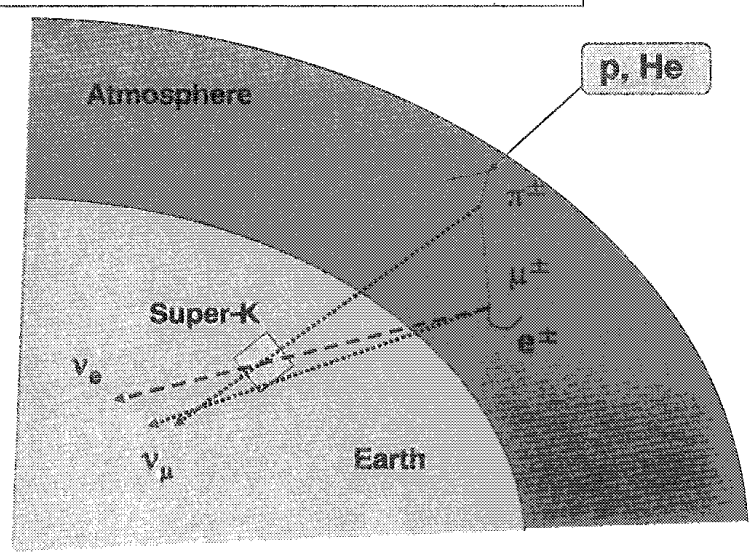
Both Shape and N_{SK} +Shape indicate consistent parameter region 50

Allowed regions



51

Atmospheric Neutrinos (1)



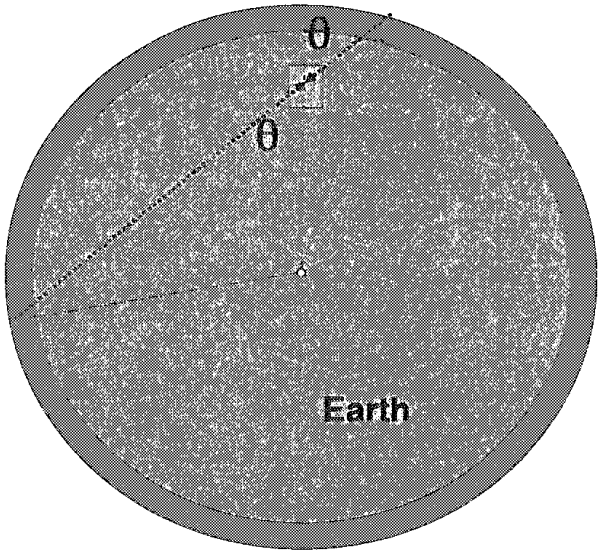
The expected event rate is uncertain by ~ 25%

$\nu_\mu : \nu_e = 2 : 1$ (low energy) known better than 5%.

52

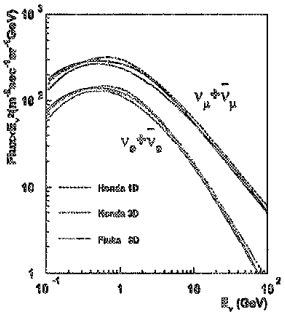
(2)

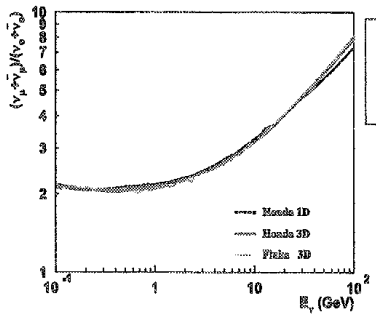
Up-down symmetry of the flux



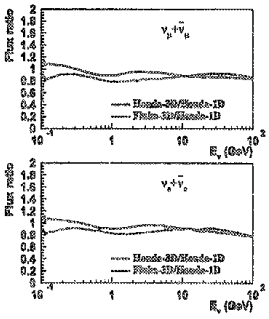
53

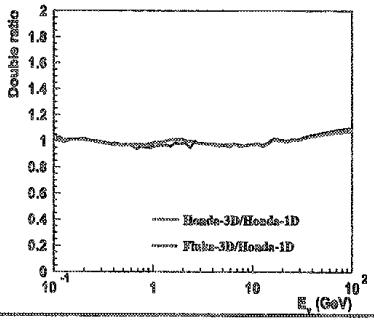
Atmospheric neutrino spectrum





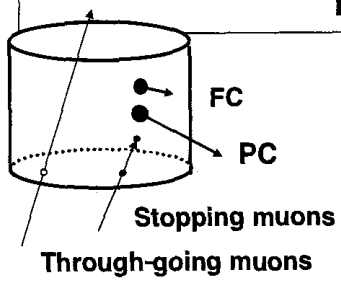
- 1D calculation will be used unless specified





54

Event topology

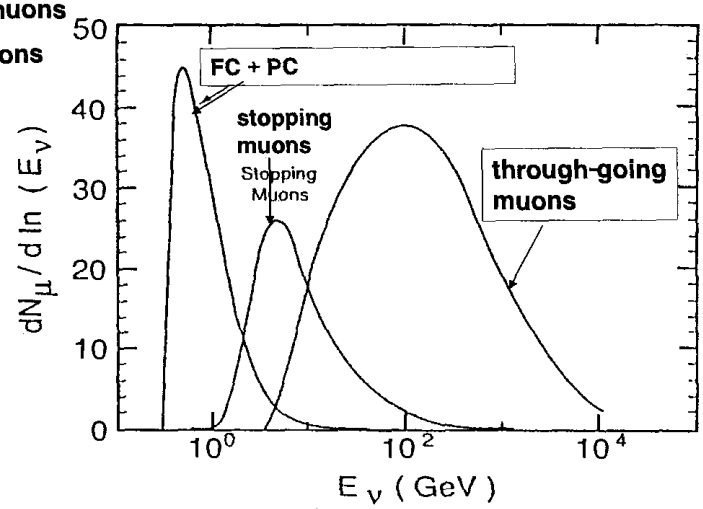


$\nu_e + N \rightarrow e + X$

$\nu_\mu + N \rightarrow \mu + X$

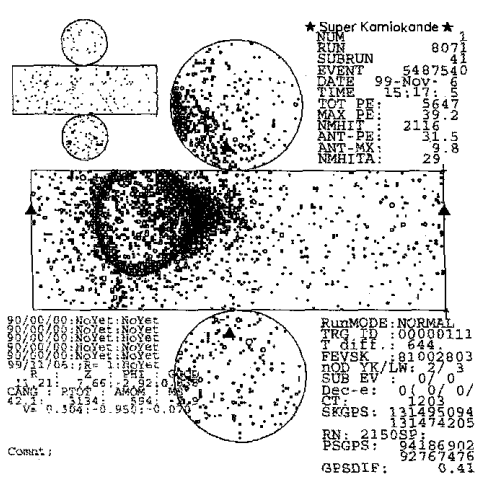
$\cancel{\nu_\tau + N \rightarrow \tau + X}$

Initial neutrino energy spectrum

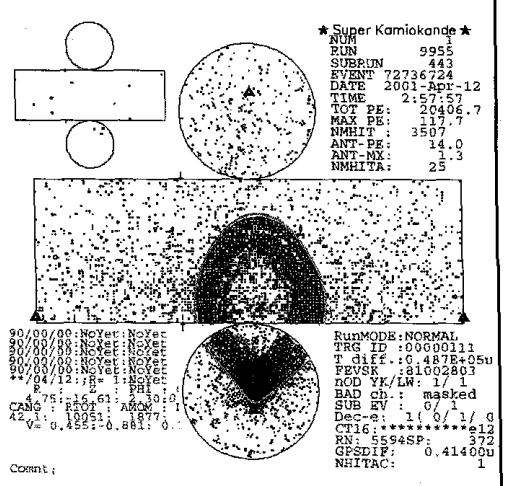


Electron-like and muon-like events

e-like



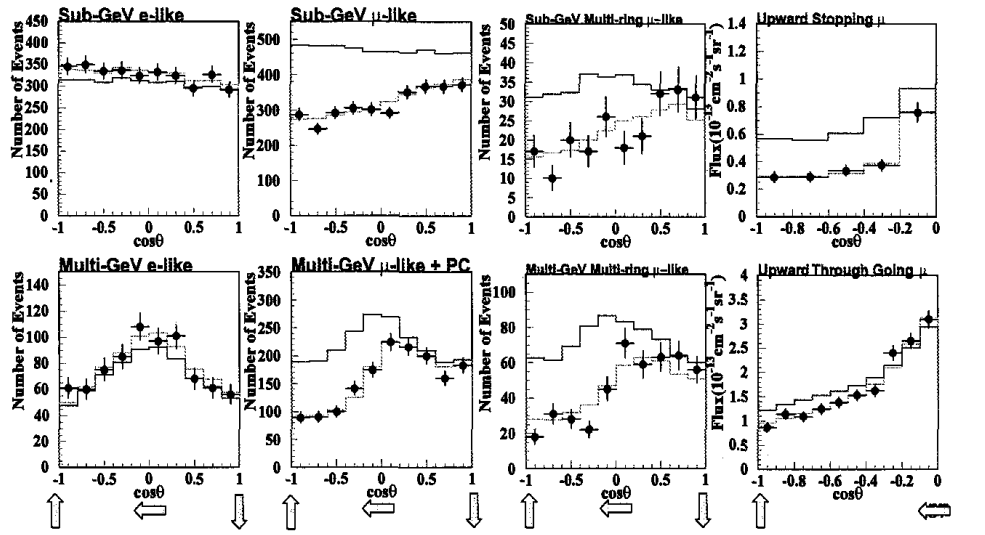
μ-like



Zenith angle distributions (FC+PC+up-μ, 1498 days)

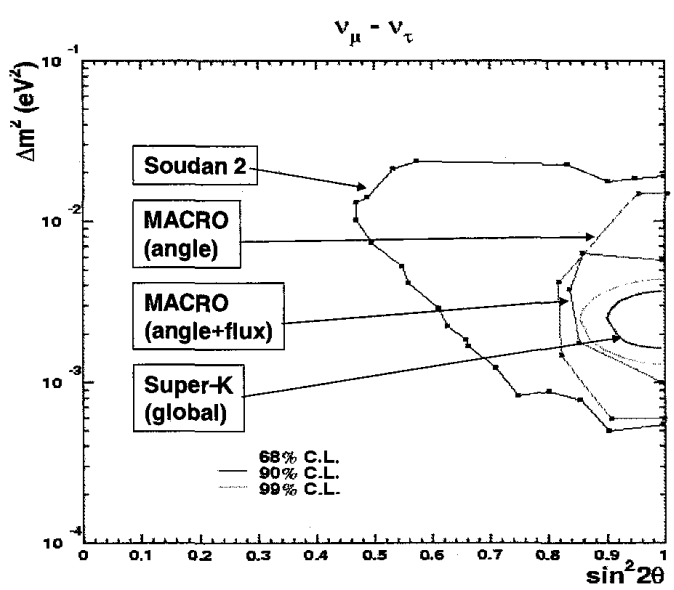
$\nu_\mu \leftrightarrow \nu_\tau$
2-flavor oscillations

Best fit (Δm_{23}^2=2.5x10^{-3}eV^2, $\sin^2 2\theta_{23}$=1.0, χ^2_{min}=163.2/170 dof)
(Δm_{23}^2=2.5x10^{-3}eV^2, $\sin^2 2\theta_{23}$=1.03, χ^2_{min}=162.7/170 dof, unphysical region allowed)
Expected without oscillation
(null osc: χ^2=456.5/172 dof)



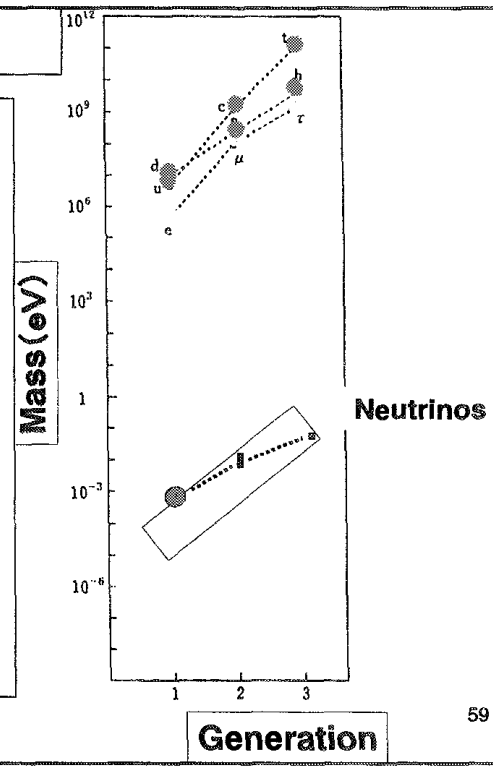
Allowed regions for $\nu_\mu \rightarrow \nu_\tau$

Allowed regions (90% CL)



Δm^2

- $\Delta m_{23}^2 \sim 1.6 - 3.9 \times 10^{-3}$ eV^2
- $\Delta m_{12}^2 \sim 3 - 10 \times 10^{-5}$ eV^2
- **Hierarchical masses ($m_3 > m_2 > m_1$):**
  - $m_3 \sim 0.04 - 0.06$ eV
  - $m_2 \sim 0.005 - 0.01$ eV
    - All 90% C.L.
- **Degenerate masses ($m_3 \sim m_2 \sim m_1$):**
  - $m_e < 2.2$ eV
    - (95% CL)
  - $\Sigma m_i < 6.6$ eV



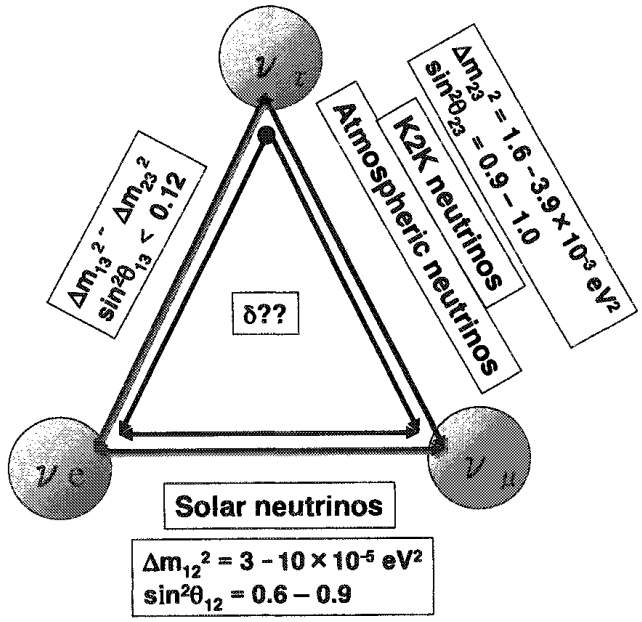
59

Summary of masses, mixings and $\Omega\nu$

- **Hierarchical masses:**
  - $m_3 \sim 0.04 - 0.06$ eV
  - $m_2 \sim 0.005 - 0.01$ eV
    - (90% CL)
  - $\Sigma m_i = 0.05 \pm 0.01$ eV
  - $\Omega_\nu = 1.0 \pm 0.3 \times 10^{-3}$
- **Degenerate masses:**
  - $m_e < 2.2$ eV
    - (95% CL)
  - $\Sigma m_i < 6.6$ eV
  - $\Omega_\nu < 0.17$
  - Note $\Omega_B = 0.04 \pm 0.01$

- **Neutrinos**
  - $\sin^2 2\theta_{12} = 0.6 - 0.9$
  - $\sin^2 2\theta_{23} = 0.92 - 1.0$
    - 90%CL
- **Quarks**
  - $\sin^2 2\theta_{12} = 0.188 \pm 0.007$
  - $\sin^2 2\theta_{23} = 0.0064 \pm 0.0010$
- **Why quarks and neutrinos are so different.**

60



61

Future

- Values of θ_{13} and CP phase δ are still unknown
  - New experiments are in preparation:
    - MINOS at Fermi Lab
    - CNGS (CERN to Gran Sasso)
    - JHF-Kamioka

62

CHAIRMAN: Y. TOTSUKA

Scientific Secretaries: G. Scioli, S. Sellitto

DISCUSSION

- *Polonektov:*
  Are there any theoretical models describing neutrino mixing like meson k?

- *Zichichi:*
  There exists a model (by Harald Fritzsch) based on the analogies between vector meson and electrons; so a neutrino is equivalent to a meson. That is a very fashionable analogy because it implies that a quark-antiquark system is equal to a lepton. This model fits with most experimental data.

- *Polonektov:*
  How can ν_e and ν_μ be distinguished?

- *Totsuka:*
  That is an experimental problem. If electrons go into the water, they are multiplied by a cascade and most electrons so generated have low energy and they diffuse Cherenkov light. In general the opening angle of Cherenkov light, in water, is 42°. Now that electron produces a diffuse Cherenkov light in a wide angle, while the muon has a very sharp edge of Cherenkov light.

  So we can distinguish between electron and muon light. We tested its accuracy at KEK in the range between 200 MeV and 1.5 GeV, and the mis-identification between electron and muon is about 1%. This is the PID system in a water Cherenkov detector.

- *Haidt:*
  What are the prospects for measuring θ_{13}?

- *Totsuka:*
  We have results from different neutrino oscillations experiments: Super-K, K2K, that is a running experiment, and MINOS, CNGS and JHF.

  At Super-K we are not able to go down to 0.1; MINOS, in future, could reach 0.03. By the way, I say that the limit for θ_{13}, coming from previous experiments, is 0.12 at $\Delta m^2 \approx 3 \cdot 10^{-3}$ eV^2. Then CNGS reaches 0.15 and at JHF-Kamioka the sensitivity could reach 0.06 and for neutrino factories it comes down to 10^{-3}. This is the estimated sensitivity for θ_{13}. However you need 10 years to find the final value of θ_{13}.

- *Bechtle:*

Can you exclude or access the LSND oscillation signals, which make the introduction of the sterile neutrino necessary?

- *Totsuka:*

If you only talk about the oscillation between, for example, ν_e and sterile neutrino, in the solar neutrino case, then we can say that it is not possible, but if you are talking about a mixture of sterile neutrino with the other kinds of neutrino, then it is still possible (the same is true for atmospheric neutrinos). To exclude completely the sterile neutrino is very difficult and we have to wait for the MINIBOONE experiment to confirm or not the LSND anomaly.

- *Ishino:*

Concerning τ appearance in the Super-K experiment; the τ production threshold is 3.5 GeV, so it is hopeless to see the τ signature in Super-K and also in JHF-Kamioka. What do you think about that?

- *Totsuka:*

Unfortunately at K2K the neutrino beam energy is low, the typical energy is only 1.5 GeV and you need a ν with a minimum energy of 3.5 GeV to produce a τ lepton. So it is hopeless to find a τ lepton in the K2K experiment.

Also for the JHF-Kamioka experiment we are going to use a very low energy neutrino beam (about 0.8 GeV) to maximize oscillation effects; so we are not very interested in the appearance of the τ lepton. The CNGS experiment will surely do it much better than the other experiments, so it is better to wait for this CNGS experiment.

- *Levy:*

Can you explain why the ν_e data are symmetric?

- *Totsuka:*

I showed a cartoon before this figure. Maybe this expresses why, first of all, if there is no oscillation then you understand that the flux must have an up-down symmetry that seems to be a good measure. Then this is just an experimental electron spectrum showing up-down symmetry which reveals no anomaly in the electron spectra. On the contrary, you can interpret this anomaly in the muon sector, assuming that the oscillation of ν_μ into ν_τ in our detector is taking place, just at these atmospheric neutrino energies. In the downward direction in our detector, these oscillations have disappeared, but oscillations are in full action in the upward direction and in this case you lose a large amount of events. Oscillations of ν_μ to ν_e do not occur here because we have that in the solar neutrino Δm^2 is much smaller than Δm^2 that we are considering now. Therefore even a path range of 10000 Km is too short.

- *Polosa:*

Is the distance from CERN to Gran Sasso sufficient to observe neutrino oscillations of the kind you described before?

- *Totsuka:*

I am not a member of this experiment but the distance is 730 Km and it would be an appearance experiment but I think they will see about 20 events per year.

- *Maas:*

You managed to hit the detector from 250 km. Could you also manage to extend the range to a few thousand kilometres by placing the source, e.g., in Russia?

- *Totsuka:*

There are two parameters: distances and energy of neutrino ($\Delta m^2 L/E$). So instead of increasing the distance you can reduce the energy (for example at a distance of 200 Km, that is the distance between KEK and Kamioka, the optimal energy of the neutrino is about 0.8 GeV and we can still reduce it). But if you are interested in the sign of Δm^2, then you have to go up to higher energy values and in that case you need a kind of neutrino factory to get these ones.

- *Strassler:*

You showed KamLand's sensitivity after three years of running. You also mentioned KamLand has promised results this month, and that they started in January. Do you have an estimate for the best possible sensitivity they might have after nine months?

- *Totsuka:*

I am not a member of KamLand but they are very close to an experimental sighting. Therefore I asked young people to tell me more about this and one guy told me that they have analysed and observed somewhat less than thirty events coming from reactors; then I guess that you see something like fifty events. If the LMA solution is the right solution of the solar neutrino problem, then they expect, by the time of conference, the 40% reduction of the events; that means the expected number of events is 50/0.6 (about 80 events). So you have to compare the 80 expected events with the 30 or 50 observed experimentally. Another guy told me that he did not find anything surprising (that may mean the reduction of events).

- *Bechtle:*

Can one obtain some values for neutrino masses from this model to implement the values of these masses in the SM?

- *Totsuka:*

Experimentally, as you know, we only measure Δm^2, so we do not know anything about the absolute mass values. However I think these are already implemented from a theoretical point of view.

- *Ishino:*

Does acrylic cover make some significant effect on your measurements?

- *Totsuka:*

We have a reduction of the Cherenkov light of a few percent, maybe 5-6%.

- *Ishino:*

Is there no change in the energy threshold?

- *Totsuka:*

There is a big change, because we have only the 47% of the number of the phototubes. This increases the threshold from 4.5 MeV, that we had before, maybe to 7-9 MeV. So we have to abandon the observation of solar neutrinos.

- *Levy:*

Please, can you tell us something more about the JHF-Kamioka?

- *Totsuka:*

JHF project is a next generation proton accelerator which is under construction about 30 km north of KEK and the distance from Super-K is 295 km. There is a 400 MeV proton LINAC, a 3 GeV PS and a 50 GeV PS.

The 3 GeV PS has a 1 MW beam power; most of them are used for neutron facilities while the 50 GeV PS has a 0.77 MW beam power. The protons are extracted by this accelerator and neutrinos are produced. Compared to the other experiments, for example the beam power is much larger than previous K2K and twice larger than MINOS. Furthermore, we foresee to increase the beam power in the future by a factor four. The sensitivity for θ_{13}, for this experiment, will come down easily to less than 0.01.

However, from a MonteCarlo simulation, we can say that the expected number of events is about 11 compared to a total of 36 for the background, but if you look at the maximum oscillation peak, then the background is not large and the signal to noise ratio is much better, that is why we are saying that we could reach 0.006 at $3 \cdot 10^{-3}$ eV^2, at neutrino energies where we expected the maximum oscillation.

In order to observe CP-violation by observing electron appearance, what we have to do is to try to observe differences of particle oscillations and antiparticle oscillations. This asymmetric parameter is expressed by Δm_{12}^2, θ_{12} and θ_{13} and this is what we want to measure. To observe this CP-violation, θ_{12} must be large and that means that solar neutrino solution must be LMA. Furthermore, θ_{13} also has to be large

in order to get a reasonable number of events for both neutrino and antineutrino appearance.

In this figure the background is assumed to be understood with an accuracy of 2% which is subtracted to obtain signals. Let us assume that the limit of θ_{13} is 0.04, then this experiment has a sensitivity $\sin\delta > 0.35$.

The JHF accelerator is under construction but unfortunately the neutrino beam line is not yet funded; but we expect beam line to be ready in 2007. This is our goal.

- *Haidt:*

You mentioned the L/E ratio. Is it also conceivable to choose L large, for instance a detector in China?

- *Totsuka:*

This is what Chinese people want. But I think if you go to a distance of about 2000 km (the distance from China), then you need much more intensity at a constant energy. Furthermore, you also have to increase the neutrino energy to compensate the oscillation length (dependence of oscillation $\propto \Delta m^2 L/E$). So in that sense I think this PS is not yet strong enough, but it could be used as an injector for a neutrino factory.

- *Hiroki:*

Will JHF be a good candidate for a neutrino factory?

- *Totsuka:*

I do not know, that is a project for 2020.

The Fermilab Experimental Physics Program

R. Tschirhart
Fermi National Accelerator Laboratory

Only the discussions are reported here, since the camera-ready copy of the contribution did not arrive in time.

CHAIRMAN: R. TSCHIRHART

Scientific Secretaries: G. Brooijmans, J. Wendland

DISCUSSION

- *Maillard:*
  What are the signatures of extra-dimensions in accelerators?

- *Tschirhart:*
  There are many possibilities. In the Arkani-Hamed-Dimopolous-Dvali (AHDD) models, a graviton can be directly produced and escapes into an extra dimension. The signature is missing transverse energy and a jet or a photon or a Z boson. A virtual graviton exchange can also take place, because it couples to the energy-momentum tensor. This leads to high mass Drell-Yan, for example with anomalous angular distribution.

  In another type of model, the Randall-Sundrum models, high mass resonances would be produced. They decay to quarks, leptons, or gauge bosons.

- *Korthals-Altes:*
  Some people have calculated cross-sections for black hole production at Fermilab, assuming that the Planck scale is in the TeV range. What are you supposed to look for?

- *Tschirhart:*
  Black holes do not conserve baryon or lepton number individually. A black hole would "evaporate" into many particles very quickly, including "easy-to-detect" leptons.

- *Maas*:
  You have hinted at a possible measurement of CPT violation by the neutrino experiments. Such a violation would have profound consequences. Could you go into more detail?

- *Tschirhart*:
  The LSND experiment has seen a 3.3 sigma signal for anti-muon-neutrino into anti-electron-neutrino oscillations. One of the ways to accommodate this with the existing results from LEP (three neutrinos), atmospheric, and solar neutrino experiments would be to invoke CPT violation. In this case the masses of neutrinos and anti-neutrinos can be different. This possibility can be explored, if the MiniBoone experiment is run with both muon-neutrino and anti-muon-neutrino beams.

- *Zichichi*:

Given three neutrinos, we cannot exclude any of the three possible transitions. LSND observed the transition anti-muon-neutrino to anti-electron-neutrino. There is nothing new in this transition. The new feature is the large difference of the squared masses. But the experiments measure the product of the squared mass difference times sin(θ). All angles of θ measured so far are large (sin(θ) = 1). If θ_{12} is large, then the difference of the squared masses of flavours 1 and 2 is small, just like the others.

- *Tschirhart*:

The oscillation probability is given by $P(\nu_\mu \rightarrow \nu_e) = \sin^2(\theta^{12}) \sin^2(1.27\ \Delta\ m^2\ L/E)$, with $\Delta\ m^2 = |m1^2 - m2^2|$. Given P, L, and E a region in the $\sin^2(\theta_{12})$ -- $\Delta\ m^2$ plane is allowed. For LSND, the lowest allowed value for $\Delta\ m^2$ is $3*10^{-2}\ eV^2$ (at $\sin^2(\theta_{12}) = 1$) while the allowed region for atmospheric neutrinos is 1 to $6*10^{-3}\ eV^2$ (Super Kamiokande, $(\nu_\mu \rightarrow \nu_x)$. This implies 3 distinct values for $\Delta\ m^2$: solar (approx. $10^{-5}\ eV^2$) atmospheric (1 to $6*10^{-3}\ eV^2$) and LSND ($> 3*10^{-2}\ eV^2$). However, with three hierarchical neutrinos, only two distinct values for $\Delta\ m^2$ are possible.

- *Polosa*:

I have a naive question about Miniboone. Suppose you produce 10,000 positive pions of which about 10 decay into an electron anti-neutrino pair. How do you distinguish between these electron-neutrinos and those coming from the oscillation of an anti-muon-neutrino?

- *Tschirhart:*

You do not. Neutrino beams are typically mainly composed of muon-neutrinos with contaminations of anti-muon-neutrinos (~5%), electron-neutrinos (~1%) and anti-electron-neutrinos (~0.1%). (The electron-neutrino component is dominated by neutrinos from Kaon decays.) The production rates for pions and kaons were measured fairly accurately and are used in beam simulations. These simulations will be validated by measuring the radial and energy spectra of all components (except the "appearing" flavour), so the search looks for an excess above this irreducible background.

- *Nobbenhuis*:

It seems to me that to go from missing transverse energy to the observation of extra dimensions is a rather big step. Basically my question is: How specific are the experiments you perform at accelerators? Would it not be better to look for deviations of the $1/r^2$ law? I think supersymmetry also has a signature of missing transverse energy. How do you discriminate?

- *Tschirhart*:

Collider detectors tend to be multi-purpose and are designed to detect a large number of signatures. They are well-equipped to detect electrons, muons, taus, photons, bottom and charm mesons and baryons, missing transverse energy, and jets.

For the compactification radii under discussion, Cavendish experiments are only sensitive to two extra dimensions. For more dimensions, the distances at which effects can be detected are extremely small (smaller than a nanometre).

In the AHDD scenario, direct graviton production has the missing transverse energy back-to-back with a jet, a photon, or a Z boson. In super-symmetric events the missing transverse energy is due to the escaping lightest super-symmetric particle, which is produced at the end of the cascade decays. These events have activity (jets, leptons) in both hemispheres.

That being said, seeing anomalous events is only the first step. Finding their correct interpretation is the second step.

- *Strassler*

Prof. Tschirhart's answer is almost true except that there is a beautiful paper by H.H. Cheng, *et al.* that describes at least one scenario where it is extremely difficult to distinguish extra dimensions from SUSY. However, I agree with Prof. Tschirhart that the first issue is to find a signal. Then we will have lots of fun looking for its interpretation.

- *Golovnev*:

Run II of Tevatron has started slowly. What problems did you face?

- *Tschirhart*:

The luminosity increase with respect to Run I is a factor of 20. Since anti-protons need to be produced, the system is very complex. Protons are shot onto a target, the created anti-protons are collected and stored in the accumulator. When the Tevatron is ready for a new store, the anti-protons have to be extracted from the accumulator and transferred and accelerated through the main injector into the Tevatron, where the proton and anti-proton beams co-exist in the same beampipe on different helices.

The three main issues limiting luminosity are: The emittance increase in the accumulator (due to intrabeam scattering); the long range beam-beam interactions in the Tevatron that lead to large anti-proton losses before acceleration to 980 GeV; the large longitudinal proton emittance, that appears to be generated during acceleration in the main injector. All of these problems have to be addressed at the same time to make progress with luminosity. Over the past year, the luminosity has increased by a factor of four, which gives us some confidence that the remaining factor of three is within reach.

- *Ishino*:

Why does a visible light photon counter (VLPC) reach such a high quantum efficiency and is able to count single photons?

- *Tschirhart*:

As opposed to standard photo-multipliers, which have typical quantum efficiencies of about 25%, VLPCs are solid-state devices built from semi-conductors with a small energy gap leading to efficiencies larger than 70%. A drawback of these devices is the cryogenic temperatures at which they have to be operated.

- *White*:

Could you comment on the running cost and difficulties of a photo-detector operated at 20 degrees Kelvin?

- *Tschirhart*:

The main cost is related to the necessary amount of research and development and understanding the properties of the new system: It basically comes down to cost for manpower in both physics and engineering.

- *Nobbenhuis*:

What happens if an extra dimension is not an extra space dimension, but an extra time dimension? Can you say something about the experimental signature or experimental prospects?

- *'t Hooft*:

All theories with extra time dimensions have tachyonic states. There is no consistent theory that makes phenomenological predictions.

- *Bechtle*:

What is the advantage of a fibre tracker with respect to a drift chamber?

- *Tschirhart*:

The very compact volume available in the detector, which is dictated by the existing, excellent calorimeter, drives the choice for a fibre tracker, which is much more radiation hard than a drift chamber. Other advantages include that high voltage or gases are unnecessary.

- *Haidt*:

What is the accuracy that CDF and D0 are aiming for in measuring the mass of the W boson and that of the top quark?

- *Tschirhart*:

The ambitious goal for the top quark is around 1 GeV, which will require excellent understanding of the jet energy scale, in particular for b jets (hence the need to identify Z decays to b and $\bar{b}$). For the W boson mass the goal is 10 to 20 MeV.

- *Bechtle*:

What technique could be used to produce a pure K_L beam to study the decay into a neutral pion and neutrino anti-neutrino pair? How could the K_L be tagged?

- *Tschirhart*:

The main contamination of K_L beams comes from neutrons. The negative effects of in-beam neutrons can be minimized by careful collimation of the neutral beam line.

Photons in the beam lines are converted to electron-positron pairs with a lead-converter plate and charged particles are removed from the neutral beam with sweeping magnets. The K_L experiment is situated far downstream of the production target, so that K_S's and hyperons in the beam decay. A promising approach to tag the K_L decay with a time-of-flight measurement following a micro-bunched proton beam striking the production target is now being explored by the KOPIO experiment at BNL.

NEW TALENTS SESSIONS

A series of special sessions was devoted to young talents: those fellows who cannot be easily identified in the constantly increasing dimensions of experimental groups. The problem is less acute in the field of theoretical research. The results of the presentations follow.

Interpretation of the Search for Neutral Higgs Bosons at OPAL in a CP Violating MSSM Scenario

Philip Bechtle

OPAL Collaboration

Abstract

This article describes preliminary results of a search for neutral Higgs bosons within a CP violating Minimal Supersymmetric Model (MSSM). The results are based on the complete dataset collected with the OPAL detector at LEP with centre-of-mass energies ranging from $\sqrt{s} = 91$ to 209 GeV. The searches for the Standard Model Higgs boson and for the neutral Higgs bosons in the CP-conserving MSSM are re-interpreted and extended. The dominant production mechanisms, Higgsstrahlung and Higgs boson pair production, are considered for the lightest (H_1) and the next-to-lightest (H_2) neutral Higgs mass eigenstates. No significant excess over the expected SM background was observed. Regions in the parameter space of a CP violating benchmark scenario are excluded.

This article summarizes preliminary OPAL results

1 Introduction to a CP violating MSSM

The MSSM assumes the existence of two complex scalar Higgs field doublets with eight degrees of freedom. As in the Standard Model (SM), three of them are used to give mass to the W^+, W^- and Z^0 gauge bosons. The remaining five degrees of freedom manifest themselves in two charged Higgs bosons H^+, H^- and in three neutral mass eigenstates H_1, H_2 and H_3, ordered with increasing mass. If all phases of MSSM parameters which are connected to the Higgs sector of the theory are zero, the three neutral Higgs mass eigenstates correspond to the two CP-even states h and H and to the CP-odd state A.

The introduction of CP violating phases into the MSSM is theoretically appealing, since CP violation is one of the three requirements to be fulfilled to generate the cosmic matter/antimatter asymmetry [1]. Moreover, the SM fails to allow a large enough rate of CP violating processes to explain the measured baryon asymmetry [2]. CP violating phases in the sector of direct soft supersymmetry breaking lead to an introduction of CP violation to the MSSM Higgs sector via first order loop corrections from third generation squarks to the otherwise CP invariant Higgs potential. This would allow the MSSM to fulfill the requirements for generating the cosmic baryon asymmetry [3].

In a CP violating MSSM scenario, the Higgs production mechanisms, Higgsstrahlung $e^+e^- \to H_iZ^0$ (with $i = 1, 2, 3$), pair production $e^+e^- \to H_iH_j$ (with $i \neq j$ due to Bose symmetry) and boson fusion $e^+e^- \to H_ie^+e^-$ and $e^+e^- \to H_i\nu\bar{\nu}$, proceed as in a CP conserving MSSM. As the Higgs mass eigenstates do not have defined CP quantum numbers, the production of all three mass eigenstates in Higgsstrahlung is allowed. Hence, the coupling of the Higgs bosons to the Z^0 is modified.

In the CP conserving MSSM scenarii, typically either the lightest CP-even Higgs state h has sufficient coupling to the Z^0 to be detected in Higgsstrahlung, or the pair production of h and A is possible. Thus the Higgs sector is experimentally accessible once the masses are light enough. In a CP violating MSSM scenario, choices of parameters exist for which the lightest Higgs mass eigenstate H_1 has a mass in the range accessible for LEP, but decouples completely from the Z^0, while the production of the second lightest Higgs H_2 also has small or vanishing cross-section, allowing for more complex experimental situations and reducing the accessible parameter regions. In large parts of the parameter space the production of both H_1 and H_2 in Higgsstrahlung is possible. The production of the heaviest Higgs state H_3 has no relevant cross-section at LEP energies in all production channels for the scenario under study.

The decay branching ratios are also changed with respect to a CP conserving scenario by the introduction of CP violation. The lightest mass eigenstate H_1 predominantly decays into $b\bar{b}$ where kinematically allowed, with only small contributions from $\tau^+\tau^-$ independent of the choice of parameters. The decay of the second lightest Higgs boson H_2 has a dominant contribution from $H_2 \to H_1H_1$, where kinematically allowed, otherwise it decays preferentially into $b\bar{b}$.

The interpretation of the OPAL Higgs searches uses the full dataset of OPAL from $\sqrt{s} = 91$ to $\sqrt{s} = 209$ GeV, corresponding to $\mathcal{L} \approx 858\,\mathrm{pb}^{-1}$. Searches for the Standard Model like production processes H_iZ^0, WW- and ZZ-fusion, presented in [4], are combined with searches for $e^+e^- \to H_iH_j$ processes, described in [5]. Extensions to these searches adapted to the CP violating MSSM are presented in this paper. The searches are predominantly sensitive to final states $b\bar{b}$ and $\tau^+\tau^-$, including decays of the type $H_2 \to H_1H_1 \to b\bar{b}b\bar{b}$. Flavour independent searches are included [5] and additional limits from decay mode independent searches [6] and Z^0 width constraints [7] are used. The Higgs boson masses, cross-sections and branching ratios of specified CP violating MSSM

scenarii are computed for each centre-of-mass energy using a modified version of the Monte Carlo generator HZHA03 [8], using SUBHPOLE [9] and CPHDECAY [10].

2 The CP-Violating Benchmark Scenario

The CP violating benchmark scenario CPX is described in [10]. Although a fundamental understanding of the origin of CP violation is still missing, many of the scenarii proposed to introduce CP violation are connected to the Higgs sector. In the SM for instance the CP symmetry is broken explicitly by complex Yukawa couplings of the Higgs boson to the quarks. Since the SM CP violation is not strong enough to describe the cosmological baryon asymmetry, other sources of CP violation in extensions to the SM are appealing. It has been realized recently that this is possible without violating experimental constraints from electric dipole moments [11].

In the MSSM, the Higgs potential is invariant under CP operations at tree level. However it is possible to explicitly or spontaneously break CP via radiative corrections, as described in [12]. In the CPX scenario, CP violation is introduced via the direct soft SUSY-breaking terms in the Lagrangian.

$$\mathcal{L}_{\text{Yuk}} = h_l \hat{H}_1^T i\tau_2 \hat{L}\hat{E} + h_d \hat{H}_1^T i\tau_2 \hat{Q}\hat{D} + h_u \hat{Q}^T i\tau_2 \hat{H}_2 \hat{U} - \mu \hat{H}_1^T i\tau_2 \hat{H}_2 \tag{1}$$

$$\begin{aligned} -\mathcal{L}_{\text{soft}} &= -\frac{1}{2}(m_{\tilde{g}} \lambda_{\tilde{g}}^a \lambda_{\tilde{g}}^a + \text{h.c.}) + \\ &\quad (h_l A_l \Phi_1^\dagger \tilde{L}\tilde{E} + h_d A_d \Phi_1^\dagger \tilde{Q}\tilde{D} - h_u A_u \Phi_2^T i\tau_2 \tilde{Q}\tilde{U} + \text{h.c.}) + \ldots \end{aligned} \tag{2}$$

The Yukawa part of the Lagrangian $\mathcal{L}_{\text{Yuk}}$ involves the Higgs superfield $\hat{H}$. The matter superfields include the left handed leptons $\hat{L}$, the right handed leptons $\hat{E}$, the left handed quarks $\hat{Q}$ and the right handed down- and up-type quarks $\hat{D}$ and $\hat{U}$. μ is the Higgs doublet mixing parameter, which plays an important role for the strength of the CP violating effects. The dimensionless parameters h_x denote the Yukawa couplings of the Higgs doublets to the matter fields, and $\mathcal{L}_{\text{Yuk}}$ corresponds to the couplings depicted in Fig. 1 (a).

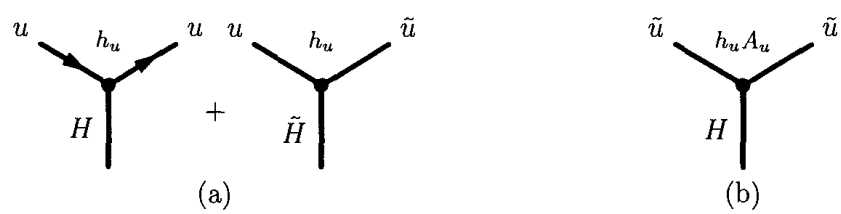


Figure 1: *(a) Examples for the SUSY conserving components of the Yukawa part of the MSSM Lagrangian. (b) SUSY violating trilinear boson coupling term in the direct soft SUSY breaking part of the Lagrangian.*

All Fields in equation (1) are superfields and therefore have correct transformation behavior under a SUSY transformation. Thus $\mathcal{L}_{\text{Yuk}}$ is invariant under SUSY transformations. The direct soft SUSY violating terms in $\mathcal{L}_{\text{soft}}$ correspond to the interaction depicted in Fig. 1 (b). They involve the supersymmetric partners of the SM fields, $\tilde{L}, \tilde{E}, \tilde{Q}, \tilde{D}$ and $\tilde{U}$. Φ_1 and Φ_2 are the bosonic Higgs doublet fields and A_x is the dimensionful trilinear boson coupling. $m_{\tilde{g}}$ is the gluino mass parameter and $\lambda_{\tilde{g}}^a$ denotes the gluino field. It can be shown that all other possible complex phases besides the phases of A_x and $m_{\tilde{g}}$ can be chosen to be 0 due to gauge symmetry [10].

The phases of A_x and $m_{\tilde{g}}$ introduce CP violation into the Higgs potential via loop effects, leading to sizeable off-diagonal contributions $\mathcal{M}^2_{SP}$ to the general Higgs boson mass matrix $\mathcal{M}^2$. Fig. 2 shows the CP conserving tree level contribution of the quartic Higgs couplings and the CP violating first order loop involving the dominant contribution from $\tilde{t}$-loops.

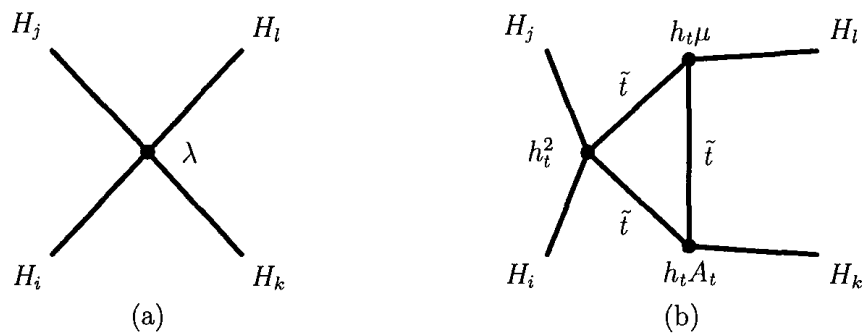


Figure 2: *CP conserving and CP violating four Higgs interactions. The H_i denote the neutral Higgs contents in the doublet fields. (a) The CP invariant tree level quartic Higgs coupling. (b) The complex coupling A_t introduces CP violation via first order loop effects, predominantly by* $\tilde{\mathrm{t}}$. *CP violating effects are dominated by graphs with one coupling involving h_tA_t and one coupling involving $h_t\mu$, since terms with A_t^2 can be real also for complex A_t.*

As a consequence the Higgs mass eigenstates $\mathrm{H}_1, \mathrm{H}_2$ and H_3 do not correspond anymore to the CP eigenstates h, H and A. This influences predominantly the couplings in the Higgs sector. In a CP conserving case, only h and H can couple to the Z^0 in Higgsstrahlung. In a CP violating case, all mass eigenstates can have CP-even contributions from h and H and therefore can couple to the Z^0 in Higgsstrahlung, depending on the choice of parameters. Fig. 3 shows the coupling of a mixed mass eigenstate H_i consisting of admixtures from h, H and A. Since the full field content of the mass eigenstate cannot couple anymore, the individual coupling of the mass eigenstate is reduced with respect to a CP conserving case.

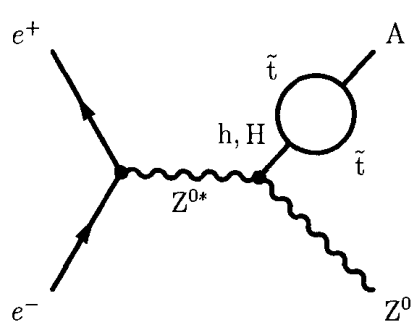


Figure 3: *An example diagram for the effective coupling of a Higgs mass eigenstate in Higgsstrahlung to the* Z^0. *The complete mass eigenstate* H_i *is composed of admixtures from* h, H *and* A. *Only the CP-even admixtures* h *and* H *couple to the* Z^0, *while the CP-odd* A *does not. Therefore the coupling of the mass eigenstate is reduced with respect to a CP conserving scenario.*

The size of CP violating off-diagonal contributions $\mathcal{M}^2_{SP,ij}$ to the Higgs boson mass

$\tan\beta$	=	0.4 – 40	ratio of Higgs v.e.v.
m_{H^+}	=	0 – 1 TeV	charged Higgs mass
μ	=	2 TeV	Higgs doublet mixing
m_{SUSY}	=	500 GeV	SUSY breaking scale = $m_{\tilde{q}}$
m_2	=	200 GeV	SU(2) gaugino mass matrix parameter
$\lvert A_q\rvert$	=	1 TeV	strength of trilinear coupling
$\arg(A_q)$	=	90°	phase of $A_q \Rightarrow$ CP-violation
$\lvert m_{\tilde{\mathrm{g}}}\rvert$	=	1 TeV	gluino mass
$\arg(m_{\tilde{\mathrm{g}}})$	=	90°	phase of $m_{\tilde{\mathrm{g}}} \Rightarrow$ CP-violation

Table 1: *Theoretical input parameters of the CPX scenario.* $\tan\beta$ *and* m_{H^+} *are being varied in the scan.*

matrix $\mathcal{M}^2$ and hence the size of CP violating effect scales qualitatively [10] as

$$\mathcal{M}^2_{SP,ij} \propto \frac{m_t^4}{v^2}\frac{\mathrm{Im}(\mu A_t)}{32\pi^2 m^2_{\mathrm{SUSY}}}. \tag{3}$$

Here m_t is the top mass, v is the SM Higgs vacuum expectation value and m_{SUSY} is the common soft SUSY-breaking scale. Large CP violating effects, and thus scenarii very dissimilar from CP conserving ones, are obtained if the SUSY-breaking scale m_{SUSY} is small and if the imaginary contribution to A_t is large.

When choosing the parameters, experimental constraints [13, 14] from electric dipole moment (EDM) measurements of the neutron and the electron have to be fulfilled. However, cancellations among different contributions to the EDM may naturally emerge [12], hence these measurements provide no strict exclusion. Therefore direct searches at LEP provide a better testing ground for a CP violating MSSM.

The parameters of the CPX scenario have been chosen such as to approximately fulfill the EDM constraints and to provide the most dissimilar features from a CP conserving scenario. The choice of parameters [10] is given in Table 1. μ and $|A_q|$ are taken to be large, while m_{SUSY} is relatively small, in order to increase CP violating effects according to equation (3). These parameters have a strong impact on the phenomenology of the model. Therefore the phase of A_q is varied explicitly in this analysis. The choice of m_2 and $m_{\tilde{\mathrm{g}}}$ has only a small effect on the Higgs sector phenomenology. Departing from the tradition in CP conserving scans, $\tan\beta$ and the mass of the charged Higgs boson m_{H^+} have been chosen as scan parameters, since m_{A} has no physical meaning anymore. The scan extends up to a value of $\tan\beta = 40$. In this region both H_1 and H_2 have a width below 1 GeV, negligible with respect to the experimental resolution of several GeV, depending on the individual search. Tachionic solutions of the model are discarded, making up most of the theoretically inaccessible region.

Additionally the complex phases of A_q and $m_{\tilde{g}}$ are varied from 0° to 90°. It will be seen that a scenario with $\arg(A_q) = 90°$ has very dissimilar features from a CP conserving case and therefore has good benchmark properties.

3 Contributing Search Channels

In the CPX scenario $\tan\beta$ and $H^{\pm}$ are varied. This yields a variety of different signatures, each dominating in different parts of the parameter space. Each of them requires different dedicated types of analysis. The following regions can be identified:

- *Region 1*: $\tan\beta < 4$
  Regime of the SM-like analysis, involving the search for Higgsstrahlung $e^+e^- \to Z^0H_1$ and boson fusion with $H_1 \to b\bar{b}$ and $H_1 \to \tau^+\tau^-$. No dedicated analysis for the CP violating MSSM is needed. The production of H_1 is stronger than the production of H_2 by almost one order of magnitude.

- *Region 2*: $4 < \tan\beta < 10$ and $m_{H_1} < 50$ GeV
  $g_{H_1Z^0Z^0}$ is close to 0, $m_{H_2} \approx 105$ GeV. Searches for the topology $e^+e^- \to Z^0H_2 \to Z^0H_1H_1 \to Z^0b\bar{b}\,b\bar{b}$ are needed. Additional contributions arise from Boson fusion processes $e^+e^- \to H_2e^+e^-$ and $e^+e^- \to H_2\nu\bar{\nu}$. Dedicated analyses for this channel are required.

- *Region 3*: $4 < \tan\beta < 10$ and $m_{H_1} > 50$ GeV
  $e^+e^- \to Z^0H_2 \to Z^0b\bar{b}$ is dominating with $m_{H_2} > 100$ GeV, which can be covered with SM type analyses. No dedicated analysis for the CP violating MSSM is needed.

- *Region 4*: $\tan\beta > 10$ and $m_{H_1} < 12$ GeV
  $e^+e^- \to H_1H_2$ pair production is strong and $H_1 \to \tau^+\tau^-$ is dominant below the $H_1 \to b\bar{b}$ threshold, while m_{H_2} is larger than 95 GeV. Due to the large mass difference between H_1 and H_2, which does not occur in the CP conserving benchmarks, a dedicated analysis is required.

- *Region 5*: $\tan\beta > 10$ and $m_{H_1} > 12$ GeV
  Pair production of the type $e^+e^- \to H_1H_2 \to b\bar{b}b\bar{b}$ is dominating. $m_{H_2} > 95$ GeV, large mass differences between H_1 and H_2. Due to the large mass difference between H_1 and H_2 a dedicated analysis for the CP violating MSSM is required.

Where searches for SM or CP conserving MSSM Higgs bosons are available [4,5], these can be reinterpreted within the CPX scenario. This is possible as only couplings and masses change while all other Higgs bosons properties used for the analyses (decay and production angular distributions) stay the same.

Dedicated analyses for signatures of the CPX scenario exist for decays of the type $H_2 \to H_1H_1$ (*Region 2*). The four-jet channel with its analysis originally dedicated to $HZ^0 \to b\bar{b}\,q\bar{q}$ is used to search for events of the type $H_2Z^0 \to H_1H_1Z^0 \to b\bar{b}\,b\bar{b}\,q\bar{q}$. Since, for heavy H_2 and light H_1, the $b\bar{b}$-systems from the H_1-decays are strongly boosted, the events exhibit a four-jet structure. Therefore the six-jet event can be grouped into four jets. The efficiency of this procedure is large at low m_{H_1} and high m_{H_2}, and decreases with decreasing mass difference $m_{H_2} - m_{H_1}$, since the event becomes more six-jet like.

The same technique is used in the missing energy channel $H\nu\bar{\nu} \to b\bar{b}\,\nu\bar{\nu}$. It is used to search for events of the type $e^+e^- \to H_2\nu\bar{\nu} \to H_1H_1 \to b\bar{b}\,b\bar{b}\,\nu\bar{\nu}$, where the b-jets are grouped as in the four-jet channel.

A CPX dedicated version of the $e^+e^- \to H_1H_2$-analysis exists for the large mass differences between H_1 and H_2 in the CPX scenario (*Region 5*). All analyses are based on cut-based preselections, followed by a multi-variant selection method. The dedicated analysis for signatures of the CPX scenario are described in [15].

4 Interpretation of the Search Results within the CPX Scenario

4.1 Combination Method

The statistical combination and interpretation of the individual search results of all OPAL neutral Higgs searches, as described in [4,5], has been performed using the calculation of confidence levels with tools described in [16], using methods described in [17,18].

In order to include both the production and decay of H_1 and of H_2 a separate interpretation is made for the primary production of H_1Z^0 and H_1H_2 with all subsequent decays, and another independent interpretation is performed for the primary production of H_2Z^0 and H_1H_2 with all subsequent decays. For the exclusion of each model point the interpretation using either H_1Z^0 or H_2Z^0 which gives the lower CL_s is chosen. A statistical combination of both H_1Z^0 and H_2Z^0 was not performed, since this was not possible for channels with Higgs boson test mass dependent background. If a model point is not excluded by either of these two combinations, constraints from the measurement of the Z^0-width are used. A point is regarded as excluded, if the following condition is true:

$$\sum_i \sigma_{H_iZ^0}(91.4\,\mathrm{GeV}) + \sum_{i,j} \sigma_{H_iH_j}(91.4\,\mathrm{GeV}) > \sigma_{lim} = \sigma_{Z^0 \to \mathrm{tot}} \frac{d\Gamma(Z^0 \to X)}{\Gamma(Z^0 \to X)}$$

using results from [7]. The nominal LEP1 centre-of-mass energy of $\sqrt{s} = 91.4$ GeV is used. $\Gamma(Z^0 \to X)$ is the total Z^0 width, and $d\Gamma(Z^0 \to X) = 6.5$ MeV is the total width allowed above the SM expectation (obtained from ZFITTER [19]) at 95% CL. For model points still not excluded the results of a decay mode independent search for $e^+e^- \to HZ^0$ [6] have been used. Points are regarded as excluded if

$$\sigma_{H_iZ^0} > k(m_{H_i})\, \sigma_{hZ^0,SM} \qquad \text{with} \qquad m_{H_i} = m_h$$

is fulfilled. This has been tested for $H_i = H_1$ and $H_i = H_2$ and at $\sqrt{s} = 91.4$, 183 and 206 GeV, where $k(m_{H_i})$ is the scale factor for the SM Higgs production cross-section excluded at the 95% CL from this search. The use of Z^0 width constraints and decay mode independent analyses is especially helpful for the range $m_{H_1} < 6$ GeV. In the scenario under study all points excluded by the Z^0-width constraint are also excluded by decay mode independent searches.

4.2 Results

The combined exclusion result for the CPX scenario is shown in Fig. 4. Fig. 4 (a) shows both the expected and observed 95% CL exclusion areas in the plane of m_{H_1} and m_{H_2}. A model point is regarded as experimentally excluded at the 95% level if its confidence level (CL) fulfills $CL = CL_s < 0.05$ [17,18]. The area of heavy m_{H_2} gives the parameter space where H_1 resembles the SM Higgs boson with very little effect from CP violation. The limit on the allowed mass of H_1 for high m_{H_2} is found to be $m_{H_1} > 108$ GeV or $m_{H_1} < 12$ GeV (95% CL). In the region below $m_{H_2} \approx 130$ GeV CP violating effects play a major role.

Fig. 4 (b) shows the 95% CL exclusion areas in the parameter space of $\tan\beta$ and m_{H_2}. One can see that $\tan\beta < 2.9$ is excluded in this model. The band at $\tan\beta < 2.9$ is excluded by searches for the SM-like H_1, while the band at $\tan\beta > 10$ and $m_{H_2} < 120$ GeV is excluded by searches for Z^0H_2 and H_1H_2 topologies.

Fig. 4 (c) displays the parameter space of $\tan\beta$ and m_{H_1}. Exclusion is reached for $\tan\beta < 2.9$ and $m_{H_1} < 108$ GeV in a SM like regime. For $4 < \tan\beta < 10$, Z^0H_2 production is dominant. Exclusion can only be achieved where $H_1H_2 \to b\bar{b}\,b\bar{b}$ is strong enough to give an additional contribution, at $m_{H_1} = 50 - 60$ GeV. Model points with $m_{H_1} < 50$ GeV are not excluded due to the $Z^0H_2 \to Z^0H_1H_1$ channel, while expected exclusion can be achieved for $m_{H_1} > 60$ GeV where $Z^0H_2 \to Z^0b\bar{b}$ is kinematically accessible. The large difference between the expected and observed exclusion regions in the area of $4 < \tan\beta < 10$ and $m_{H_1} > 60$ GeV is mainly due to a less than 2σ excess of the 1999 missing energy channel near the kinematic limit.

In Fig. 4 (d) the exclusion area is shown in the parameter space of the theoretical input parameters $\tan\beta$ and $m_{H^\pm}$, which are varied during the scan. Since the scenario yields $m_{H_2} \approx m_{H^\pm}$ in most cases, this is very similar to Fig. 4 (b). Note that no searches for charged Higgs bosons are included.

In the region around $\tan\beta \approx 6$, $m_{H_1} \approx 40$ GeV and $m_{H_2} \approx 105$ GeV an excess of candidate events is observed from the 1999 $Ah \to b\bar{b}\tau^+\tau^-$ and the 1999 $Ah \to b\bar{b}b\bar{b}$ channels, with small contributions from the 2000 missing energy channel. This excess was not observed in previous MSSM scans, since this mass combination of H_1 and H_2 is not present in CP conserving models. In this region there is no sensitivity to the CPX scenario, since cross-sections are small. The probability of an excess of the measured order of magnitude to show up from a fluctuation of the background somewhere in the accessible m_{H_1}, m_{H_2} plane was estimated to be around 5% for the 1999 $Ah \to b\bar{b}\tau^+\tau^-$ channel and the 1999 $Ah \to b\bar{b}b\bar{b}$ channel individually.

Since the CPX scenario has maximum CP violation due to maximum imaginary parts of $A_{t,b}$ and $m_{\tilde{g}}$, the effect of different phases has been surveyed using scans with different values of $\arg A_{t,b} = \arg m_{\tilde{g}}$ from 90° to 0°, gradually decreasing CP violating effects. Fig. 5 shows exclusion regions in the parameter space of $\tan\beta$ and m_{H_1} for $\arg A_{t,b} = \arg m_{\tilde{g}} = 90°, 60°, 30°$ and 0°. At $\arg A_{t,b} = \arg m_{\tilde{g}} = 30°$ and at $\arg A_{t,b} = \arg m_{\tilde{g}} = 0°$ all areas for low m_{H_1} and low $\tan\beta$ are excluded. The exclusion for the maximally CP violating scenario CPX with $\arg A_{t,b} = \arg m_{\tilde{g}} = 90°$ is very different from the exclusion of a CP conserving scenario with $\arg A_{t,b} = \arg m_{\tilde{g}} = 0°$. A variation of the second main parameter governing the size of CP violating effects in equation (3), m_{SUSY}, has similar effects on the exclusion to those of a variation of $\arg A_{t,b} = \arg m_{\tilde{g}}$.

5 Summary

The searches for Higgs bosons at $\sqrt{s} = 91$ GeV to $\sqrt{s} = 209$ GeV have been interpreted within a CP violating MSSM scenario. The searches have not revealed any significant excess over the expected backgrounds. For the first time limits on the masses of Higgs bosons and on $\tan\beta$ have been set in a CP violating interpretation. Due to the presence of two different Higgs bosons coupling to the Z^0 simultaneously, due to reduced couplings of the individual Higgs bosons to the Z^0, and due to regions in the parameter space with dominant $H_2 \to H_1H_1$ decay channels the obtained exclusion areas are smaller than the exclusion areas of the standard CP conserving MSSM interpretations. For the CPX scenario, $\tan\beta < 2.9$ can be excluded at 95% confidence level. No $\tan\beta$-independent limit on m_{H_1} can be set. The complete interpretation is described in [15].

References

[1] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32 [JETP Lett. **5** (1967) 24].

[2] E. Kolb, M. Turner, *The Early Universe*, Addison-Wesley, Redwood City, CA (1990)

[3] M. Carena, J. M. Moreno, M. Quiros, M. Seco and C. E. Wagner, Nucl. Phys. B **599** (2001) 158

[4] G. Abbiendi *et al.* [OPAL Collaboration], Phys. Lett. B **499** (2001) 38

[5] G. Abbiendi *et al.* [OPAL Collaboration], OPAL Physics Note PN472, 27 February 2001

[6] G. Abbiendi *et al.* [OPAL Collaboration], arXiv:hep-ex/0206022, to be published in Eur. Phys. J.

[7] The LEP Collaborations ALEPH, DELPHI, L3, OPAL, the LEP Electroweak Working Group, and the SLD Heavy Flavour and Electroweak Groups, arXiv:hep-ex/0112021.

[8] HZHA Generator: P. Janot in *Physics at LEP2,* edited by G. Altarelli, T. Sjöstrand and F. Zwirner, CERN 96-01 Vol. 2 p. 309.
S. Mrenna, private communication

[9] M. Carena, M. Quiros and C. E. Wagner, Nucl. Phys. B **461** (1996) 407

[10] M. Carena, J. R. Ellis, A. Pilaftsis and C. E. Wagner, Phys. Lett. B **495** (2000) 155

[11] A. Pilaftsis, Phys. Rev. D **58** (1998) 096010

[12] M. Carena, J. R. Ellis, A. Pilaftsis and C. E. Wagner, Nucl. Phys. B **586** (2000) 92

[13] E. D. Commins, S. B. Ross, D. DeMille and B. C. Regan, Phys. Rev. A **50** (1994) 2960.

[14] P. G. Harris *et al.*, Phys. Rev. Lett. **82** (1999) 904.

[15] G. Abbiendi *et al.* [OPAL Collaboration], OPAL Physics Note PN505, 15th July 2002,
`http://opalinfo.cern.ch/opal/doc/physnote/html/pn505.html`

[16] T. Junk, Nucl. Instrum. Meth. A **434** (1999) 435

[17] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. 2, Ch. 23 5th Ed., Oxford University Press, New York, (1991).

[18] ALEPH, DELPHI, L3, OPAL Collaborations, and the LEP working group for Higgs boson searches, CERN-EP/2000-055 (2000)

[19] D. Y. Bardin, P. Christova, M. Jack, L. Kalinovskaya, A. Olchevski, S. Riemann and T. Riemann, Comput. Phys. Commun. **133** (2001) 229

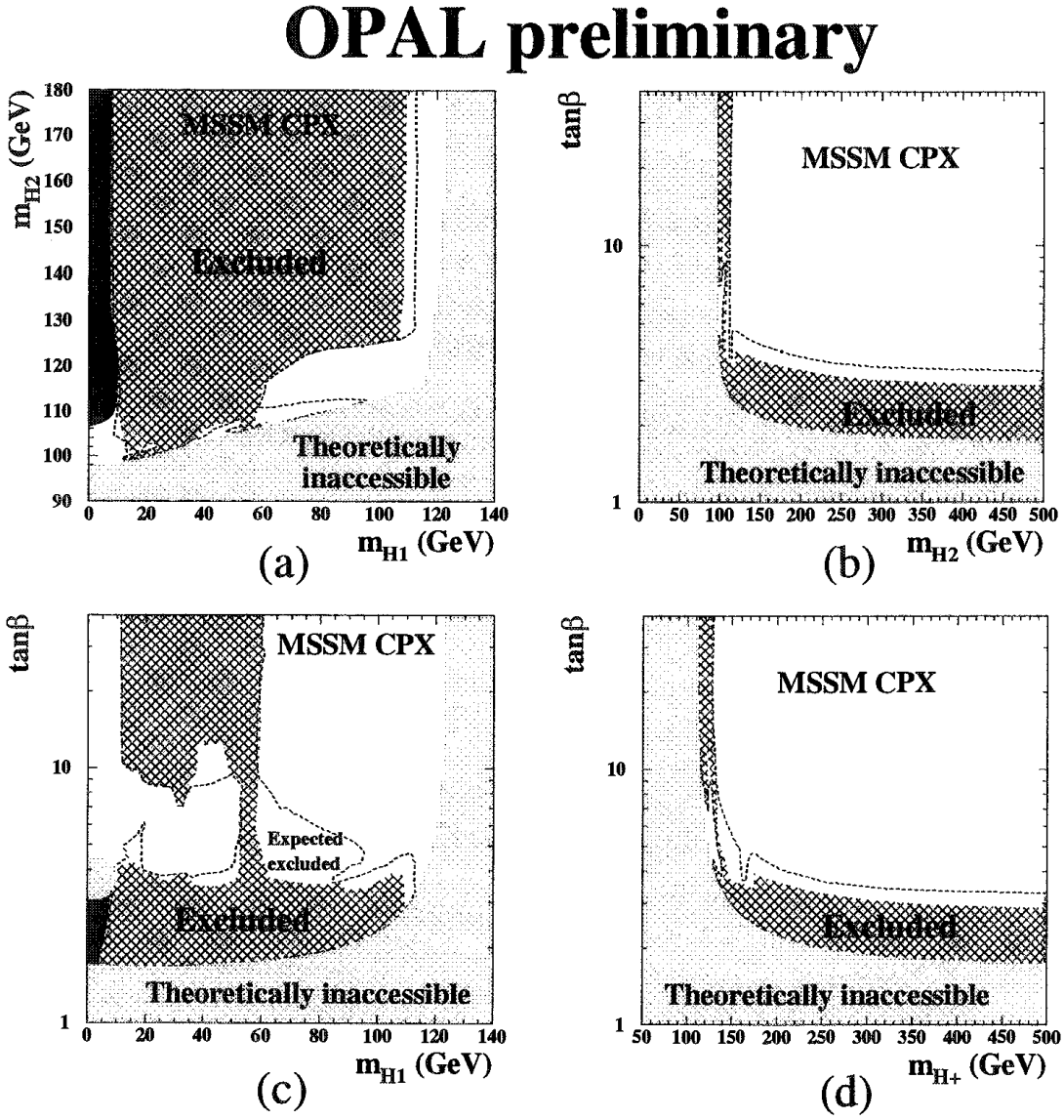


Figure 4: *The CPX MSSM* 95 % CL *exclusion areas. Excluded regions are shown for (a) the* (m_{H_1},m_{H_2}) *plane, (b) the* (m_{H_2},tanβ) *plane, (c) the* (m_{H_1},tanβ) *plane and (d) the* ($m_{H^\pm}$,tanβ) *plane. The observed excluded region is hatched, the expected excluded region is indicated by the dashed line and the theoretically inaccessible region is light shaded. Regions excluded either by* Z^0 *width constraints or by decay mode independent searches are dark shaded. In (b) and (d) the area excluded by* Z^0 *width constraints or by decay independent searches is too small to be displayed.*

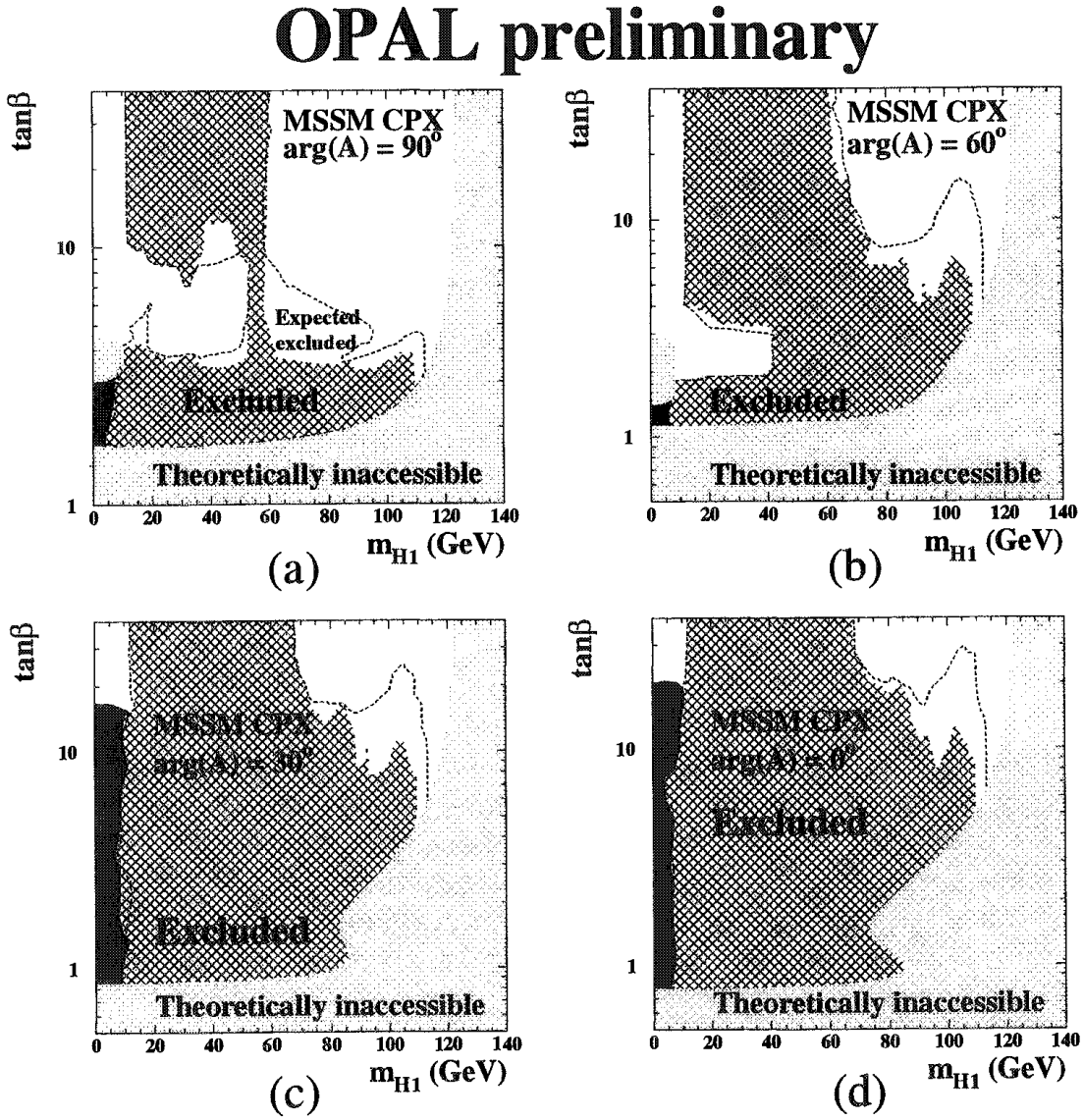


Figure 5: *The CPX MSSM* 95 % CL *exclusion areas in the* ($m_{\mathrm{H_1}}$,tanβ) *plane, using scans with (a)* $\arg A_{\mathrm{t,b}} = \arg m_{\tilde{\mathrm{g}}} = 90°$, *(b)* $\arg A_{\mathrm{t,b}} = \arg m_{\tilde{\mathrm{g}}} = 60°$, *(c)* $\arg A_{\mathrm{t,b}} = \arg m_{\tilde{\mathrm{g}}} = 30°$ *and (d)* $\arg A_{\mathrm{t,b}} = \arg m_{\tilde{\mathrm{g}}} = 0°$. *While the CP violating phases decrease, effects from CP violation like the strong* $\mathrm{H_2} \rightarrow \mathrm{H_1H_1}$ *contribution vanish. The observed excluded region is hatched, the expected excluded region is indicated by the dashed line and the theoretically inaccessible region is light shaded. Regions excluded either by* $\mathrm{Z^0}$ *width constraints or by decay mode independent searches are dark shaded.*

Application of the Large–N_c limit to a Chiral Lagrangian with Resonances

Oscar Catà*

Abstract

It is shown that the implementation of the Large–N_c approximation helps to get insight into the structure of, in principle, any QCD-like theory. As an example, we will compute the NLO corrections to L_{10} in the chiral limit with a Lagrangian with Resonances.

1. Introduction

The QCD Lagrangian is assumed to encode the whole description of the strong interactions. At least, QCD has proven to be successful in describing its perturbative regime, corresponding to the high energy limit of the theory. However, a useful description of the strong interactions at low energies is still needed. The problem is two-fold: first of all, the strong coupling constant α_s blows up, thus excluding all perturbative techniques. Secondly, quark and gluon fields are not the appropriate variables at low energies, where one observes hadrons, and there is not even a clue as how to connect these two sets of variables, or, in other words, understand what lies behind the process of hadronisation.

The efforts devoted to these topics for the last 40 years or so have led to remarkable progress. On the one hand, the low energy regime of QCD has been successfully approached by means of Chiral Perturbation Theory. On the other hand, Lattice QCD and Large–N_c QCD are two firmly-established attempts to fill the gap between low and high energies.

In particular, the Large-N_c limit is currently profusely used not only to establish a link between String theory and the Standard Model, but also to extract information from the Standard Model itself. Our aim in this paper is to show how the Large–N_c approximation can be used to provide a solid framework to QCD-like theories, both by enabling the use of perturbation theory consistently and by constraining the (often) too large number of free parameters.

*UAB & IFAE, Edifici Cn, 08193 Bellaterra (Barcelona), Spain

2. Chiral Perturbation Theory

Chiral Perturbation Theory (χPT) is the Effective Field Theory (EFT) of QCD in the sector of the light quarks (u,d,s). It describes the interactions of the pseudoGoldstone bosons[1] under the spontaneously broken chiral symmetry

$$SU_L(3) \times SU_R(3) \to SU_V(3) \tag{1}$$

as a power series in momenta and masses. The first two terms read [1]:

$$\mathcal{L}_\chi^{(2)}(U, DU) = \frac{f_\pi^2}{4}\langle D_\mu U^\dagger D^\mu U + U^\dagger \chi + \chi^\dagger U\rangle \tag{2}$$

$$\begin{aligned}
\mathcal{L}_\chi^{(4)}(U, DU) &= L_1\langle D_\mu U^\dagger D^\mu U\rangle^2 + L_2\langle D_\mu U^\dagger D_\nu U\rangle\langle D^\mu U^\dagger D^\nu U\rangle + \\
&+ L_3\langle D_\mu U^\dagger D^\mu U D_\nu U^\dagger D^\nu U\rangle + L_4\langle D_\mu U^\dagger D^\mu U\rangle\langle U^\dagger\chi + \chi^\dagger U\rangle + \\
&+ L_5\langle D_\mu U^\dagger D^\mu U(U^\dagger\chi + \chi^\dagger U)\rangle + L_6\langle U^\dagger\chi + \chi^\dagger U\rangle^2 + \\
&+ L_7\langle U^\dagger\chi - \chi^\dagger U\rangle^2 + L_8\langle\chi^\dagger U\chi^\dagger U + U^\dagger\chi U^\dagger\chi\rangle - \\
&- iL_9\langle F_R^{\mu\nu} D_\mu U D_\nu U^\dagger + F_L^{\mu\nu} D_\mu U^\dagger D_\nu U\rangle + L_{10}\langle U^\dagger F_R^{\mu\nu} U F_{L\mu\nu}\rangle + \\
&+ H_1\langle F_{R\mu\nu}F_R^{\mu\nu} + F_{L\mu\nu}F_L^{\mu\nu}\rangle + H_2\langle\chi^\dagger\chi\rangle
\end{aligned} \tag{3}$$

f_π being the pion decay constant in the chiral limit, $f_\pi = 87\pm3.5$ MeV, and the L_i low-energy couplings encoding the information of the heavy degrees of freedom (hadrons lying above the pions) compatible with their quantum numbers.

As with any EFT, the previous expansion is only valid up to some threshold ($\Lambda_{\chi PT} \sim 1\,\mathrm{GeV}$). In going to higher energies one has to resort to more general frameworks such as the Lattice or Large–N_c. Still, not so ambitious an approach would be to build up another EFT by inserting new dynamical degrees of freedom (hadrons), thus taking over χPT at a threshold μ^* and covering a wider range of energies.

3. Adding Resonances

The key ingredient is to notice that the L_i couplings entering the chiral Lagrangian are also order parameters of Chiral Symmetry Breaking, i.e., they do not get contributions from the continuum of QCD. Therefore, there is hope to saturate them with a discrete number of resonances, the most favoured ones being the lowest-lying hadrons. This is supported by the following argument, taken from [2]: consider a set of resonances, whose propagators are given generically by

$$\frac{1}{q^2 - M_{\mathcal{R}}^2} \tag{4}$$

[1] i.e., π, k and η. To be generically referred to as pions thereafter.

At low energies, they can be expanded in powers of q^2, contributing to L_i like

$$L_i \sim \sum_{\mathcal{R}} \frac{F_{\mathcal{R}}^2}{M_{\mathcal{R}}^2} \tag{5}$$

Thus, the natural choice would be to insert the lightest hadrons, since they bear the biggest impact on L_i. One such proposal is [3]:

$$\begin{aligned}
\mathcal{L}_{\mathcal{R}}(V,A,S,P) &= \sum_{R=V,A} \left[-\frac{1}{2}\langle \nabla^\lambda R_{\lambda\mu} \nabla_\nu R^{\nu\mu} - \frac{1}{2} M_R^2 R_{\mu\nu} R^{\mu\nu} \rangle \right] - \\
&- \sum_{R=V,A} \left[\frac{1}{2} \partial^\lambda R_{1,\lambda\mu} \partial_\nu R_1^{\nu\mu} + \frac{M_{R_1}^2}{4} R_{1,\mu\nu} R_1^{\mu\nu} \right] \\
&+ \sum_{R=S,P} \left[\frac{1}{2} \langle \nabla^\mu R \nabla_\mu R - M_R^2 R^2 \rangle + \frac{1}{2} \left(\partial^\mu R_1 \partial_\mu R_1 - M_{R_1}^2 R_1^2 \right) \right] \\
&+ \frac{F_V}{2\sqrt{2}} \langle V_{\mu\nu} f_+^{\mu\nu} \rangle + i \frac{G_V}{\sqrt{2}} \langle V_{\mu\nu} u^\mu u^\nu \rangle + \frac{F_A}{2\sqrt{2}} \langle A_{\mu\nu} f_-^{\mu\nu} \rangle \\
&+ c_d \langle S u_\mu u^\mu \rangle + c_m \langle S \chi_+ \rangle + \tilde{c}_d S_1 \langle u_\mu u^\mu \rangle + \tilde{c}_m S_1 \langle \chi_+ \rangle \\
&+ i d_m \langle P \chi_- \rangle + i \tilde{d}_m P_1 \langle \chi_+ \rangle
\end{aligned} \tag{6}$$

to be added to the chiral Lagrangian, such that the enlarged chiral theory looks like:

$$\mathcal{L}_{\chi\mathcal{R}}(V,A,S,P,U,DU) = \mathcal{L}_\chi^{(2)}(U,DU) + \tilde{\mathcal{L}}_\chi^{(4)}(U,DU) + \mathcal{L}_{\mathcal{R}}(V,A,S,P) \tag{7}$$

One should be careful in distinguishing L_i from $\tilde{L}_i$: the former contain information about resonances lying above the pions, whereas the latter have the information of the resonances lying above the V, A, S and P multiplets[2,3].

4. Integrating out Resonances

Once the Lagrangian $\mathcal{L}_{\chi\mathcal{R}}$ is chosen, the natural step is to integrate out the resonances. We are left with the original chiral lagrangian, the coefficients being functions of hadronic parameters (masses and couplings), to be related to the L_i's through a matching procedure. The results are as follows [3]:

$$\begin{array}{lcccccc}
L_1 = & \frac{G_V^2}{8M_V^2} & + & - \frac{c_d^2}{6M_S^2} & + & \frac{\tilde{c}_d^2}{2M_{S_1}^2} & + & + \\
L_2 = & \frac{G_V^2}{4M_V^2} & + & + & + & & + & + \\
L_3 = & -\frac{3G_V^2}{4M_V^2} & + & + \frac{c_d^2}{2M_S^2} & + & & + & +
\end{array}$$

[2]Caveat: Eq.(6) is not intended to be a complete Lagrangian in any sense. It is the most general *chirally-symmetric* theory with *lowest-lying* hadronic multiplets and interaction terms *linear* in the resonance fields. But obviously, there is no compelling reason to do so. Indeed there are several other proposals for such an EFT.

[3]V, A, S and P stand for Vector, Axial, Scalar and Pseudoscalar sectors.

$$
\begin{array}{lcccccccccc}
L_4 = & & + & & - & \frac{c_d c_m}{3M_S^2} & + & \frac{\tilde{c}_d \tilde{c}_m}{M_{S_1}^2} & + & & + \\
L_5 = & & + & & + & \frac{c_d c_m}{M_S^2} & + & & + & & + \\
L_6 = & & + & & - & \frac{c_m^2}{6M_S^2} & + & \frac{\tilde{c}_m^2}{2M_{S_1}^2} & + & & + \\
L_7 = & & + & & + & & + & & + & \frac{d_m^2}{6M_P^2} & - & \frac{\tilde{d}_m^2}{2M_{\eta_1}^2} \\
L_8 = & & + & & + & \frac{c_m^2}{2M_S^2} & + & & - & \frac{d_m^2}{2M_P^2} & + \\
L_9 = & \frac{F_V G_V}{2M_V^2} & + & & + & & + & & + & & + \\
L_{10} = & -\frac{F_V^2}{4M_V^2} & + & \frac{F_A^2}{4M_A^2} & + & & + & & + & & +
\end{array}
\tag{8}
$$

Upon comparison with experiment one gets[4]:

$L_i(M_\rho)$	experimental values ($\times 10^{-3}$)	V	A	S	S_1	η_1	total
L_1	0.7 ± 0.3	0.6	0	-0.2	0.2	0	0.6
L_2	1.3 ± 0.7	1.2	0	0	0	0	1.2
L_3	-4.4 ± 2.5	-3.6	0	0.6	0	0	-3.0
L_4	-0.3 ± 0.5	0	0	-0.5	0.5	0	0.0
L_5	1.4 ± 0.5	0	0	1.4	0	0	1.4
L_6	-0.2 ± 0.3	0	0	-0.3	0.3	0	0.0
L_7	-0.4 ± 0.15	0	0	0	0	-0.3	-0.3
L_8	0.9 ± 0.3	0	0	0.9	0	0	0.9
L_9	6.9 ± 0.2	6.9	0	0	0	0	6.9
L_{10}	-5.2 ± 0.3	-10.0	4.0	0	0	0	-6.0

which shows a remarkable agreement, strongly supporting resonance saturation. Still, there is a flaw in the argument: the value $\mu^* = M_\rho$ is, though very suggestive, completely arbitrary. The right way to proceed would be to determine μ^* from $\mathcal{L}_{\chi R}$ at one loop level. However, contrary to $\mathcal{L}_\chi$, $\mathcal{L}_{\chi R}$ has no obvious expansion parameter. One way out is the following: since $\mathcal{L}_{\chi R}$ is meant to be an EFT of QCD, we can resort to Large–N_c methods.

5. The Large–N_c limit[5]

As mentioned above, the Large–N_c limit of QCD is a powerful non-perturbative tool to explore QCD and, in a broader sense, the Standard Model. The crucial point is to realise that, surprisingly as it is, a $SU(N_c)$ gauge theory of quarks and gluons is dynamically simplified if one takes N_c to infinity while keeping $\alpha_s N_c$ fixed. This last condition ensures that the theory is well-behaved and, even further, that it admits a systematic expansion in powers of $1/N_c$. Among the several additional features of the Large–N_c limit we shall be interested only in a few ones:

[4]Numerical values for masses and couplings are given in [3].

[5]The interested reader is referred to [4].

- The dynamical simplicity shows up in the fact that it is a theory of free, stable and non-interacting mesons, where the interactions are restricted to tree-level exchanges of physical mesons. This in turn implies that the only singularities of Green functions are poles at the masses of physical mesons.

- The mass difference between singlets and octets is $1/N_c$-suppressed, so that in the $N_c \to \infty$ limit hadronic multiplets are nonets.

Despite the fact that even the leading term in the $1/N_c$ expansion entails an unsurmountable calculational effort, those qualitative features highlighted above provide, as we will see in a moment, a great deal of information.

6. Large–N_c applied to the model

Both the chiral Lagrangian $\mathcal{L}_\chi$ and the Resonance Lagrangian $\mathcal{L}_\mathcal{R}$ can be seen in the light of Large–N_c[6]. The first step is to know the N_c-scaling of the parameters, which can be determined straightforwardly by a matching procedure.

$$\begin{array}{rl} L_{10} & \mathcal{O}(N_c) \\ f_\pi, F_V, F_A, G_V, c_d, \tilde{c}_d & \mathcal{O}(\sqrt{N_c}) \\ M_V, M_A, M_S, M_{S_1} & \mathcal{O}(1) \end{array} \tag{9}$$

The next step is to put singlets and octets together into nonets. In the chiral limit it simply amounts to replace [7]

$$U \to U\, e^{-i\frac{\sqrt{2}}{\sqrt{3}}\frac{\eta_1}{f_\pi}} \tag{10}$$

while the other terms in $\mathcal{L}_{\chi R}$ remain unaltered. This furnishes $\mathcal{L}_{\chi R}$ with a consistent Large–N_c framework, as first shown in [6]. As a by-product, the very particular singularity structure of Large–N_c Green functions provides a set of constraining equations for the hadronic parameters[7]

$$F_V = 2G_V = \sqrt{2}F_A = \sqrt{2}f_\pi; \qquad M_A = \sqrt{2}M_V = 4\pi f_\pi \left(\frac{2\sqrt{6}}{5}\right)^{1/2} \tag{11}$$

which induce a *prediction* for the values of L_i.

$$6L_1 = 3L_2 = -\frac{8}{7}L_3 = \frac{3}{4}L_9 = -L_{10} = \frac{3}{8}\frac{f_\pi^2}{M_V^2} = \frac{15}{8\sqrt{6}}\frac{1}{16\pi^2} \tag{12}$$

[6]The reader is referred to [5] and [6] for details.

[7][8] and references therein.

This is remarkable, since it provides a testing ground for the accuracy of the Large–N_c approximation. Alternatively, Eq.(12) can be read as the imprint of a (still) hidden dynamical symmetry of the Large–N_c limit. To make the point clearer, one can set out an analogy with $SU(5)$ GUT. The parallelism is as follows:

SU(5)	
$\sin^2\theta_W = \frac{3}{8}$	$\alpha_{SU(3)}(M_{GUT}) = \alpha_{SU(2)}(M_{GUT}) = \alpha_{U(1)}(M_{GUT})$

Hypothetical Symmetry of Large–N_c	
$L_{10} = -\frac{15}{8\sqrt{6}}\frac{1}{16\pi^2}$	$6L_1 = 3L_2 = -\frac{8}{7}L_3 = \frac{3}{4}L_9 = -L_{10} = \frac{3}{8}\frac{f_\pi^2}{M_V^2} = \frac{15}{8\sqrt{6}}\frac{1}{16\pi^2}$

Getting insight into this underlying symmetry would be of great interest in trying to understand what lies behind the Large–N_c approximation.

7. An Example: L_{10}

In this section we will illustrate how theories equipped with the Large–N_c approximation work with a particularly simple application: the determination of L_{10} with $\mathcal{L}_{\chi R}$ at one loop level[8,9]. Additionally, this simple example may help clarify whether resonance saturation, as advocated in [3], survives beyond the leading order. The way to proceed is to integrate out resonances in $\mathcal{L}_{\chi R}$ up to one loop. Equivalently, one can select an appropriate Green function (*i.e.*, L_{10} should appear without any mixing with the remaining L_i) and compute it to one loop with both Lagrangians at hand, $\mathcal{L}_\chi$ and $\mathcal{L}_{\chi R}$. For the present work, the latter method will be adopted.

Let us start with the two-point function

$$\Pi^{\mu\nu}_{LR}(q)\delta_{ab} = 2i\int d^4x\, e^{iq\cdot x}\langle 0|\mathrm{T}\left(L^\mu_a(x)R^\nu_b(0)^\dagger\right)|0\rangle\,, \tag{13}$$

where

$$R^\mu_a\left(L^\mu_a\right) = \bar{q}(x)\gamma^\mu\,\frac{\lambda_a}{\sqrt{2}}\,\frac{(1\pm\gamma_5)}{2}\,q(x)\,, \tag{14}$$

are the helicity currents of QCD. In the chiral limit, Lorentz invariance implies

$$\Pi^{\mu\nu}_{LR}(Q^2) = (g^{\mu\nu}q^2 - q^\mu q^\nu)\Pi_{LR}(Q^2) \tag{15}$$

Moreover, $\Pi_{LR}(Q^2)$, computed with χPT, admits the following power expansion in Q^2

$$\Pi_{LR}(Q^2) = \frac{f_\pi}{Q^2} + 4L_{10} + \mathcal{O}(Q^2) \tag{16}$$

[8]We will follow closely [9] from here onwards. For further explanation the reader is referred to [9] and references therein.

[9]We shall adopt the $1/N_c$ expansion throughout as our underlying power expansion.

Therefore, a suitable Green function to determine L_{10} is given by

$$\frac{1}{4}\frac{d}{dQ^2}(Q^2\Pi_{LR}(Q^2))_{Q^2=0} \tag{17}$$

One can now compute (17) up to one loop with $\mathcal{L}_\chi$ and $\mathcal{L}_{\chi R}$, thereby getting automatically the matching conditions. We will skip the calculations and refer the reader to [9] for further details. The expressions one ends up with are

$$L_{10} = \frac{F_A^2}{4M_A^2} - \frac{F_V^2}{4M_V^2} + \tilde{L}_{10} = -\frac{1}{4}\left(\frac{15}{32\pi^2\sqrt{6}}\right) + \tilde{L}_{10} \tag{18}$$

at tree level, which is in agreement with the last line of (8), and

$$\begin{aligned}
4\,L_{10}^r(\mu) \;=\; & -\frac{15}{32\pi^2\sqrt{6}} \\
& -\frac{3}{2}\frac{F_A^2}{f_\pi^2}\frac{1}{(4\pi)^2}\left(\frac{1}{2}-\log\frac{M_A^2}{\mu^2}\right)+\frac{3}{2}\frac{F_V^2}{f_\pi^2}\frac{1}{(4\pi)^2}\left(\frac{1}{2}-\log\frac{M_V^2}{\mu^2}\right) \\
& -\frac{5}{(4\pi)^2}\frac{G_V^2}{f_\pi^2}\left(-\frac{17}{30}-\log\frac{M_V^2}{\mu^2}\right) \\
& +\frac{3}{2}\frac{1}{(4\pi)^2}\left(-\frac{1}{3}-\log\frac{M_A^2}{\mu^2}\right)+\frac{3}{2}\frac{1}{(4\pi)^2}\left(-\frac{1}{3}-\log\frac{M_V^2}{\mu^2}\right) \\
& -\frac{4}{3}\left(\frac{\tilde{c}_d}{f_\pi}\right)^2\frac{1}{(4\pi)^2}\left(\frac{1}{6}+\log\frac{M_{S_1}^2}{\mu^2}\right)-\frac{10}{9}\left(\frac{c_d}{f_\pi}\right)^2\frac{1}{(4\pi)^2}\left(\frac{1}{6}+\log\frac{M_S^2}{\mu^2}\right) \\
& +\frac{1}{2}\frac{1}{(4\pi)^2}\left(1+\log\frac{M_S^2}{\mu^2}\right)- \\
& -\frac{4}{9}\left(\frac{c_d}{f_\pi}\right)^2\frac{1}{(4\pi)^2}\left[\frac{1}{6}+\log\frac{M_S^2}{\mu^2}+2B+2B^2-(2B^3+3B^2)\log\frac{M_S^2}{M_{\eta_1}^2}\right] \\
& +4\,\tilde{L}_{10}^r(\mu)\;,
\end{aligned} \tag{19}$$

up to one loop[10]. Notice that we have kept the $\tilde{L}_{10}$ coupling all the way. Now we are in a position to assess whether resonance saturation takes place or not. L_{10}^r is plotted in figure 1 (solid lines), together with the L_{10} running (dashed lines) and L_{10} tree level value (dot-dashed lines). Since the slope of L_{10}^r does not follow that of runnning L_{10}, one concludes that $\tilde{L}_{10}^r$ cannot be dropped (in fact, its μ-dependence has to supply the right slope for L_{10}^r). However, there can still be a value for μ (say μ^*) at which resonance saturation is not spoiled (intersection of solid and dashed lines). Yet this value lies at $\mu^* \sim 380$ MeV, too low a value to be realistic[11].

[10] $B = \frac{M_{\eta_1}^2}{(M_S^2 - M_{\eta_1}^2)}$.

[11] Remember that this μ^* was to be interpreted as the threshold for $\mathcal{L}_{\chi R}$ to take over the chiral Lagrangian.

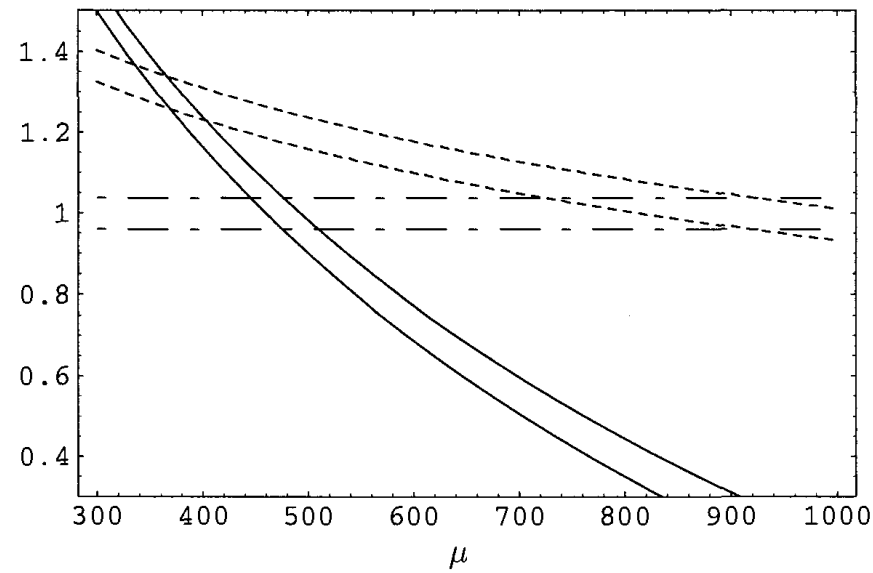


Figure 1: $L_{10}(\mu)$ as a function of μ. The solid lines represent $L_{10}^r(\mu)$ as given by Eq.(19) when imposing $\tilde{L}_{10}^r = 0$. The dashed curves are the running of $L_{10}(\mu)$ as dictated by χPT. For comparison the tree level contribution to $L_{10}(\mu)$ is also given. All three curves are normalised to the central value of the tree level resonance contribution.

Therefore, the Large–N_c approximation has allowed us to survey $\mathcal{L}_{\chi R}$ at one loop level, showing, through a non-zero $\tilde{L}_{10}$, that for resonance saturation to be fulfilled one has to supply additional structure (*i.e.*, new couplings and most likely new resonance fields) to our starting Lagrangian $\mathcal{L}_{\chi R}$.

8. Conclusions

Taking Large–N_c for granted, any QCD-like theory admits, in principle, a perturbative series expansion in powers of $1/N_c$. This turns out to be extremely useful when the theory hasn't got a natural perturbative parameter, as in the example studied. At the same time, the Large–N_c limit is able to constrain the values of the free parameters of the theory, these constraints hinting at an underlying symmetry principle of Large–N_c.

Acknowledgements

I would like to thank the organisers of the International School of Subnuclear Physics (Erice '02), Professors A. Zichichi and G.'t Hooft, for giving me the opportunity to deliver a talk.

References

[1] J. Gasser and H. Leutwyler, Ann. of Phys. **158** (1984) 142. J. Gasser and H. Leutwyler, Nucl. Phys. **B250** (1985) 465.

[2] A. Pich, [hep-ph/0205030]. To appear in the proceedings of The Phenomenology of Large $N(c)$ QCD, Tempe, Arizona, 9-11 Jan 2002, Ed. R. Lebed, World Scientific, Singapore, 2002.

[3] G. Ecker, J. Gasser, A. Pich and E. de Rafael, Nucl. Phys. **B321** (1989) 311.

[4] G. 't Hooft, Nucl. Phys. **B72** (1974) 461. E. Witten, Nucl. Phys. **B160** (1979) 57.

[5] H. Leutwyler, Nucl. Phys. Proc. Suppl. **64** (1998) 223 [hep-ph/9709408]. R. Kaiser and H. Leutwyler, Eur. Phys. J. **C17** (2000) 623 [hep-ph/0007101].

[6] S. Peris, M. Perrottet and E. de Rafael, JHEP**9805** (1998) 011 [hep-ph/9805442].

[7] E. Witten, Nucl. Phys. **B156** (1979) 269.

[8] M. F. Golterman and S. Peris, Phys. Rev. D **61** (2000) 034018 [hep-ph/9908252].

[9] O. Catà and S. Peris, Phys. Rev. **D65** (2002) 056014 [hep-ph/0107062].

Towards the Finite Temperature Gluon Propagator in Landau Gauge Yang-Mills Theory

A. Maas

Institute for Nuclear Physics, Darmstadt University of Technology, Schloßgartenstraße

Abstract

Yang-Mills theories undergo a deconfining phase transition at a critical temperature. In lattice calculations the temporal Wilson loop and Z_3 order parameter show above this temperature a behavior typical of deconfinement. A quantity of interest in the study of this transition is the gluon propagator and its evolution with temperature. This contribution describes the current status of an investigation of the finite temperature gluon propagator in Landau gauge. It analyzes the high temperature case. The resulting equations are compared to the corresponding ones of three-dimensional Yang-Mills theory. Under certain assumptions it is found that a kind of spatial "confinement" is still present, even at very high temperatures.

1 Introduction

The gluon propagator at zero and finite temperature is of many possible uses, although it is a gauge-dependent quantity. In Landau gauge the ghost propagator is of equal importance. Knowledge of these propagators can be used as input to phenomenological calculations, e.g. in analysis of heavy ion collisions and the properties of the produced state of matter. On the other hand these propagators can be used to compare analytical calculations with lattice simulations to estimate the effect of approximations and to gain insight into the physical mechanisms underlying the lattice results and elucidate the dynamics of confinement.

This contribution describes the approach and the status of the calculation of the finite-temperature gluon propagator in Landau gauge Yang-Mills (YM) theory. The gluon propagator in the energy regime of interest for heavy ion collisions is non-perturbative, and so must be the calculation. As infrared singularities are anticipated, a continuum method is desirable. Therefore Dyson-Schwinger equations are employed as a non-perturbative approach. This will be described in section 2. The reason for choosing this method are the encouraging results at vanishing temperature which will be exemplified in section 3. Section 4 describes the necessary extensions to access the equilibrium properties of gluons at finite temperature. Section 5 will deal with the high-temperature limit and discuss the comparison with a dimensionally reduced theory. Finally this contribution ends with a summary and an outlook.

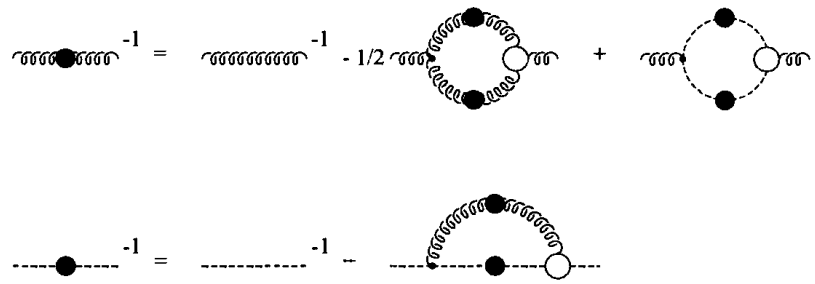


Figure 1: The truncated DSE used in the present work. The propagators with large full dots denote the fully dressed propagators while the ones without are the perturbative ones. A vertex denoted by a small full dot represents a perturbative vertex while the open circles indicate the fully dressed ones. In addition to the employed truncation all vertex functions are approximated by bare ones.

2 Gluon and ghost propagator equations

2.1 Dyson-Schwinger equations

Dyson-Schwinger equations [1] (DSE) are a well-known tool for non-perturbative calculations in several areas of physics. The DSE for YM theories can be obtained by using functional derivatives [2, 3]. However, this approach produces an infinite hierarchy of equations: each equation of a given n-point function depends on higher n-point functions. Since it is not possible to solve such an infinite coupled system, it is necessary to truncate it. There is no a-priori information about which truncation is sufficient to keep the relevant physics. The truncation scheme employed here is justified by comparison of the results at zero temperature with lattice results, which show that all qualitative features and many quantitative features survive. The truncation scheme is to keep only the equations for the 2-point functions and also only up to three-point vertex level as is depicted in figure 1.

In addition, it is assumed that multiplicative renormalization holds also in the non-perturbative regime and that therefore wavefunction renormalization and vertex renormalization can be cast into renormalization constants Z_i. Renormalization is performed using a momentum subtraction scheme. The resulting equations after inserting the color structure and removing an overall factor of δ_{ab} are[1]

$$D_G^{-1}(k) = -\widetilde{Z}_3 k^2 + \widetilde{Z}_1 N_c \int \frac{d^4q}{(2\pi)^4} i k_\mu D_{\mu\nu}(k-q) G_\nu(q,k) D_G(q) \,, \tag{1}$$

$$\begin{aligned} D_{\mu\nu}^{-1}(k) = {} & Z_3 D_{\mu\nu}^{tl\,-1}(k) - \widetilde{Z}_1 N_c \int \frac{d^4q}{(2\pi)^4} i q_\mu D_G(k+q) D_G(q) G_\nu(k+q,q) \\ & + \frac{1}{2} Z_1 N_c \int \frac{d^4q}{(2\pi)^4} \Gamma^{tl}_{\mu\rho\alpha}(k,-(k+q),q) \\ & D_{\alpha\beta}(q) D_{\rho\sigma}(k+q) \Gamma_{\beta\sigma\mu}(-q,q+k,-k) \end{aligned} \tag{2}$$

To render the equations dimensionless, the dressing functions G and Z are

[1] All quantities are defined using the conventions of ref. [2].

defined via their relation to the propagators

$$D_G(k) = -\frac{G(k)}{k^2}\,, \quad D_{\mu\nu}(k) = P_{\mu\nu}(k)\frac{Z(k)}{k^2} \tag{3}$$

for the ghost and the gluon, respectively, and the transverse projector is given by $P_{\mu\nu} = \delta_{\mu\nu} - k_\mu k_\nu / k^2$.

It remains to fix the fully dressed vertices, since they are unknown as long as the equations for the three point functions are discarded. It is possible to construct these vertices using the Slavnov-Taylor identities to constrain them as much as possible and this has been done [4]. At zero temperature, explicit calculations show that vertices constructed in this way only slightly improve the result compared to perturbative vertices [2]. The effect of other non-perturbatively dressed vertices has been studied in ref. [5]. Since constructed vertices induce a significant complexity to the problem it is therefore more reasonable in this first approach to finite temperature to keep only the perturbative vertices.

2.2 Gauge symmetry

To obtain only scalar equations it is useful to contract the gluon equation with the projector $P_{\mu\nu}$. However, it turns out that such a contraction produces spurious quadratic divergencies. They stem from the fact that the truncation of the DSE violates gauge symmetry. To remove these divergencies and therefore the effects of gauge symmetry violation it is possible to use instead of the transverse projector $P_{\mu\nu}$ other projectors which project onto states without gauge symmetry violations. This can be accomplished e.g. by using a generalized projector

$$P^{\zeta}_{\mu\nu} = \delta_{\mu\nu} - \zeta\frac{k_\mu k_\nu}{k^2} \tag{4}$$

where ζ is a real parameter. Choosing $\zeta = d = 4$ removes the spurious quadratic divergencies [6]. This procedure is not unique and introduces ambiguities in the value of numerical coefficients. This effect has been studied by varying the value of ζ while removing manually the spurious divergencies. The results at zero temperature demonstrate that these effects are small [7].

A second item with respect to gauge symmetry in the non-perturbative regime are Gribov copies. It has been shown that for the solutions of DSE in Landau gauge it is sufficient to require positive semi-definiteness of the dressing functions G and Z for all momenta to stay within the first Gribov horizon [8]. This is however not yet a full solution to the problem, since there are also Gribov copies within the first Gribov horizon. It is possible, in principle, to solve the problem by either introducing a second set of ghost fields to fully fix the gauge or by adding a new term to the DSE [8, 9]. But, since lattice calculations indicate that the influence of Gribov copies inside the first Gribov horizon is small [10], these considerable complifications are neglected for now.

2.3 Kugo-Ojima confinement criterion

When studying the finite-temperature gluon and ghost propagator one of the main goals to be achieved is the question of the fate of confinement. It is therefore necessary to find a criterion to test for the presence of confinement. Confinement means the absence of gluons from the physical spectrum. This is

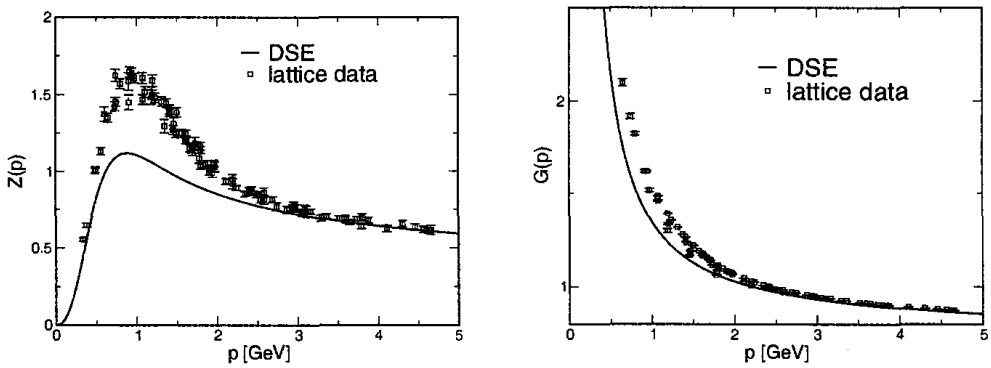


Figure 2: Comparison of lattice calculations [14] with results from DSE for $N_c = 2$. The left panel shows the dressing function of the gluon and the right panel the one for the ghost

for example possible if they do not have a Källen-Lehmann representation. Kugo and Ojima were able to construct a criterion to test for such a realization of confinement by assuming that the gluons carry BRST charge [11]. The criterion can be cast in a simple form in Landau gauge [12]: Does the euclidean ghost propagator diverge stronger than a particle pole as the euclidean four momentum approaches zero, i.e. does G diverge for $k \to 0$? To utilize this criterion and to simplify the calculations, the analysis will be performed in euclidean space.

3 Results at zero temperature

The calculations at zero temperature have been carried out in [2, 4, 7, 13]. Figure 2 shows a comparison of lattice calculations with results of the DSE. These are calculations for $N_c = 2$, but the dressing functions have the same form for $N_c = 3$, if 't Hooft scaling is employed, since the DSE in this truncation order depend only on the combination $g^2 N_c$.

Both propagators compare well with lattice results [14], although the gluon propagator misses some strength around 1 GeV. It is likely that this is due to the neglect of the two-loop diagram in the equations for the two-point functions [2]. The results show confinement according to the Kugo-Ojima criterion with a divergence of G with an exponent of $-\kappa$ in k^2 with $\kappa = 0.5953$. In addition, the gluon dressing function vanishes at zero momentum with an exponent of 2κ [5, 7, 8]. It is also possible to define a running coupling constant and it turns out that according to these calculations there is an infrared stable fixed point in YM theory. The good agreement with lattice calculations and the fact that the Kugo-Ojima criterion is satisfied motivates to use this method also at finite temperature.

4 Extension to finite temperature

4.1 Formulation

Since the primary interest are the equilibrium properties of gluons at finite temperature, it is convenient to use the Matsubara formalism. This induces that there are now two independent variables, the discrete zero component of the momentum and the absolute value of the three momentum. Additionally there are now two possible tensor structures for the gluon propagator [15] instead of only $P_{\mu\nu}$. Both have to be four dimensional-transverse due to gauge symmetry, but one is three-dimensional longitudinal and the other three-dimensional transverse. They can be expressed as

$$\begin{aligned} P_{T\mu\nu} &= \delta_{\mu\nu} - \frac{k_\mu k_\nu}{k_3^2} + \delta_{\mu 0}\frac{k_0 k_\nu}{k_3^2} + \delta_{0\nu}\frac{k_\mu k_0}{k_3^2} - \delta_{\mu 0}\delta_{0\nu}(1+\frac{k_0^2}{k_3^2}) \\ P_{L\mu\nu} &= P_{\mu\nu} - P_{T\mu\nu} \end{aligned} \tag{5}$$

where k_3 denotes the magnitude of the three momentum and the zero component of the momentum $k_0 = 2\pi n T$ is the bosonic Matsubara frequency. These projectors are orthogonal to each other and satisfy $P_{T\mu\mu} = 2$ and $P_{L\mu\mu} = 1$. There are therefore two independent dressing functions, Z_L and Z_T, and the gluon propagator is defined as

$$D_{\mu\nu}(k) = P_{T\mu\nu}\frac{Z_T(k)}{k^2} + P_{L\mu\nu}\frac{Z_L(k)}{k^2} \ . \tag{6}$$

By this definition Z_L and Z_T have to become equal at zero temperature and equal to Z.

It is possible to obtain two equations for Z_L and Z_T by contracting the gluon DSE with $P_{T\mu\nu}$ and with $P_{L\mu\nu}$ respectively. After inserting the perturbative vertices, this results in the following set of equations for the three dressing functions:

$$\begin{aligned} \frac{1}{G(k)} &= \widetilde{Z}_3 + \widetilde{Z}_1 \frac{g^2 T N_c}{(2\pi)^2}\sum_n \int d\theta dq_3 \\ &\qquad (A_L(k,q)G(q)Z_L(k-q) + A_T(k,q)Z_T(k-q)) \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{1}{Z_L(k)} &= Z_3 - \widetilde{Z}_1 \frac{g^2 T N_c}{(2\pi)^2}\sum_n \int d\theta dq_3 P(k,q)G(q)G(k+q) \\ &+ Z_1 \frac{g^2 T N_c}{(2\pi)^2}\sum_n \int d\theta dq_3 \\ &\qquad (N_L(k,q)Z_L(q)Z_L(q+k) + N_1(k,q)Z_L(q)Z_T(q+k) \\ &\qquad + N_2(k,q)Z_L(k+q)Z_T(q) + N_T(k,q)Z_T(q)Z_T(k+q)) \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{2}{Z_T(k)} &= 2Z_3 - \widetilde{Z}_1 \frac{g^2 T N_c}{(2\pi)^2}\sum_n \int d\theta dq_3 R(k,q)G(q)G(k+q) \\ &+ Z_1 \frac{g^2 T N_c}{(2\pi)^2}\sum_n \int d\theta dq_3 \\ &\qquad (M_L(k,q)Z_L(q)Z_L(q+k) + M_1(k,q)Z_L(q)Z_T(q+k) \\ &\qquad + M_2(k,q)Z_L(k+q)Z_T(q) + M_T(k,q)Z_T(q)Z_T(k+q)) \end{aligned} \tag{9}$$

where T is the temperature and the sum runs over the bosonic Matsubara frequencies of the gluon and the ghost. The integral kernels A_i, P, R, N_i and M_i are quite lengthy and will not be quoted here. The trivial angular integration has been performed.

Since no additional divergencies arise at finite temperature [16], the renormalization procedure can be kept by employing temperature-independent Z_i.

4.2 Gauge symmetry

In the same sense as at zero temperature, spurious quadratic divergencies arise due to the truncation and projection of the DSE. It is possible to remove such artifacts in the same, and therefore also ambiguous, way as at zero temperature. Since there are now two independent functions for the gluon dressing functions, both projectors have to be modified and it is necessary to introduce two real parameters. This can be accomplished using the following generalized projectors

$$\begin{aligned} P^{\xi}_{L\mu\nu} &= \xi P_{L\mu\nu} + (1-\xi)(1+\frac{k_0^2}{k_3^2})\delta_{\mu 0}\delta_{0\nu} \\ P^{\zeta}_{T\mu\nu} &= \zeta P_{T\mu\nu} + (1-\zeta)(\delta_{\mu\nu} - (1+\frac{k_0^2}{k_3^2})\delta_{\mu 0}\delta_{0\nu})\,. \end{aligned} \tag{10}$$

It turns out that all spurious divergencies are removed for $\xi = 0$ and $\zeta = 3$. These values are interesting, since they force the longitudinal projector to live only in the compactified dimension, while the transverse part becomes a structure which is reminiscent of the three-dimensional transverse projector.

5 The high-temperature limit

The finite-temperature regime is quite complicated and it is therefore useful to start at some limiting cases and evolve their solutions to finite temperatures. One possibility is to start at zero temperature and to evolve the known solution to non-zero temperatures. This approach is currently under investigation. The other approach is to start at high temperatures. As this method provided already some insights it will be described in the following.

5.1 Definition of the high-temperature limit

For the high-temperature limit, it is assumed that only the lowest Matsubara frequency contributes. It is therefore described by first letting the loop Matsubara frequency go to zero and then the exterior frequency go to zero. The appearance of the explicit temperature dependence can be removed by either measuring the momentum in units of temperature or by defining a temperature dependent coupling constant $g^2T = g_3^2$, which is dimensionful and its usefulness will be discussed in the next subsection.

By performing this limit, several remarkable observations can be made. The first is, that many of the integral kernels vanish and that the complete system becomes independent of ξ. Their exact values are

$$A_L = 0,\ A_T = \frac{q_3^2 \sin^3\theta}{u_-^2},\ P = 0,$$

$$N_L = 0,\ N_1 = -\frac{2q_3^2 \sin^3\theta}{u_+^2},\ N_2 = -\frac{2\sin^3\theta}{u_+},\ N_T = 0$$

$$R = \frac{(q_3^2 - (\zeta-1)w_3 - \zeta q_3^2 \cos^2\theta)\sin\theta}{k^2 u_+} \tag{11}$$

$$M_L = \frac{((\zeta-1)k_3^2 - 4q_3^2 + 4(\zeta-1)w_3 + 4\zeta q_3^2\cos^2\theta)\sin\theta}{2k_3^2 u_+},\ M_1 = 0,\ M_2 = 0$$

$$M_T = \frac{\sin\theta}{2k_3^2 u_+^2}((k_3^2 + 2q_3^2)((\zeta-9)k_3^2 - 4q_3^2) + 8(\zeta-3)w_3(k_3^2+q_3^2)$$

$$+((8\zeta q_3^4 + (\zeta+7)k_3^4 + 4(5\zeta-1)k_3^2 q_3^2) + 4(4\zeta q_3^2 + k_3^2(\zeta+3))w_3 + 4\zeta w_3^2)\cos^2\theta)$$

where $u_\pm = k_3^2 + q_3^2 \pm 2k_3 q_3 \cos\theta$ and $w_3 = k_3 q_3 \cos\theta$. Secondly, the three-dimensional transverse projector now becomes the transverse projector of a theory with only three dimensions, while the longitudinal projector only acts in the compactified dimension. In addition, it is interesting to note that the value of ζ coincides with that of the three-dimensional Brown-Pennigton projector. But the most striking feature is, that if Z_L^2 can be neglected compared to Z_T^2, then the equation for Z_L decouples. This leads to an interesting comparison to the three-dimensional theory.

5.2 Comparison to three-dimensional Yang-Mills theory

Compactified four-dimensional YM theory is not the same as three-dimensional YM theory, since the zeroth component of the four-dimensional gauge field becomes an additional Higgs field [17]. However, lattice calculations indicate that this Higgs field is unimportant and produces only a small effect on the gluon propagator [18]. It is therefore interesting to compare to the three-dimensional YM theory only. The DSE can be calculated in the same way as in four dimensions and yield

$$\frac{1}{G_{3d}(k_3)} = \widetilde{Z}_3 + \widetilde{Z}_1 \frac{g_3^2 N_c}{(2\pi)^2}\int d\theta dq_3 A_3(k_3,q_3) G_{3d}(q_3) Z_{3d}(k_3 - q_3) \tag{12}$$

$$\begin{aligned}\frac{2}{Z_{3d}(k_3)} &= 2Z_3 - \widetilde{Z}_1 \frac{g_3^2 N_c}{(2\pi)^2}\int d\theta dq_3 P_3(k_3,q_3) G_{3d}(q_3) G_{3d}(k_3+q_3) \\ &\quad + Z_1 \frac{g_3^2 N_c}{(2\pi)^2}\int d\theta dq_3 (N_3(k_3,q_3) Z_{3d}(q_3) Z_{3d}(q_3+k_3))\end{aligned} \tag{13}$$

with the integral kernels

$$\begin{aligned}A_3 &= \frac{q_3^2\sin^3\theta}{u_-^2},\ P_3 = \frac{\sin\theta}{2k_3^2 u_+}(2(\zeta-1)w_3 + q_3^2(\zeta - 2 + \zeta\cos 2\theta)) \\ N_3 &= \frac{\sin\theta}{2k_3^2 u_+^2}((k_3^2+2q_3^2)(k_3^2(\zeta-9) - 4q_3^2 + 8w_3(\zeta-3)(k_3^2+q_3^2) \\ &\quad +((8\zeta q_3^4 + (7+\zeta)k_3^4 + 4k_3^2 q_3^2(5\zeta-1)) \\ &\quad\quad +4w_3(4\zeta q_3^2 + (\zeta+3)k_3^2) + 4\zeta w_3^2)\cos^2\theta)\end{aligned} \tag{14}$$

using the three-dimensional Brown-Pennigton projector. It turns out, that the resulting coupled equations for the three-dimensional ghost and gluon are very similar. In fact, if Z_L can be neglected, the three-dimensional coupling constant g_3^2 is identified with g^2T and the three-dimensional dressing function Z_{3d} with Z_T, then the above equations become identical to the ones for three-dimensional YM theory provided $\zeta = 3$ is chosen. It is therefore tempting to identify the longitudinal part of the four-dimensional gluon propagator with the contribution from the Higgs field. This is supported by the fact that the three-dimensional gluon propagator has also to be transverse due to gauge symmetry and the three-dimensional longitudinal part Z_L therefore cannot be part of it. This in turn would justify the assumption that Z_L is indeed small and can be neglected compared with Z_T. However, only a detailed calculation of these quantities can prove this assumption. Additional support is found by the fact that in another approximation scheme, the Mandelstam approximation, the same situation arises. Here is, however, no space to detail these other calculations which have been done by us. Moreover, recent lattice results support the finding of an infrared divergent ghost dressing function at finite temperature [14] which diverges weaker than at zero temperature as would be expected from the above, since in three dimensions κ is smaller [8].

If this turns out to be the case, it would have remarkable consequences, since the three-dimensional YM theory indeed does confine [8], and therefore a (spatial) kind of confinement would still be present even at very high temperatures in YM theory. This corresponds to lattice findings in that only the temporal Wilson loop shows deconfinement while the Wilson loops in the not compactified spatial dimensions do not [19]. This would underline the fact that the magnetic sector of YM theory is non-trivial even at high temperatures.

6 Summary and outlook

This contribution describes our recent progress in a calculation of the finite-temperature gluon propagator in Landau-gauge YM theory. It is a first step generalizing a self-consistent zero-temperature formulation within the Matsubara formalism. Possibilities to circumvent the problem of gauge symmetry violations due to the truncation of the DSE have been discussed. A comparison of the high-temperature equations of the compactified theory with the corresponding equations for a lower-dimensional theory reveals interesting properties. If the assumption of a negligible influence of the three-dimensional longitudinal part of the gluon propagator at high temperature turns out to be correct, then this proves the non-triviality of the magnetic sector of the YM theory even at high temperatures. Note that this would imply the presence of some confinement effects at very high temperatures. The physical realization of such a spatial "confinement" has still to be understood. This may be possible as soon as the solutions of the DSE are obtained. The corresponding calculations are under way.

Acknowledgments

This work has been done in collaboration with Burghard Grüter and Reinhard Alkofer from the University of Tübingen and with Jochen Wambach from the Darmstadt University of Technology.

A. M. thanks the organizers of the International School of Subnuclear Physics for the very interesting meeting and for the opportunity to present this talk. The authors are grateful to C. S. Fischer for helpful discussions.

This work is supported by the BMBF under grant number 06DA917 and by the European Graduate School Basel-Tübingen (DFG contract GRK683).

References

[1] F. J. Dyson, Phys. Rev. **75** (1949) 1736, J. S. Schwinger, Proc. Nat. Acad. Sci. **37** (1951) 452, Proc. Nat. Acad. Sci. **37** (1951) 455.

[2] R. Alkofer and L. von Smekal, Phys. Rept. **353** (2001) 281.

[3] C. D. Roberts and S. M. Schmidt, Prog. Part. Nucl. Phys. **45** (2000) S1.

[4] L. von Smekal, A. Hauck and R. Alkofer, Annals Phys. **267** (1998) 1; L. von Smekal, R. Alkofer and A. Hauck, Phys. Rev. Lett. **79** (1997) 3591.

[5] C. Lerche and L. von Smekal, Phys. Rev. D **65** (2002) 125006.

[6] N. Brown and M. R. Pennington, Phys. Rev. D **38** (1988) 2266.

[7] C. S. Fischer and R. Alkofer, Phys. Lett. B **536** (2002) 177.

[8] D. Zwanziger, Phys. Rev. D **65** (2002) 094039.

[9] D. Zwanziger, Nucl. Phys. B **412** (1994) 657.

[10] A. Cucchieri, Nucl. Phys. B **521** (1998) 365.

[11] T. Kugo and I. Ojima, Prog. Theor. Phys. Suppl. **66** (1979) 1.

[12] T. Kugo, arXiv:hep-th/9511033.

[13] R. Alkofer, C. S. Fischer and L. von Smekal, Acta Phys. Slov. **52** (2002) 191.

[14] K. Langfeld, J. C. Bloch, J. Gattnar, H. Reinhardt, A. Cucchieri and T. Mendes, arXiv:hep-th/0209173.

[15] J. I. Kapusta, "Finite Temperature Field Theory", *Cambridge University Press (1989)* .

[16] A. K. Das, "Finite Temperature Field Theory", *Singapore: World Scientific (1997)* .

[17] K. Kajantie, M. Laine, K. Rummukainen and M. E. Shaposhnikov, Nucl. Phys. B **458** (1996) 90.

[18] A. Cucchieri, F. Karsch and P. Petreczky, Phys. Rev. D **64** (2001) 036001.

[19] G. S. Bali, J. Fingberg, U. M. Heller, F. Karsch and K. Schilling, Phys. Rev. Lett. **71** (1993) 3059.

HERMES MEASUREMENTS OF THE NUCLEON SPIN STRUCTURE

J. WENDLAND

Department of Physics, Simon Fraser University
Burnaby, British Columbia, V5A 1S6 Canada
E-mail: juergen.wendland@triumf.ca

FOR THE HERMES COLLABORATION

The HERMES collaboration has measured double spin asymmetries using polarized deep inelastic scattering. The spin structure functions $\mathbf{g_1}$ of the proton, the neutron, and the deuteron calculated from inclusive asymmetries are in agreement with world data. Semi-inclusive asymmetries of hadrons were measured from proton and deuteron targets with high precision. Asymmetries of pions from the proton and of pions and kaons from the deuteron were measured for the first time. Quark polarizations of the u, $\bar{u}$, d, $\bar{d}$, $(s + \bar{s})$ flavours were extracted from the inclusive and semi-inclusive spin asymmetries in a LO analysis. The polarizations of the up and the down quarks were determined with good precision to be positive and negative, respectively. The polarization of the sea quarks is consistent with zero.

1. Introduction

Polarized deep inelastic scattering (DIS) is a powerful tool for the investigation of the nucleon spin structure. Measurements by the European Muon Collaboration (EMC) first indicated that only a small fraction of the nucleon spin is due to the spin of the quarks[1]. Several experiments have been carried out since[2,3,4,5,6]. Their results also indicate a significantly smaller value for the quark polarization than naively expected[7].

The HERMES experiment[8] was designed to perform precise measurements of the quark polarizations. The experiment is carried out at DESY in Germany, utilizing the polarized 27.5 GeV electron or positron beam of the HERA accelerator in combination with an internal polarized gaseous target of 3-helium, hydrogen, or deuterium. With its large forward acceptance and good particle identification, the spectrometer is well suited to measure inclusive reactions where only the scattered positron is detected, and semi-inclusive processes where a hadron is detected in coincidence. A threshold Čerenkov detector provided identification of pions in the hydrogen data. A ring imaging Čerenkov (RICH) detector installed in 1998 allows the iden-

tification of pions, kaons, and protons over a wide momentum range in the data taken with the deuterium target.

2. Inclusive Spin Asymmetries

In inclusive DIS the measurable kinematic variables are the initial and final momenta $k = (E, \vec{k})$ and $k' = (E', \vec{k'})$ of the incident and scattered lepton respectively and the initial momentum of the target nucleon, $P = (M, \vec{0})$ in the case of a fixed target. Common DIS variables are the negative squared invariant mass of the virtual photon $Q^2 = -(k - k')^2$, and the Bjørken scaling variable $x = Q^2/(2\,M\nu)$, where $\nu = E - E'$.

The inclusive spin asymmetry of cross sections with aligned and anti-aligned target and beam spins gives access to the spin structure function $\mathbf{g_1}$. The measured lepton-nucleon asymmetry $A_{\parallel}$ is related to the photon-nucleon asymmetry A_1 and the nucleon structure functions $\mathbf{g_1}$, $\mathbf{g_2}$ and $\mathbf{F_1}$ in the following way:

$$\frac{A_{\parallel}(x, Q^2)}{D\,(1 - \eta\gamma)} \simeq A_1(x, Q^2) = \frac{\mathbf{g_1}(x, Q^2) - \gamma^2\,\mathbf{g_2}(x, Q^2)}{\mathbf{F_1}(x, Q^2)}. \tag{1}$$

Here D is the depolarization factor, and η and $\gamma = \sqrt{Q^2/\nu^2}$ can be calculated from the scattering kinematics. The function $\mathbf{g_2}$ was measured to be very small at SLAC[9]. Contributions from $\mathbf{g_2}$, which are further suppressed by the small kinematic factor γ^2, are thus neglected. The spin structure function $\mathbf{g_1}$ can then be computed using equation (1) with a parameterization of world data for $\mathbf{F_1}$.

The preliminary data on the inclusive deuteron asymmetry $A_{1,d}$, shown in the left hand panel in Fig. 1 are based on 10 million DIS events taken from a polarized deuterium target and in the kinematic range $Q^2 > 0.1\text{GeV}^2$ and $W^2 > 3.24\text{GeV}^2$. The HERMES data agree with earlier measurements carried out by SMC[10] and at SLAC[5,6], even though the SMC data were collected at a significantly higher average Q^2. This is a clear indication that both the polarized and unpolarized nucleon structure functions evolve very similarly with Q^2. The recent world data[2,4,5,6,10,11,12] of the spin structure functions of the proton, the deuteron, and the neutron (extracted from 3-helium) are shown on the right hand side of Fig. 1. The data from the various experiments are presented at their measured $\langle Q^2 \rangle$. The proton spin structure function is positive in the measured range of x, because it is dominated by the positively polarized up quark (see below). The polarized neutron structure function is found to be negative in the measured range of x, as it largely probes the polarization of the isospin-rotated down quark

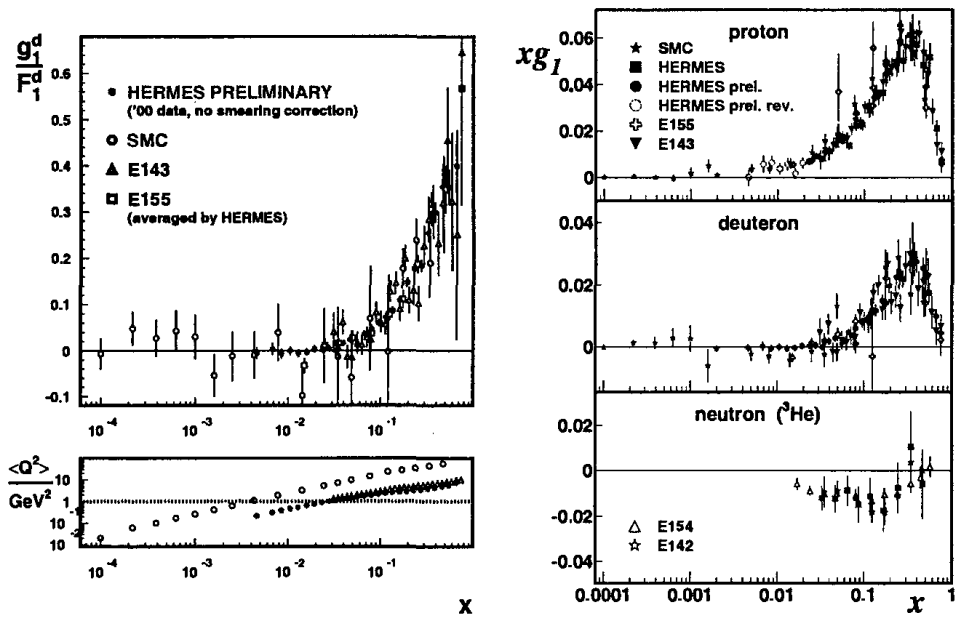


Figure 1. Left: Preliminary HERMES results on $\mathbf{g_1^d/F_1^d}$ as a function of x. Results by the SMC, E143, and E155 collaborations are also shown. The lower panel compares the average values of Q^2 in each bin of x. Right: World data on the spin structure function $\mathbf{g_1}$. The top panel shows $x\,\mathbf{g_1}$ of the proton, the middle panel $x\,\mathbf{g_1}$ of the deuteron, and the bottom panel presents $x\,\mathbf{g_1}$ of the neutron, all at their respective measured Q^2.

of the proton. The polarized structure function $\mathbf{g_{1d}}$, which is essentially the average of the proton and the neutron functions, is positive but smaller than $\mathbf{g_{1p}}$.

3. Semi-Inclusive Spin Asymmetries

In the quark parton model the structure functions $\mathbf{F_1}$ and $\mathbf{g_1}$ can be written in terms of the unpolarized and polarized quark densities q_f and Δq_f respectively:

$$\mathbf{F_1}(x,Q^2) = \sum_f e_f^2\, q_f(x,Q^2), \qquad \mathbf{g_1}(x,Q^2) = \sum_f e_f^2\, \Delta q_f(x,Q^2), \qquad (2)$$

where e_f^2 is the charge of the quark f, and the sum runs over the quark flavours u, d, s, $\bar{u}$, $\bar{d}$, and $\bar{s}$. The densities q_f and Δq_f are defined in terms of the quark densities q_f^+ and q_f^- for quarks with spins aligned and anti-aligned with the nucleon spin respectively: $q_f = (q_f^+ + q_f^-)$ and $\Delta q_f = (q_f^+ - q_f^-)$.

In semi-inclusive deep inelastic scattering (SIDIS), additional information from the final state hadron with fractional energy $z = (E_h/\nu)$ allows

a determination of the polarized quark densities[13,14]. The semi-inclusive double spin asymmetry A_1^h of hadron type h reads in terms of the quark densities q_f and Δq_f

$$A_1^h(x,Q^2) \overset{g_2=0}{\simeq} \frac{1+R(x,Q^2)}{1+\gamma^2} \frac{\sum_f e_f^2\, \Delta q_f(x,Q^2) \int_{z_{\min}}^{z_{\max}} \mathrm{d}z\, D_f^h(z,Q^2)}{\sum_{f'} e_{f'}^2\, q_{f'}(x,Q^2) \int_{z_{\min}}^{z_{\max}} \mathrm{d}z\, D_{f'}^h(z,Q^2)}, \quad (3)$$

where $R = \sigma_L/\sigma_T$ is the ratio of the longitudinal to transverse photo-absorption cross sections. The fragmentation functions D_f^h which give the probability that a quark of flavour f fragments into a hadron h in the final state, are integrated over the range $z_{\min} = 0.2$ to $z_{\max} = 0.8$. The upper limit rejects exclusive events from the SIDIS sample.

The quark polarizations are isolated by rewriting equation (3) and introducing purities P_f^h,

$$\begin{aligned} A_1^h(x) &= \sum_f P_f^h(x)\, \frac{\Delta q_f}{q_f}(x), \\ P_f^h(x) &\equiv \frac{1+R(x)}{1+\gamma^2} \frac{e_f^2\, q_f(x) \int_{z_{\min}}^{z_{\max}} \mathrm{d}z\, D_f^h(z)}{\sum_{f'} e_{f'}^2\, q_{f'}(x) \int_{z_{\min}}^{z_{\max}} \mathrm{d}z\, D_{f'}^h(z)}, \end{aligned} \quad (4)$$

where all quantities were integrated at each x over the corresponding range of Q^2 covered by the experiment. The purities P_f^h describe the probability that a hadron h originates from an event where a quark of flavour f was struck. They depend on the fragmentation functions, the unpolarized quark densities and in case of the deuterium target on the relative fluxes of hadrons originating from the various nucleons. The purities also include effects from the limited acceptance of the spectrometer. The fragmentation was modelled in the LUND string model implemented in the JETSET 7.4 package[15]. The string breaking parameters of the LUND model were tuned to fit the hadron multiplicities measured at HERMES, in order to achieve an accurate description of the fragmentation at HERMES energies. The CTEQ5LO parton distributions[16] were used to model the unpolarized quark densities.

The quark polarizations are obtained by combining the hydrogen and deuterium inclusive and semi-inclusive asymmetries of positive and negative hadrons, pions, and kaons into an over-constrained system of linear equations and solving for $\Delta q_f/q_f$. The flavour decomposition was carried out with the parameters $\frac{\Delta u}{u}$, $\frac{\Delta \bar{u}}{\bar{u}}$, $\frac{\Delta d}{d}$, $\frac{\Delta \bar{d}}{\bar{d}}$, $\frac{\Delta s}{s} \equiv \frac{\Delta \bar{s}}{\bar{s}}$. In contrast to earlier analyses[13,17], the only remaining symmetry assumption is that the strange sea be symmetric $\frac{\Delta s}{s} \equiv \frac{\Delta \bar{s}}{\bar{s}}$.

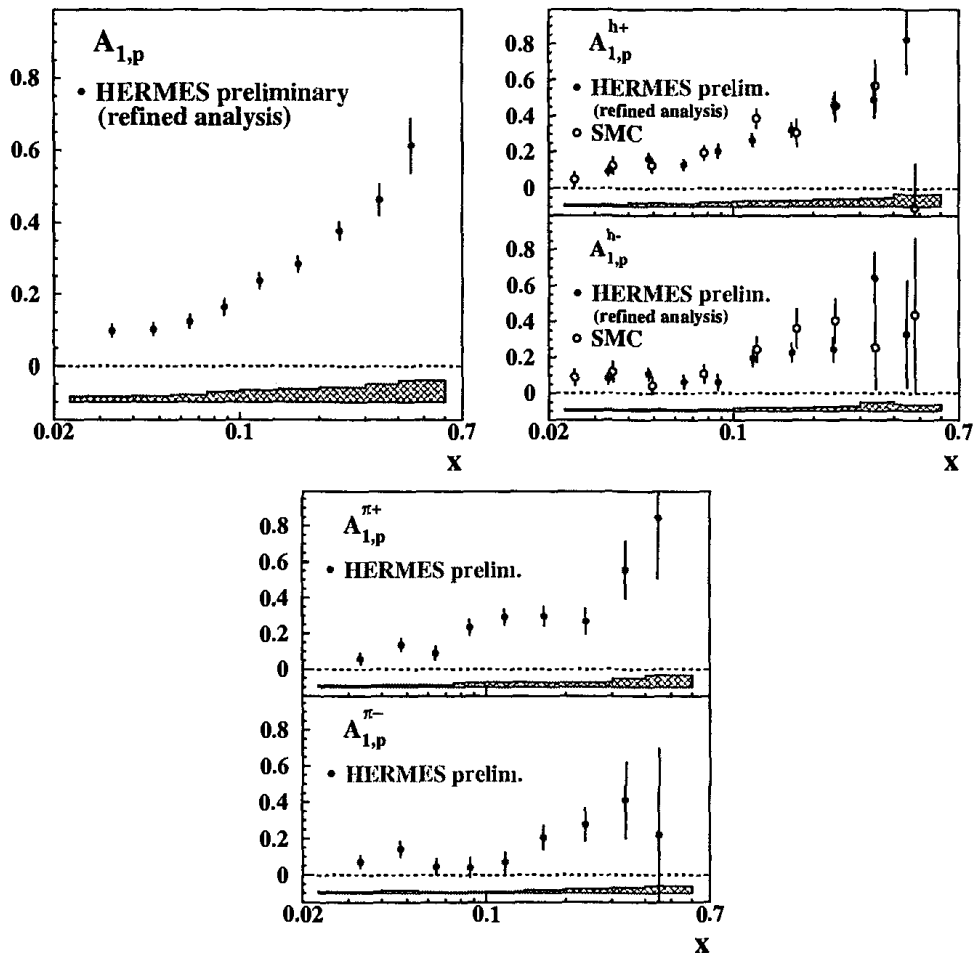


Figure 2. Preliminary HERMES results on the inclusive and semi-inclusive hadron and pion asymmetries from the proton. The panels of the positive and negative hadron asymmetries also show asymmetries measured by SMC[17] for comparison. The error bars represent the statistical uncertainties and the error band the systematic uncertainty of the asymmetries.

The preliminary inclusive and semi-inclusive asymmetries presented in Figs. 2 and 3 are based on 1.8 and 6.5 million DIS events respectively. The data were collected in the kinematic range $Q^2 > 1\,\mathrm{GeV}^2$ and $W^2 > 10\,\mathrm{GeV}^2$. Semi-inclusive hadrons were selected by requiring $0.2 < z < 0.8$ and $x_F \simeq 2p_L/W > 0.1$ where p_L is the longitudinal momentum of the hadron with respect to the virtual photon direction in the photon-nucleon center-of-mass frame. The lower limits suppress hadrons from the target fragmentation region. The upper cut on z rejects hadrons from exclusive events. The hydrogen data were collected with a threshold

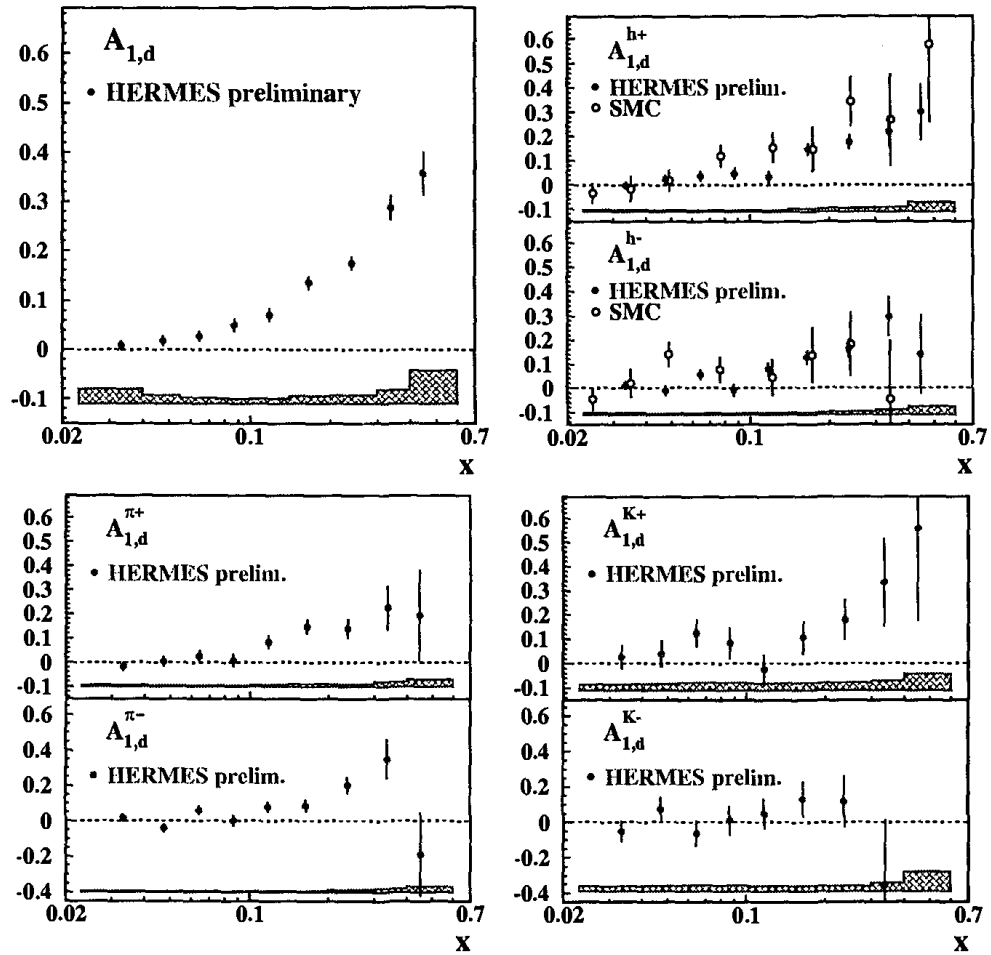


Figure 3. Preliminary HERMES results on the inclusive and semi-inclusive hadron and pion asymmetries on the deuteron. The panels of the positive and negative hadron asymmetries also show asymmetries measured by SMC[17] for comparison. The pion and kaon asymmetries shown in the lower panels were measured for the first time using a ring imaging Čerenkov detector. The error bars represent the statistical uncertainties and the error band the systematic uncertainty of the asymmetries.

Čerenkov detector installed which allowed the identification of pions with momenta above approximately 4 GeV from the hadron sample. In the case of the deuterium data, the RICH detector provided kaon identification in addition. The asymmetries on the proton are positive over the entire range of x reaching the largest values at high x. The deuteron asymmetries are generally smaller. For comparison the positive and negative hadron asymmetries measured by SMC are also shown. The pion and kaon asymmetries were measured for the first time by the HERMES collaboration.

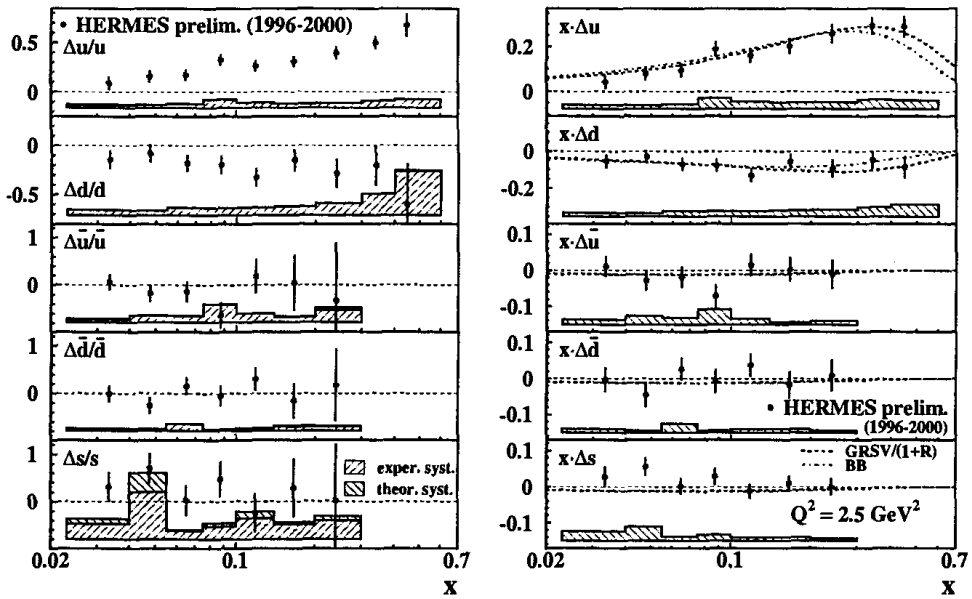


Figure 4. The left hand panel shows the quark polarizations in the flavour decomposition as a function of x extracted from HERMES inclusive and semi-inclusive asymmetries on hydrogen, and deuterium targets. The panel on the right shows the polarized quark densities for the up and down quarks, $x\,\Delta u(x)$ and $x\,\Delta d(x)$, and for the sea quarks, $x\,\Delta\bar{u}(x)$, $x\,\Delta\bar{d}(x)$, and $x\,\Delta\bar{s}(x)$. The data are presented at a common Q^2 of $2.5\,\mathrm{GeV}^2$. Data from SMC also evolved to $Q^2 = 2.5\,\mathrm{GeV}^2$ are shown for comparison. The dashed lines and the dashed-dotted lines show parameterizations of references [18,19]. In both panels the error bars represent the statistical and the bands the systematic uncertainties. The contribution labelled "theor. syst." in the left hand panel is the uncertainty due to differences when using different sets of unpolarized parton densities[16,20].

The polarizations of the five quark flavours extracted using the presented inclusive and semi-inclusive asymmetries are shown in Fig. 4. The systematic uncertainties include those due to the asymmetry measurements, the unpolarized parton densities, and an estimate of the uncertainties due to the fragmentation functions. At values larger than $x = 0.3$ the polarization of the sea flavours was set to zero. The resulting small uncertainties in the up and down quark polarizations were also included in the systematic uncertainty. The polarization of the up quark is positive in the measured range of x and increases with x up to 0.7 at $x = 0.47$. The down quark polarization is negative and shows no statistically significant dependence on x. The polarizations of the light sea quarks are compatible with zero, whereas a slightly positive polarization of the strange quark is favoured in contrast to results based on inclusive data alone (cf. e.g. [18,19]). However, within the total uncertainty the strange quark polarization is also zero.

4. Conclusion

The HERMES collaboration has collected a wealth of inclusive and semi-inclusive deep inelastic scattering data on polarized hydrogen and deuterium targets. The inclusive data on the deuteron allowed for a precise determination of the asymmetry $A_{1,d}$. The data are in excellent agreement with results by E155, E143, and SMC, indicating that the asymmetry is Q^2 independent within the experimental uncertainty. The spin structure functions $\mathbf{g_1}$ on the proton, the deuteron, and the neutron were measured over a large range of x. The HERMES data, which are in good agreement with results from previous experiments, significantly improve the world data set.

Semi-inclusive hadron asymmetries on the proton and the deuteron were measured with good precision. Pion asymmetries on the proton and pion and kaon asymmetries on the deuteron were measured for the first time. In a leading order QCD analysis quark polarizations were extracted from these semi-inclusive and inclusive asymmetries. The polarizations of the up and down quark were determined with good precision to be positive and negative respectively. The polarized quark sea was decomposed for the first time. Within the experimental uncertainty the polarizations of the sea quarks, $\Delta\bar{u}$, $\Delta\bar{d}$, and $\Delta s \equiv \Delta\bar{s}$, are compatible with zero.

References

1. J. Ashman *et al.* [EMC], Phys. Lett. B **206**, 364 (1988).
2. P. L. Anthony *et al.* [E142], Phys. Rev. D **54**, 6620 (1996) [arXiv:hep-ex/9610007].
3. B. Adeva *et al.* [SMC], Phys. Lett. B **412**, 414 (1997).
4. K. Abe *et al.* [E154], Phys. Rev. Lett. **79**, 26 (1997) [arXiv:hep-ex/9705012].
5. K. Abe *et al.* [E143], Phys. Rev. D **58**, 112003 (1998) [arXiv:hep-ph/9802357].
6. P. L. Anthony *et al.* [E155], Phys. Lett. B **463**, 339 (1999) [arXiv:hep-ex/9904002].
7. B. W. Filippone *et al.*, arXiv:hep-ph/0101224.
8. K. Ackerstaff *et al.* [HERMES], Nucl. Instrum. Meth. A **417**, 230 (1998) [arXiv:hep-ex/9806008].
9. P. L. Anthony *et al.* [E155], arXiv:hep-ex/0204028.
10. B. Adeva *et al.* [SMC], Phys. Rev. D **58** (1998) 112001.
11. B. Adeva *et al.* [SMC], Phys. Rev. D **60**, 072004 (1999) [Erratum-ibid. D **62**, 079902 (2000)].
12. A. Airapetian *et al.* [HERMES], Phys. Lett. B **442**, 484 (1998) [arXiv:hep-ex/9807015].
13. K. Ackerstaff *et al.* [HERMES], Phys. Lett. B **464**, 123 (1999) [arXiv:hep-ex/9906035].
14. M. Beckmann, DESY-THESIS-2000-029.

15. T. Sjostrand, Comput. Phys. Commun. **82**, 74 (1994).
16. H. L. Lai *et al.* [CTEQ], Eur. Phys. J. C **12**, 375 (2000) [arXiv:hep-ph/9903282].
17. B. Adeva *et al.* [SMC], Phys. Lett. B **420**, 180 (1998) [arXiv:hep-ex/9711008].
18. M. Glück, E. Reya, M. Stratmann and W. Vogelsang, Phys. Rev. D **63**, 094005 (2001) [arXiv:hep-ph/0011215].
19. J. Blümlein and H. Böttcher, Nucl. Phys. B **636**, 225 (2002) [arXiv:hep-ph/0203155].
20. M. Glück, E. Reya and A. Vogt, Eur. Phys. J. C **5**, 461 (1998) [arXiv:hep-ph/9806404].

DIPLOMAS

Eighteen Diplomas were open for competition among the participants. They have been awarded as follows:

- John S. BELL Diploma to:

  Oscar CATÀ
  Universitat Autònoma de Barcelona, Bellaterra, Spain

- Patrick M.S. BLACKETT Diploma to:

  Mark RAMTOHUL
  University of Edinburgh, UK

- James CHADWICK Diploma to:

  Juergen WENDLAND
  Simon Fraser University, Burnaby, Canada

- Paul A.M. DIRAC Diploma to:

  Alexander NESTERENKO
  Joint Institute for Nuclear Research, Dubna, Russia

- Vladimir N. GRIBOV Diploma to:

  Vladimir MARKOV
  Politechnical University, PNPI, St. Petersburg, Russia

- Robert HOFSTADTER Diploma to:

  Alessandro TORRIELLI
  Università degli Studi di Padova, Italy

- Gunnar KÄLLEN Diploma to:

  Axel MAAS
  Technische Universität Darmstadt, Germany

- Giuseppe P.S. OCCHIALINI Diploma to:

  Philip BECHTLE
  DESY, Hamburg, Germany

- Bruno PONTECORVO Diploma to:

  Mauro PAPINUTTO
  DESY, Hamburg, Germany

- Oreste PICCIONI Diploma to:

  Tristan MAILLARD
  ETH, Institut for Theoretical Physics, Zurich, Switzerland

- Isidor I. RABI Diploma to:

  Anton POLOUEKTOV
  Budker Institute for Nuclear Physics, NOVOSIBIRSK, Russia

- Giulio RACAH Diploma to:

  Dipankar CHAKRABARTI
  Saha Institute of Nuclear Physics, Kolkata, India

- Bruno ROSSI Diploma to:

  Hiroaki OHNISHI
  CERN, Geneva, Switzerland

- Andreij D. SAKHAROV Diploma to:

  Maxim ZAVATSKY
  Politechnical University, PNPI, St. Petersburg, Russia

- Victor F. WEISSKOPF Diploma to:

  Emmanuel KOHLPRATH
  CERN, Geneva, Switzerland

- Eugene P. WIGNER Diploma to:

  Michele PAPUCCI
  Scuola Normale Superiore, Pisa, Italy

- Bjorn H. WIIK Diploma to:

  Rainer STAMEN
  DESY, Hamburg, Germany

- Chien Shiung WU Diploma to:

  Cristiana SEBU
  Université Montpellier II, Montpellier, France

AWARDS

The Awards for the best "New Talents"-"Speaker" is divided into three parts:

i) for the best theoretical presentation, to:

Mauro PAPINUTTO
DESY, Hamburg, Germany

ii) for the best experimental presentation, to:

Juergen WENDLAND
Simon Fraser University, Burnaby, Canada

iii) for the original work, to:

Rainer STAMEN
DESY, Hamburg, Germany

- Award for BEST STUDENT given to:

Philip BECHTLE
DESY, Hamburg, Germany

- Award for BEST SCIENTIFIC SECRETARY given to:

Axel MAAS
Technische Universität Darmstadt, Germany

PARTICIPANTS

Halina ABRAMOWICZ Israel	Tel Aviv University Ramat Aviv 69978 TEL AVIV, Israel
Serguei AFONINE Russia	V.A. Fock Institute of Physics University of St. Petersburg RU-198504 ST. PETERSBURG, Russia
Alexander AKINDINOV Russia	Institute for Theoretical and Experimental Physics B. Cheremushkinskaya ulitsa 25 RU-117 259 MOSCOW, Russia
Andrea ALICI Italy	Università degli Studi di Bologna Via Irnerio 46 I-40126 BOLOGNA, Italy
Giovanni AMELINO CAMELIA Italy	Università degli Studi di Roma La Sapienza Piazzale Aldo Moro 2 I-00135 ROMA, Italy
Philip BECHTLE Germany	DESY FLC Notkestrasse 85 D-22607 HAMBURG, Germany
Wolf BEIGLBÖCK Austria	Springer-Verlag Heidelberg Tiergartenstrasse 17 D-69121 HEIDELBERG, Germany
Alessandro BETTINI Italy	Laboratori Nazionali del Gran Sasso - LNGS I- 67010 ASSERGI, L'Aquila, Italy
Francesco BIGAZZI Italy	International Centre for Theoretical Physics Strada Costiera 11 I-34014 TRIESTE
Gustaaf BROOIJMANS The Netherlands	Fermi National Accelerator Laboratory P.O. Box 500 BATAVIA, IL 60510-0500, USA
Giuseppe Eugenio BRUNO Italy	Università degli Studi di Bari Via Amendola 173 I-70126 BARI, Italy
Concetta CARTARO Italy	Università di Napoli Federico II Complesso Universitario Monte Sant'Angelo Via Cinthia I-80126 NAPOLI, Italy
Oscar CATÀ Spain	Universitat Autònoma de Barcelona E-08193 BELLATERRA (Barcelona), Spain
Dipankar CHAKRABARTI India	Saha Institute of Nuclear Physics Bidhannagar - KOLKATA 700 064, India

Francesco CONVENTI
Italy

Università di Napoli Federico II
Complesso Universitario Monte Sant'Angelo
Via Cinthia
I-80126 NAPOLI, Italy

Christopher DAWSON
UK

RIKEN/BNL Research Center
Brookhaven National Laboratory
UPTON, NY 11973, USA

Antonino DI PIAZZA
Italy

Università degli Studi di Trieste
Strada Costiera 11
I-34014 MIRAMARE-GRIGNANO, Italy

Erik DVERGSNES
Norway

University of Bergen
Allegaten 55
N-5007 BERGEN, Norway

François ENGLERT
Belgium

Université Libre de Bruxelles
Campus Plaine, CP 225 - Blvd du Triomphe
B-1050 BRUSSELS, Belgium

Fabrizio FERRO
Italy

University of Genoa
Via Dodecaneso 33
I-16146 GENOVA, Italy

Alexey GOLOVNEV
Russia

University of St. Petersburg
RU-198504 ST. PETERSBURG, Russia

Michel GOURDIN
France

Laboratoire de Physique Théorique et Hautes Energies
Université Pierre et Marie Curie
4 Place Jussieu
F-75252 PARIS Cedex 05, France

Dan Radu GRIGORE
Romania

Institute for Theoretical Physics
Zurich University
Winterthurerstr. 190
CH-8057 ZURICH, Switzerland

Nikolay GROMOV
Russia

University of St. Petersburg
RU-198504 ST. PETERSBURG, Russia

Roberto GUIDA
Italy

INFN – Sezione di Pavia
Università di Pavia
Via A. Bassi 6
I-27100 PAVIA, Italy

Dieter HAIDT
Germany

EPJC - DESY
Notkestrasse 85
D-22603 HAMBURG, Germany

Timothy J. HALLMAN
USA

Brookhaven National Laboratory
20 Pensylvania Street
UPTON, NY 11973, USA

Sonja HILLERT
Germany

DESY
Notkestrasse 85
D-22607 HAMBURG, Germany

Fedor IGNATOV
Russia

The Budker Institute of Nuclear Physics
Academy of Sciences Lavrentiev pr.11
RU-630090 NOVOSIBIRSK 90, Russia

Masaya ISHINO
Japan

CERN
CH-1211 GENEVA 23, Switzerland

Richard D. KENWAY
UK

The University of Edinburgh
The King's Buildings
EDINBURGH, EH4 3JZ, Scotland, UK

Emmanuel KOHLPRATH
Austria + Switzerland

CERN
CH-1211 GENEVA 23, Switzerland

Christian P. KORTHALS-ALTES
The Netherlands

Centre de Physique Théorique
CNRS Luminy
F-13288 MARSEILLE Cedex 9, France

Jan KOTANSKI
Poland

Marian Smoluchowski - Institute of Physics
Jagellonian University
Reymonta 4
PL 30-059 KRAKÓW, Poland

Tomas LASTOVICKA
Czech

DESY
Platanenallee 6
D-15738 ZEUTHEN, Germany

Jonathan LEVELL
UK

Durham University
South Road
DURHAM DH1 3LE, UK

Aharon LEVY
Israel

Tel Aviv University
Ramat Aviv
69978 TEL AVIV, Israel

Martin LÜBCKE
Sweden

Uppsala University
S-75232 UPPSALA, Sweden

Axel Torsten MAAS
Germany

Technische Universität Darmstadt
Schlossgartenstrasse 9
D-64289 DARMSTADT

Tristan MAILLARD
Switzerland

ETH - Zurich
Institut for Theoretical Physics, Hönggerberg
CH-8093 ZURICH, Switzerland

Dmitri MALKEVICH
Russia

Institute for Theoretical and Experimental Physics
B. Cheremushkinskaya ulitsa 25
RU-117 259 MOSCOW, Russia

Guido MARANDELLA
Italy

Scuola Normale Superiore
Piazza dei Cavalieri 7
I-56126 PISA

Vladimir MARKOV
Russia

PNPI
Orlova Rosha 1
RU-188300, ST. PETERSBURG, Russia

Maxim MATVEEV
Russia

PNPI
Orlova Roscha 1
RU-188300, ST. PETERSBURG, Russia

Radostaw MATYSZKIEWICZ
Poland

Institute of Theoretical Physics
Warsaw University
Ulica Hoza 69
PL-00681 WARSZAWA, Poland

Alexander NESTERENKO
Russia

Joint Institute for Nuclear Research (JINR)
Bogoliubov Laboratory of Theoretical Physics
Juliot Curie Str, 6
RU-141 980 DUBNA, Moscow Region, Russia

Stefan NOBBENHUIS
The Netherlands

University of Utrecht
Leuvenlaan 4, P.O. Box 8019
NL-3508 TD UTRECHT, The Netherlands

Hiroaki OHNISHI
Japan

RIKEN
2-7 Hirosawa
WAKO, Japan

Mauro PAPINUTTO
Italy

DESY
Notkestrasse 85
D-22603 HAMBURG, Germany

Michele PAPUCCI
Italy

Scuola Normale Superiore
Piazza dei Cavalieri 7
I-56126 PISA

Jogesh PATI
India

University of Maryland
COLLEGE PARK, MD 20742, USA

John PEOPLES
USA

Fermi National Accelerator Laboratory
P.O. Box 500 - Wilson Road
BATAVIA, IL 60510-0500, USA

Dmitri PERESSOUNKO
Russia

RRC Kurchatov Institute
Kurtachov sq 1
RU-12382 MOSCOW, Russia

Claudio PICA
Italy

Università degli Studi di Pisa
Via Bonarroti 2
I-56127 PISA, Italy

Antonio POLOSA
Italy

CERN
CH-1211 GENEVA 23, Switzerland

Anton POLOUEKTOV
Russia

Budker Institute for Nuclear Physics
Lavrentiev prospect, 11
RU-630090 NOVOSIBIRSK 90, Russia

Mark RAMTOHUL
UK

University of Edinburgh
James Clerk Maxwell Building
The King's Buildings - Mayfield Road
EDINBURGH, EH9 3JZ, Scotland, UK

Takao SAKAGUCHI
Japan

Center for Nuclear Study
University of Tokyo
RIKEN 2-1 Hirosawa, Wakoshi
SAITAMA 351-0198 Japan

Gilda SCIOLI
Italy

Università degli Studi di Bologna
Via Irnerio 46
I-40126 BOLOGNA, Italy

Cristiana SEBU
Romania

Laboratoire de Physique Mathématique
Université Montpellier II
F-34095 MONTPELLIER Cedex 05, France

Stefano SELLITTO
Italy

Università degli Studi di Bologna
Via S. Allende
I-84081 BARONISSI-SALERNO, Italy

Edward V. SHURYAK
USA

State University of New York at Stony Brook
STONY BROOK, NY 11794, USA

Sergey SLIZOVSKIY
Russia

University of St. Petersburg
RU-198504 ST. PETERSBURG, Russia

Rainer STAMEN
Germany

DESY
Notkestrasse 85
D-22603 HAMBURG, Germany

Matthew STRASSLER
USA

University of Washington
PO Box 351560
SEATTLE 98195, USA

Michael TEPER
UK

University of Oxford
1 Keble Road
OXFORD, OX1 3NP, UK

Gerardus 't HOOFT
The Netherlands

Instituut voor Theoretische Fysica
Rijksuniversiteit Utrecht
NL-3584 CC UTRECHT, The Netherlands

Alessandro TORRIELLI
Italy

Università degli Studi
Via F. Marzolo 8
I-35131 PADOVA, Italy

Yoji TOTSUKA
Japan

Kamioka Observatory
University of Tokyo
Higashimozumi, Kamioka
Yoshiki-gun, GIFU-KEN 506-1205, Japan

Robert TSCHIRHART
USA

Fermi National Accelerator Laboratory
P.O. Box 500
Wilson Road
BATAVIA, IL 60510-0500, USA

Erik VERLINDE
The Netherlands

Princeton University
P.O. Box 708
PRINCETON, NJ 08544, USA

Francesca VOLPE
Italy

Università degli Studi di Bari
Via Amendola 173
I-70126 BARI, Italy

Juergen WENDLAND
Germany

Simon Fraser University
8888 University Drive
BURNABY, BC V5A 1S6, Canada

Sebastian WHITE
USA

Brookhaven National Laboratory
UPTON, NY 11973-5000, USA

Maxim ZAVATSKY
Russia

PNPI
RU-188350, Gatchina, ST. PETERSBURG, Russia

Andrea ZIROTTI
Italy

Università degli Studi di Bologna
Via Irnerio 46
I-40126 BOLOGNA, Italy

Antonino ZICHICHI
Italy

CERN
CH-1211 GENEVA 23, Switzerland

Dipartimento di Fisica - Università di Bologna
I-40126 BOLOGNA, Italy