P.W. ANDERSON

Stud. Hist. Phil. Mod. Phys., Vol. 32, No. 3, pp. 487–494, 2001
Printed in Great Britain

Pergamon

ESSAY REVIEW

Science: A 'Dappled World' or a 'Seamless Web'?

*Philip W. Anderson**

Nancy Cartwright, *The Dappled World: Essays on the Perimeter of Science* (Cambridge: Cambridge University Press, 1999), ISBN 0-521-64411-9, viii + 247 pp.

In their much discussed recent book, Alan Sokal and Jean Bricmont (1998) deride the French deconstructionists by quoting repeatedly from passages in which it is evident even to the non-specialist that the jargon of science is being outrageously misused and being given meanings to which it is not remotely relevant. Their task of 'deconstructing the deconstructors' is made far easier by the evident scientific illiteracy of their subjects.

Nancy Cartwright is a tougher nut to crack. Her apparent competence with the actual process of science, and even with the terminology and some of the mathematical language, may lead some of her colleagues in the philosophy of science and even some scientists astray. Yet on a deeper level of real understanding it is clear that she just does not get it.

Her thesis here is not quite the deconstruction of science, although she seems to quote with approval from some of the deconstructionist literature. She seems no longer to hold to the thesis of her earlier book (Cartwright, 1983) that 'the laws of physics lie'. But I sense that the present book is almost equally subversive, in that it will be useful to the creationists and to many less extreme anti-science polemicists with agendas likely to benefit from her rather solipsistic views. While allowing science some very limited measure of truth—which she

(*Received 24 January 2000; revised 3 March 2000*)
*Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544, U.S.A. (*e-mail:* pwa@pupgg.princeton.edu).

PII: S1355-2198(01)00011-9

defines as truth of each 'model' in its individual subject area within its 'shield' of carefully defined conditions set by the scientists themselves—she insists that there is no unity, no sense in which there are general laws of physics with which all must be compatible. She sees all of the laws of physics as applying only *ceteris paribus*, all other things being equal. Whenever other things are not equal, she seems to abandon the laws as any guide. In this sense, she advocates a 'dappled world' in which each aspect of reality has its separate truth and its separate 'model'.

Reading further, one senses that the problem may be that she is bogged down in eighteenth- and nineteenth-century (or even older) epistemology while dealing with twentieth-century science. The central chapters seem to depend heavily on such outdated, anthropomorphic notions as cause and effect, recurrent regularities, capacities, etc. To me, the epistemology of modern science seems to be basically Bayesian induction with a very great emphasis on its Ockham's razor consequences, rather than old-fashioned deductive logic. One is searching for the simplest schematic structure which will explain all the observations. In particular, what seems to be missing in the thinking of many philosophers of science—even the great Tom Kuhn, according to Steven Weinberg—is the realisation that the logical structure of modern scientific knowledge is not an evolutionary tree or a pyramid but a multiply-connected web.

The failure to recognise this interconnectedness becomes obvious when we are presented with 'classical Newtonian mechanics, quantum mechanics, quantum field theory, quantum electrodynamics, Maxwell's electromagnetic theory' and, in a separate place, 'fluid dynamics', as logically independent and separate rather than as, what they are, different aspects of the same physical theory, the deep interconnections among them long since solidly cemented.

Another part of the problem with this book is that the two dominant examples chosen are physics and economics, the rationale being that both sciences have 'imperialistic ambitions', the physicists aiming to provide a 'theory of everything' in the physical world, and some economists claiming universal validity in the social sphere. These two sciences, however, are on different levels of the epistemological ladder. Physicists search for their 'theory of everything', acknowledging that it will in effect be a theory of almost nothing, because it would in the end have to leave all of our present theories in place. We already have a perfectly satisfactory 'theory of everything' in the everyday physical world, which only crazies such as those who believe in alien abductions (and perhaps Bas van Fraassen) seriously doubt. The problem is that the detailed consequences of our theories are often extraordinarily hard to work out, or even in principle impossible to work out, so that we have to 'cheat' at various intermediate stages and look in the back of the book of Nature for hints about the answer. For instance, there is nothing in the quantum mechanics of the chemical bond which implies the genetic code in its detailed form, yet there is equally nothing in the operations of molecular biology which is incompatible with our quantum-mechanical understanding of

the chemical bond, intermolecular forces, and so on. In fact in the defining discovery of the field, the double helix, that understanding played a crucial role. Thus the consequences often imply the laws without the laws implying a particular set of consequences. Physics is well embedded in the seamless web of cross-relationships which is modern physical science.

Economics, on the other hand, is an example of a field which has not yet achieved interconnection of enough related information to have real objective validity. It resembles medicine before the germ theory, or biology before genetics: there are a lot of facts and relationships, but there are probably some unifying principles and connections to other sciences which are yet to be found. Yes, the Chicago school makes ambitious claims—so did the Marxists in the past. Both, to my mind, qualify as ideologies rather than theories and belong in the political sphere. There are also serious economists—with several of whom I happen to have worked, including one of the creators of the mainstream Arrow–Debreu theory—who are doing their best to discover the deeper realities that may be there, and are in conflict with the dominant school. In science as in every other field of human endeavour the rule must be *caveat emptor*: science as a whole cannot be responsible for the temporary gullibility of the public to the claims of cold fusion, Freudianism or monetarism: these are just bad, or at best incomplete, science. In sum, whenever a school of scientists creates an intellectually isolated structure which claims validation only within its own area and on its own terms—that is, does exactly what Cartwright is claiming *all* scientists do—that science no longer has the force of dynamic, self-correcting growth which is characteristic of modern science. Cartwright's 'cocoons' are an excellent description of Freudianism or behaviourism in psychology, or of the response of electrochemists to cold fusion, but do not describe healthy science.

I have some particular reason for unhappiness about the message of the book: in a very early chapter she quotes me as being opposed to my clearly stated position. My best-known work on these subjects begins with these words: 'The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of scientists it is accepted without question. The workings of our minds and bodies, and of all matter [...], are assumed to be controlled by the same set of fundamental laws, which [...] we know pretty well'. Since it is clear that I was and am one of that 'great majority', it is disingenuous of Cartwright, who is one of those 'controversial philosophers', to quote succeeding paragraphs in such a way as to arrogate me to the opposite side.

There was a second place where I can fairly competently fault her understanding. In Chapter 8 she states that she will 'take as the central example' the BCS theory of superconductivity, an area which has been extensively studied by the 'London School of Economics Modelling Project'. I have been involved with the theory (and practice) of superconductivity for 43 years—for instance, I supplied a crucial proof which is referred to in the original BCS paper as a personal communication. In 1987 I gave a lecture

studying an important and neglected part of the history of this theory, which was written up and published by Lillian Hoddeson as part of the American Institute of Physics's history of solid-state physics (Hoddeson, 1992).

My contribution was called 'It Isn't Over Till the Fat Lady Sings'. I used that crude American metaphor from the world of sport to characterise the somewhat confused period (1957–1965) between the original BCS paper, which indeed proposed a 'model', and the approximate end of the verification and validation process which made the model into what physicists properly call the 'theory' of phonon-mediated (ordinary, or BCS) superconductivity. (My usage of 'model' may be rather different from that of the LSE modelling project. What I, and physicists in general, usually mean by the term is a simplified example which is in the right universality class—for the meaning of this term, read on.) At the end of that period we were in possession (i) of a microscopic theory controlled by a small parameter, hence described by a convergent perturbation expansion about a mean-field solution which is rather like BCS, and (ii) of a detailed understanding of a much wider range of phenomenology than the Ginsburg–Landau equations could provide. This is such that the theory is no longer confined to its 'cocoon' but deals well with all kinds of messy dirt effects. (The best books describing this outcome may be Parks' two-volume compendium (1969) and de Gennes' slim book (1966), both published in the late 1960s.)

Apparently, the LSE project accepts, for much of its account of BCS and G-L, a late pedagogical description (Orlando and Delin, 1990), by two engineering-oriented authors who had no part in the above history. It is known to many historians of science that textbooks tend to caricature the real process of discovery and validation, and this is an error I regret finding here. The only original literature quoted (except for BCS and for Gor'kov's early, model-based derivation of G-L from BCS) are unsuccessful previous attempts at a theory by Bardeen himself and by Frohlich, as well as others by such luminaries as Heisenberg, Salam, Wentzel and Tomonaga. (In the process, she renames my old friend Herbert Frohlich 'Hans'.)

So: in 1957 BCS may have been describable as a *ceteris paribus* model, with no adequate account of a wide range of phenomena, or of its own limitations. It was made, by 1965, into an enormously flexible instrument with a high degree of *a priori* predictive power, and even more explanatory power. In fact, one of the deep difficulties of theorists of the new high Tc superconductors is persuading the community that, flexible as BCS is, new principles of physics must be invoked to explain the new phenomena. But as is almost always the case, the new ideas do not destroy, but instead supplement, the old. Just as the discovery of quantum chromodynamics left quantum electrodynamics firmly in place, no sensible theory of high Tc will displace BCS from its validity in ordinary metals.

The story that Cartwright misses entirely, however, is the unifying and interleaving effect the theory of superconductivity had on very widely separated areas of physics. Far from being an isolated 'model' applying only in its

shielded cocoon (as in the misfit metaphor she uses of the SQUID magnetoencephalograph in its shielded room) it was an explosive, unifying, cocoon-breaking event. First, in its own field: it showed us solid-state physicists that we could no longer safely ignore the general structure of physical theory: our familiar electrons were no longer little particles but unquestionably were quanta of a quantum field. True, in accepting the exclusion principle we should long since have realised how implausible it would be for 'particles' to be *absolutely* identical, but we had come to make casual assumptions about them. Then our particle physics friends began speculating how the vacuum itself might be a BCS state, a speculation ending in the electroweak theory. Finally, the nuclear physicists realised that we might have found the explanation for a series of puzzling phenomena observed in nuclei, and made the nucleus into a paired state. Yet another epiphany came when we predicted, and found, that the rare isotope of He would be a BCS superfluid of a new kind. So even though, in terms of the fundamental, unifying, microscopic laws, BCS made not the slightest change, it taught us a new way in which quantum fields could act, and also called our attention to the very general phenomenon of broken symmetry which is one of the key ways in which complexity can emerge from those laws.

Let us get back to the book. One of the basic epistemological points on which I differ radically from Cartwright is a very common misconception. Like many others, she maintains that the primary goal of science is prediction, prediction in the sense of being able—or at least wishing—to exactly calculate the outcome of some determinate set of initial conditions. But that is not, for instance, what an archaeologist is doing when he measures a carbon date, or a fluid dynamicist when he studies the chaotic outcome of convection in a Benard cell. Rather, each is searching for understanding. In the one case, he wishes to correlate different measurements to get some idea of the sequence of past events, which surely could never have been predicted in the quantitative sense but may enlighten him as to fundamental human behaviours. In the second case, he knows to a gnat's eyelash the equations of motion of his fluid but also knows, through the operation of those equations of motion, that the details of the outcome are fundamentally unpredictable, although he hopes to get to understand the gross behaviour. This aspect is an example of a very general reality: the existence of universal law does not, in general, produce deterministic, cause-and-effect behaviour.

Of course, in some sense there is always an aspect of prediction, in the sense that correct predictions—but by no means detailed ones—are very strong validations of the theory. If the archaeologist sees a particular kind of pottery, he may predict and then verify the carbon date; then next time he will not have to check the date. But in the epistemology which describes at least the natural sciences, I believe that the goal is exactly what Cartwright is trying to convince us is impossible: to achieve an accurate, rational, objective, and unified view of external reality. In the final section of her Chapter 2, asking 'Where Do Laws of Nature Come From?', she gives as her answer, 'always the source must be

the books of human authors and not the book of Nature'. On crucial matters she is a social constructionist.

I have argued elsewhere that this is not a tenable position, at least unless one is willing to accept total solipsism. Our perception of the everyday world is based on fragmentary, unreliable data which we only put together by creating a 'schema' or theory about the actual nature and objective existence of the various objects—chairs, mothers-in-law, teddy bears, or whatever—which we hypothesise to be in it. Then we correct, verify and validate the theory by making predictions from it (if I reach out and touch that brown thing, it will be fuzzy). Or I ask someone else to confirm my idea. Thus if we reject the inductive methods of science, we reject our only way of dealing with reality. In order to maintain our daily lives we have to accept the objective reality of the world and that it is the same world for everyone.

Why is this necessarily the case? Because we have so many cross-checks, so many consistency conditions. In the end, the schema contains many, many fewer bits of information than the data our senses gather, so we can be sure that no other theory could possibly fit. Now, we see that we can think of science as simply a somewhat more abstract, somewhat more comprehensive extension of our schema, describing the external world and compressing the enormous variety of our observations into a comprehensible whole.

The process of deconstructing the rest of the book in detail is beyond my budget of patience. The last chapter, in which she deals with the quantum measurement problem, for instance, seems to advocate one of the thousands of alternative incorrect ways of thinking about this problem that retain the quantum-classical dichotomy. My main test, allowing me to bypass the extensive discussion, was a quick, unsuccessful search in the index for the word 'decoherence' which describes the process that used to be called 'collapse of the wave function'. The concept is now experimentally verified by beautiful atomic beam techniques quantifying the whole process.

Another key word missing from the index and from the book—I checked—is renormalisation. This is not just a technical trick but a central concept in the philosophy of physics, underpinning the physicists' use of model Hamiltonians, the passage to the limit of continuum equations, and even the modern view of statistical mechanics. A 'modelling project' which has anything to do with physics should hardly ignore the way in which we build and justify our models. The renormalisation group is a way to expand the scale from the atomic to the macroscopic which shows that often the result is an enormous simplification, a division of systems into 'universality classes' all of which behave the same way in this limit; hence we may pick the simplest model for further study.

Returning to Cartwright's other exemplary subject, there is another contrast here. It is a great advantage of physics over economics that we physicists can often actually justify our use of models in this way, whereas use of the same idea in economics is almost never justifiable. An economy cannot be sorted out into macroscopic vs microscopic, with the former constructed by simply

aggregating the latter: the individual agents have foresight and are of such widely different sizes and characteristics that averaging is meaningless, even if they behaved in any mechanistic or even rational way.

There is an attack on the entire science of molecular biology in the Introduction, making the hardly very philosophical plea that the allocation of funds for genetics should be slashed in favour of preventive medicine, childcare, and other worthy causes. I could agree that a very bad glitch in the patent laws—based on not very good science—has led to a frantically accelerated search for the 'gene for this and that disease', where almost all phenomena involve the collective contributions of many genes and perhaps even of the entire genome. But while we are being feminist, are we willing to give up DNA testing? Or the heavily molecular and surprisingly successful research programme on AIDS? These are political and moral questions and have no place in a book about epistemology. Science advances by looking under the streetlight where the light is, not by 'crusades' against socially acceptable targets. The political direction of scientific strategy which she appears to advocate here has a very bad historical record in which Lysenko is only the worst recent disaster.

In summary, this book seems to show that what may have happened in the philosophy of science—or at least in this corner of the field—is precisely the kind of intellectual isolation from outside sources which elsewhere leads to bad science. There is a reluctance to accept the fact that science has become a dynamic, growing web of interrelationships both within and across fields, and that even philosophers can no longer make do without taking into account modern insights into the workings of nature and of our own mentalities. The description in this book of the process of scientific discovery, in the chapter called 'Where Do Laws of Nature Come From?', is just false from the point of view of one who has participated in it. Scientists have increasingly, and in some cases consciously, had to invent for themselves the epistemology they actually use. Scientists are not particularly able philosophers, as the case of Bohr demonstrates, but at least they are in touch with reality at first hand, and their insights into the matter have profoundly changed our understanding of how we make discoveries. In the modern state of science, no discovery lives in a cocoon, rather it is built within and upon the entire interconnected structure of what we already know.

In a sense, this is a valuable book, in that it serves as a wake-up call telling me that it is time scientists themselves examined epistemology in the light of their experience of the reality of scientific discovery. When challenged on these subjects, many of us cite Popper's ideas. Though basically right as far as they go, these now seem out of date and naive. Two scientists who have addressed these matters are Murray Gell-Mann and E. O. Wilson, and my remarks above are strongly influenced by what they have had to say. Gell-Mann, in *The Quark and the Jaguar* (1994), and even more in remarks at various workshops of the Santa Fe Institute, has emphasised the role of 'compression', while Wilson (1998) proposes the term 'consilience' for the web of interrelationships. But it is

time to take a more definitive look at why—and, of course, when—science is right.

References

Cartwright, N. (1983) *How the Laws of Physics Lie* (Oxford: Oxford University Press).
Gell-Mann, M. (1994) *The Quark and the Jaguar* (New York: W.H. Freeman).
de Gennes, P.-G. (1966) *Superconductivity of Metals and Alloys* (New York: Benjamin).
Hoddeson, L. *et al*. (eds) (1992) *Out of the Crystal Maze: Chapters from the History of Solid-State Physics* (Oxford: Oxford University Press).
Orlando, T. and Delin, K. (1990) *Foundations of Applied Superconductivity* (Reading, MA: Addison-Wesley).
Parks, R. (1969) *Superconductivity*, *2 vols* (New York: Marcel Dekker).
Sokal, A. and Bricmont, J. (1998) *Fashionable Nonsense* (New York: Picador).
Wilson, E. O. (1998) *Consilience: The Unity of Knowledge* (New York: Knopf).

1992 March Meeting of the American Physical Society
16-20 March 1992, Indianapolis, IN

March 10, 1992

When Did Dirty Solids Become Different From Clean Solids?

P.W. ANDERSON

Joseph Henry Laboratories of Physics
Jadwin Hall, Princeton University
Princeton, NJ 08544

Most of the materials you encounter in your daily life—liquids, glass, plastics, gels, alloys, colloids—are highly irregular in structure. Even the properties of the true solids of technology are controlled primarily by defects in their structure—e.g. dislocations in metals, flux lines in superconductors, domain walls in magnets. Yet 30–40 years ago, when the previous lectures left off, solid state physics—now known as condensed matter—was overwhelmingly the study of the pure crystalline state or of as good an approximation as could be managed, or of certain very simple types of defects such as color centers or shallow impurity centers.

I looked in the two major textbooks as written or revised even as late as 1976, and found no reference in one under "disorder"'—Kittel's 5th edition—and in the other (Ashcroft and Mermin) the following dismissal in a footnote: "The problem of electronic structure in disordered potentials.....is the subject of lively discussion". True enough.

The very next year a Nobel Prize was awarded in just this area, and this year's prize, awarded to Pierre Gilles de Gennes for work on almost everything but regular solids, represents perhaps the full maturation of a new field, involving complexly irregular materials, as part of the mainstream of physics. Clearly, at some time a revolution must have happened: some landmarks to mark the time scale being the first session on localization at the major semiconductor meeting in 1980, and mention—and a picture—in Kittel's '86 edition. But the existence of this new field of physics has yet to catch the attention of the world outside CM: for instance, a visitor at one of the country's oldest and most prestigious physics departments, Berkeley, was told of some bewilderment at the Nobel choice of de Gennes in 1991. And in an otherwise excellent book called "The New Physics" dedicated to postwar developments, the field was completely ignored by Paul Davies.

A book which summarizes this revolution as it was taking place was produced as the lectures from a Les Houches summer school in 1978 called "Ill-Condensed Matter"'—"La Matiére Mal Condensée"— which would have been a much better title for my talk here—and in the introduction to these lectures I wrote some words which may serve as the manifesto for this revolution.

"Multiple scattering is the paradigm of the old attitude; localization and percolation

of the new. These are phenomena which are specific to disordered systems ..... closely related is the concept of non-(or broken) ergodicity... This revolution has left us asking a whole new set of questions..... 'How do disordered systems differ from regular ones?', not, 'How can they be reduced to them?'....."

To cover even a fraction of the history related to that book is far beyond my capacity, even given indefinite time, which I'm not.. I will focus on what I see to be a few characteristic, and/or seminal bits of history, and of course I can only talk authoritatively about things I was more or less involved with so I apologize in advance for a self-centered point of view.

But first I am going to give you an overall map of some of this territory, much of which may be unfamiliar to you.

I will borrow Davies' phrase and essentially lay out for you a "Table of Contents" of a possible history of The New Physics of Ill – Condensed Matter. The table, is, hopefully, to a great extent self-explanatory; it reveals an enormous burgeoning of activity, involving the generation of many very active new fields of physics, some of which, such as mesoscopics, spin glass, esoteric phases, and defect-dominated phase transitions, are very active today; while the problem of glass itself remains, to my mind, the most profound problem, which is both well-posed and unsolved, in modern physics.

After guiding you through the general structure of this table, I'd like to focus on what I see as important and seminal events. Most of these I shall just list in Table II, and then finally try to discuss one or two of them in a truly historical manner in the sense of "how and why did they happen?"

TABLE OF CONTENTS:

THE NEW PHYSICS OF ILL-CONDENSED MATTER

II. DEFECTS IN CONDENSED PHASES

A. CLASSICAL CRYSTALLOGRAPHY : GRAIN BOUNDARIES, DISLOCATIONS, GROWTH, ETC.

'20's BURGERS, G.I. TAYLOR, ETC.

'40's THE BRISTOL SCHOOL: FRANK, MOTT, NABARRO, COTTRELL, READ, ETC.: SLIP, CREEP, FLOW

B. DEFECTS IN QUANTUM CONDENSATES

'30' – 50' (1) DOMAIN WALLS + STRUCTURES IN FM, BLOCH, LANDAU-LIFSCHITZ

'50's LANDAU-GINSBURG + THE SUPERCONDUCTING MIXED PHASE

'50's (2) VORTICES AND FLUX LINES: FEYNMAN, ('55), ONSAGER, ABRIKOSOV ('56)

C. MOTION OF DEFECTS AS DISSIPATION

'55 FEYNMAN

'62 JOSEPHSON

'63 – '65 ANDERSON

(NEUTRON STARS '74)

D. DEFECT THEORY OF PHASE TRANSITIONS

'40's SHOCKLEY - NABARRO

'55 FEYNMAN

'72 KOSTERLITZ, THOULESS

'80's HEXATICS: NELSON, ET AL

E. TOPOLOGICAL THEORY : THE REVIVAL OF LIQUID CRYSTALS

'60's F.C. FRANK; DE GENNES ('30)

'74 – '75' TOULOUSE-KLEMAN, VOLOVIK-MINEEV

III. SELF-ORGANIZED CRITICALITY ('88)???

So let's see how many of these seminal events we have time for. Clearly the earliest, and the most interesting historically, was localization.

TABLE II:

SOME SEMINAL SOURCES AND EVENTS
IN "THE NEW PHYSICS OF ILL – CONDENSED MATTER"

(1) ACCEPTANCE OF NON – ERGODICITY IN DISORDERED QUANTUM SYSTEMS

HAMMERSLEY'S PERCOLATION
+PWA+MOTT'S 20-YEARS WAR → MODERN MESOSCOPICS

(2) ACCEPTANCE OF BROKEN ERGODICITY IN SPIN GLASSES

'63–'70 EXPERIMENTS

'69'–'70 WHAT'S IN A NAME? COLES, PWA

'75 EDWARDS, THE REPLICA TRICK
(ALSO DE GENNES $n \to 0$)
FRUSTRATION THE KEY: PWA, TOULOUSE

'78 PARISI'S BRILLIANT ANSATZ

'77–'80 KIRKPATRICK, HOPFIELD, KAUFFMANN, LINK TO "COMPLEXITY"
~'80 CONFIRMATION: BOUCHIAT + OTHERS, Non-linear χ

(3) FIRST SUCCESSFUL DEFECT TH. OF PHASE TRANS :

'72 KOSTERLITZ-THOULESS

(4) TOPOLOGICAL THEORY OF DEFECTS :

'75 DE GENNES, TOULOUSE, VOLOVIK, (RELATION TO DISSIPATION AND GENERALIZED RIGIDITY: PWA, JOSEPHSON, TOULOUSE)

TABLE OF CONTENTS:

THE NEW PHYSICS OF ILL-CONDENSED MATTER

I. BROKEN ERGODICITY PHENOMENA

A. PERCOLATION : THE PARADIGM

~ '54 (J.C. Hammersley)

'75–'85 (Much later)
FRACTALS IN PHYSICS: DLA, ETC.
STRENGTH OF CONGLOMERATES, PERMEABILITY

B. LOCALIZATION

'40's MULTIPLE SCATTERING THEORY: LAX, FOLDY, EDWARDS, ETC.
VAN HOVE-KOHN-LUTTINGER TRANSPORT THEORY.

'58 FAILURE OF TRANSPORT THEORY ⟶ LOCALIZATION

'60–'80 MOTT'S 20 YEARS' WAR
HEROES: ABRAHAMS, MOTT, THOULESS, FRITSCHE AND DAVIS

'78 - - SCALING, MESOSCOPICS, CLASSICAL, QHE.

C. GLASS

'20s DEEP ROOTS: SIMON, GIAUQUE, VOGEL-FULCHER

'50's KAUZMANN

'60's RECOGNIZED AS PHYSICS PROB:
COHEN ETAL → '80's ANGELL, STEIN, SETHNA, ETC.
SIMULATIONS↓
UNSOLVED
LOW TEMPERATURE GLASS: "TLS"

'60's O.K. ANDERSON, POHL

'70's A.H.V., PHILLIPS
LEGGETT, ETC.
UNSOLVED

.....I. BROKEN ERGODICITY PHENOMENA *Continued*

D. SPIN GLASS

'55-'65 PREHISTORY AND EARLY MEASUREMENTS
OWEN, MARSHALL, BECK; WHEATLEY, MATHIAS SUHL, TOURNIER.

'69-'70 T_{SG} VIOLET-BORG; MYDOSH ETAL (NLχ)

'75-'80 SG THeory, EDWARDS-ANDERSON, TOULOUSE, PARISI
"BROKEN ERGODICITY" - PALMER

'80-'90 INFORMATION SCIENCES
SIM. ANNEAL. '77 (KIRKPATRICK), HOPFIELD ('81); GARDNER;

E. GELS, POLYMER PHASES, MICROEMULSIONS FOAMS, ETC :

'70's THE PHYSICS OF "SOFT MATTER" (DE GENNES), (ROOTS: FLORY, '50) EDWARDS, FRANK, ETC.

II. DEFECTS IN CONDENSED PHASES

A. CLASSICAL CRYSTALLOGRAPHY : GRAIN BOUNDARIES, DISLOCATIONS, GROWTH, ETC.

'20's< BURGERS, G.I. TAYLOR, ETC.

'40's THE BRISTOL SCHOOL: FRANK, MOTT, NABARRO, COTTRELL, READ, ETC.: SLIP, CREEP, FLOW

B. DEFECTS IN QUANTUM CONDENSATES

30'-50' (1) DOMAIN WALLS + STRUCTURES IN FM, BLOCH, LANDAU-LIFSCHITZ

'50's LANDAU-GINSBURG + THE SUPERCONDUCTING MIXED PHASE

'50's (2) VORTICES AND FLUX LINES: FEYNMAN, ('55), ONSAGER, ABRIKOSOV ('56)

C. MOTION OF DEFECTS AS DISSIPATION

'55 FEYNMAN

'62 JOSEPHSON

'63–'65 ANDERSON

(NEUTRON STARS '74)

.....II. DEFECTS IN CONDENSED PHASES *Continued*

D. DEFECT THEORY OF PHASE TRANSITIONS

'40's SHOCKLEY - NABARRO

'55 FEYNMAN

'72 KOSTERLITZ, THOULESS

'80's HEXATICS: NELSON, ET AL

E. TOPOLOGICAL THEORY : THE REVIVAL OF LIQUID CRYSTALS

'60's F.C. FRANK; DE GENNES ('30)

'74-75' TOULOUSE-KLEMAN, VOLOVIK-MINEEV

III. SELF-ORGANIZED CRITICALITY ('88)???

TABLE II:

SOME SEMINAL SOURCES AND EVENTS IN "THE NEW PHYSICS OF ILL-CONDENSED MATTER"

(1) ACCEPTANCE OF NON-ERGODICITY IN DISORDERED QUANTUM SYSTEMS:

HAMMERSLEY'S PERCOLATION
+PWA+MOTT'S 20-YEARS WAR → MODERN MESOSCOPICS

(2) ACCEPTANCE OF BROKEN ERGODICITY IN SPIN GLASSES:

'63–'70 EXPERIMENTS

'69–'70 WHAT'S IN A NAME? COLES, PWA

'75 EDWARDS, THE REPLICA TRICK
(ALSO DE GENNES $n \to 0$)
FRUSTRATION THE KEY: Pwa, TOULOUSE

'78 PARISI'S BRILLIANT ANSATZ

'77–'80 KIRKPATRICK, HOPFIELD, KAUFFMANN, LINK TO "COMPLEXITY"

~'80 CONFIRMATION: BOUCHIAT + OTHERS, Non-linear χ

(3) FIRST SUCCESSFUL DEFECT TH. OF PHASE TRANS:

'72 KOSTERLITZ-THOULESS

(4) TOPOLOGICAL THEORY OF DEFECTS:

'75 DE GENNES, TOULOUSE, VOLOVIK,
(RELATION TO DISSIPATION AND GENERALIZED RIGIDITY: PWA, JOSEPHSON, TOULOUSE)

October 10, 1989

Some Thoughtful Words (not mine) on Research Strategy for Theorists

P.W. ANDERSON

Joseph Henry Laboratories of Physics

Jadwin Hall, Princeton University

Princeton, NJ 08544

I quote from one of the greatest theoretical physicists of the postwar era:

"The principal error I see in most current theoretical work is that of imagining that a theory is really a good model for ... nature rather than being merely a demonstration (of possibility)—a 'don't worry' theory. Theorists almost always become too fond of their own ideas... It is difficult to believe that one's cherished theory, which really works rather nicely, may be completely false. The basic trouble is that many quite different theories can go some way to explaining the facts. If elegance and simplicity are... dangerous guides, what constraints can be used as a guide through the jungle of possible theories?... The only useful constraints are contained in the experimental evidence. Even this information is not without its hazards, since experiment 'facts' are often misleading or even plain wrong. It is thus not sufficient to have rough acquaintance with the evidence, but rather a deep and critical knowledge of many different types, since one never knows knows what type of fact is likely to give the game away...

Theorists... should realize that it is extremely unlikely that they will produce a useful theory just by having a bright idea distantly related to what they imagine to be the facts. Even more unlikely is that they will produce a good theory at their first attempt... they have to produce theory after theory... The very process of abandoning theories gives them a degree of critical detachment which is almost essential."

The missing words indicated by dots would give the game away, that this is Sir Frances Crick talking about theory in biology, at the conclusion of his autobiography, "What Mad Pursuit". He, in fact, distinguishes biological theory from physical theory on the basis that the mechanisms arise from the complex process of evolution. But in the absence of definitive advice on this matter from

such other successful theorists as Crick's contemporaries, Richard Feynman and Murray Gell-Mann, it seems to me that one should, perhaps, take him more seriously as a guide to how theory is actually done than he may himself do. After all, in physical theory, we now know that whether or not the original cosmic egg was as scrambled as some astrophysicists such as Linde seem to think it was, almost all the phenomena we study, both in condensed matter and in particle theory, are the result of emergent processes and broken symmetries nearly as complex and evolutionary as biology.

My own experience has certainly been that most successful theories are the result of successive corrections to errors that may verge on the ludicrous, corrections normally dictated by a careful look at experiment. The long and tortuous tale I have told elsewhere of spin glass is one example; another is localization—who could have guessed, even in 1978 after certain prizes had been given out, that potential scattering, spin-orbit scattering, and magnetic scattering would turn out to give qualitatively different localization phenomena? Localization, in the presence of a magnetic field, seemed simple at first—until the experimentalists showed us that it led to the utterly unlikely phenomenon of Hall resistance quantization, leaving us theorists scurrying to catch up. In another example familiar to me, at least, the right A phase of superfluid helium three was predicted by solving the wrong Hamiltonian in the wrong way. Yet that is, too, a delightful example of Crick's "demonstration" theory: that paper demonstrated that phases of different symmetry were possible, which, in the end, turned out to be the really useful and important conceptual result.

Young theorists in my field, especially, would do well also to take Crick's words about experiment to heart. They often seem to believe that there is some

kind of "Miranda rule" about what kind of evidence is admissible. Theorists discuss theory either in an experimental vacuum, or in relation to experiments endorsed by some previous paper or produced by the most fashionable experimental methods, rather than searching out the anomalies which are the real guide to the truth.

As I see it, even the "standard model" of particle theory—like it or not—was arrived at by the same kind of random walk guided at every stage by experiment, and many of its features still seem to have been as unpredictable on the basis of general principles of elegance or simplicity as the convolutions of biological evolution.

In conclusion, it appears that in all its branches physics is still an experimental science. Its basic goal is not mathematical elegance or the achievement of tenure, but learning the truth about the world around us, and Sir Francis Crick's words are as good a guide to that end as I have seen.

D.P. AROVAS

Fractional Statistics and the Quantum Hall Effect

Daniel Arovas
Department of Physics, University of California, Santa Barbara, California 93106

and

J. R. Schrieffer and Frank Wilczek
Department of Physics and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

(Received 18 May 1984)

The statistics of quasiparticles entering the quantum Hall effect are deduced from the adiabatic theorem. These excitations are found to obey fractional statistics, a result closely related to their fractional charge.

PACS numbers: 73.40.Lq, 05.30.−d, 72.20.My

Extensive experimental studies have been carried out[1] on semiconducting heterostructures in the quantum limit $\omega_0\tau >> 1$, where $\omega_0 = eB_0/m$ is the cyclotron frequency and τ is the electronic scattering time. It is found that as the chemical potential μ is varied, the Hall conductance $\sigma_{xy} = I_x/E_y = \nu e^2/h$ shows plateaus at $\nu = n/m$, where n and m are integers with m being odd. The ground state and excitations of a two-dimensional electron gas in a strong magnetic field B_0 have been studied[2-4] in relation to these experiments and it has been found that the free energy shows cusps at filling factors $\nu = n/m$ of the Landau levels. These cusps correspond to the existence of an "incompressible quantum fluid" for given n/m and an energy gap for adding quasiparticles which form an interpenetrating fluid. This quasiparticle fluid in turn condenses to make a new incompressible fluid at the next larger value of n/m, etc.

The charge of the quasiparticles was discussed by Laughlin[2] by using an argument analogous to that used in deducing the fractional charge of solitons in one-dimensional conductors.[5] He concluded for $\nu = 1/m$ that quasiholes and quasiparticles have charges $\pm e^* = \pm e/m$. For example, a quasihole is formed in the incompressible fluid by a two-dimensional bubble of a size such that $1/m$ of an electron is removed. Less clear, however, is the statistics which the quasiparticles satisfy; Fermi, Bose, and fractional statistics having all been proposed. In this Letter, we give a direct method for determining the charge and statistics of the quasiparticles.

In the symmetric gauge $\vec{A}(\vec{r}) = \frac{1}{2}\vec{B}_0 \times \vec{r}$ we consider the Laughlin ground state with filling factor $\nu = 1/m$,

$$\psi_m = \prod_{j<k}(z_j - z_k)^m \exp(-\tfrac{1}{4}\textstyle\sum_l |z_l|^2), \tag{1}$$

where $z_j = x_j + iy_j$. A state having a quasihole localized at z_0 is given by

$$\psi_m^{+z_0} = N_+ \textstyle\prod_i (z_i - z_0)\psi_m, \tag{2}$$

while a quasiparticle at z_0 is described by

$$\psi_m^{-z_0} = N_- \textstyle\prod_i (\partial/\partial z_i - z_0/a_0^2)\psi_m, \tag{3}$$

where $2\pi a_0^2 B_0 = \phi_0 = hc/e$ is the flux quantum and $N_\pm$ are normalizing factors.

To determine the quasiparticle charge e^*, we calculate the change of phase γ of $\psi_m^{+z_0}$ as z_0 adiabatically moves around a circle of radius R enclosing flux ϕ. To determine e^*, γ is set equal to the change of phase,

$$(e^*/\hbar c)\oint \vec{A}\cdot d\vec{l} = 2\pi(e^*/e)\phi/\phi_0, \tag{4}$$

that a quasiparticle of charge e^* would gain in moving around this loop. As emphasized recently by Berry[6] and by Simon[7] (see also Wilczek and Zee[8] and Schiff[9]), given a Hamiltonian $H(z_0)$ which depends on a parameter z_0, if z_0 slowly transverses a loop, then in addition to the usual phase $\int^t E(t')\,dt'$, where $E(t')$ is the adiabatic energy, an extra phase γ occurs in $\psi(t)$ which is independent of how slowly the path is traversed. $\gamma(t)$ satisfies

$$d\gamma(t)/dt = i\langle\psi(t)|d\psi(t)/dt\rangle \ . \tag{5}$$

From Eq. (2),

$$\frac{d\psi_m^{+z_0}}{dt} = N_+ \textstyle\sum_i \frac{d}{dt}\ln[z_i - z_0(t)]\psi_m^{+z_0}, \tag{6}$$

so that

$$\frac{d\gamma}{dt} = iN_+^2\left\langle\psi_m^{+z_0}\middle|\frac{d}{dt}\textstyle\sum_i \ln(z_i - z_0)\middle|\psi_m^{+z_0}\right\rangle. \tag{7}$$

Since the one-electron density in the presence of

the quasihole is given by

$$\rho^{+z_0}(z) = \langle \psi_m^{+z_0} | \sum_i \delta(z_i - z) | \psi_m^{+z_0} \rangle, \tag{8}$$

we have

$$\frac{d\gamma}{dt} = i \int dx\, dy\, \rho^{+z_0}(z) \frac{d}{dt} \ln[z - z_0(t)], \tag{9}$$

where $z = x + iy$. We write $\rho^{+z_0}(z) = \rho_0 + \delta\rho^{+z_0}(z)$, with $\rho_0 = \nu\phi/\phi_0$. Concerning the ρ_0 term, if z_0 is integrated in a clockwise sense around a circle of radius R, values of $|z| < R$ contribute $2\pi i$ to the integral while $|z| > R$ contributes zero. Therefore, this contribution to γ is given by

$$\gamma_0 = i \int_{|r|<R} dx\, dy\, \rho_0 2\pi i = -2\pi \langle n \rangle_R = -2\pi\nu\phi/\phi_0, \tag{10}$$

where $\langle n \rangle_R$ is the mean number of electrons in a circle of radius R. Corrections from $\delta\rho$ vanish as $(a_0/R)^2$, where $a_0 = (\hbar c/eB)^{1/2}$ is the magnetic length. This term corresponds to the finite size of the hole.

Comparing with Eq. (4), we find $e^* = \nu e$, in agreement with Laughlin's result. A similar analysis shows that the charge of the quasiparticle $\psi_m^{-z_0}$ is $-e^*$.

To determine the statistics of the quasiparticles, we consider the state with quasiholes at z_a and z_b,

$$\psi_m^{z_a, z_b} = N_{ab} \prod_i (z_i - z_a)(z_i - z_b) \psi_m. \tag{11}$$

As above, we adiabatically carry z_a aroound a closed loop of radius R. If z_b is outside the circle $|z_b| = R$ by a distance $d >> a_0$, the above analysis for γ is unchanged, i.e., $\gamma = -2\pi\nu\phi/\phi_0$. If z_b is inside the loop with $|z_b| - R << -a_0$, the change of $\langle n \rangle_R$ is $-\nu$ and one finds the extra phase $\Delta\gamma = 2\pi\nu$. Therefore, when a quasiparticle adiabatically encircles another quasiparticle an extra "statistical phase"

$$\Delta\gamma = 2\pi\nu \tag{12}$$

is accumulated.[10] For the case $\nu = 1$, $\Delta\gamma = 2\pi$, and the phase for interchanging quasiparticles is $\Delta\gamma/2 = \pi$ corresponding to Fermi statistics. For ν noninteger, $\Delta\gamma$ corresponds to fractional statistics, in agreement with the conclusion of Halperin.[11] Clearly, when ν is noninteger the change of phase $\Delta\gamma$ when a third quasiparticle is in the vicinity will depend on the adiabatic path taken by the quasiparticles as they are interchanged and the pair permutation definition used for Fermi and Bose statistics no longer suffices.

A convenient method for including the statistical phase $\Delta\gamma$ is by adding to the actual vector potential $\vec{A}_0$ a "statistical" vector potential $\vec{A}_\phi$ which has no independent dynamics. $\vec{A}_\phi$ is chosen such that

$$(e^*/\hbar c) \oint \vec{A}_\phi \cdot d\vec{l} = \Delta\gamma = 2\pi\nu, \tag{13}$$

when z_a encirlces z_b. One finds this fictious $\vec{A}_\phi$ to be

$$\vec{A}_\phi(\vec{r} - \vec{r}_b) = \frac{\phi_0 \hat{z} \times (\vec{r} - \vec{r}_b)}{2\pi |\vec{r} - \vec{r}_b|^2} \tag{14}$$

if the quasiparticles are treated as bosons and $\phi_0 \rightarrow \phi_0(1 - 1/\nu)$ if they are treated as fermions. Thus, the peculiar statistics can be replaced by a more complicated effective Lagrangian describing particles with conventional statistics.[12]

Finally, we note that if one pierces the plane with a physical flux tube of magnitude ϕ, the above arguments suggest that a charge $\nu e\phi/\phi_0$ is accumulated around the tube, regardless of whether ϕ/ϕ_0 is equal to the ratio of integers.

This work was supported in part by the National Science Foundation through Grant No. DMR82-16285 and No. PHY77-27084, supplemented by funds from the National Aeronautics and Space Administration. One of us (D.A.) is grateful for the support of an AT&T Bell Laboratories Scholarship.

[1]K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

[2]R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).

[3]F. D. M. Haldane, Phys. Rev. Lett. **51**, 605 (1983).

[4]B. I. Halperin, Institute of Theoretical Physics, University of California, Santa Barbara, Report No. NSF-ITP-83-34, 1983 (to be published).

[5]W. P. Su and J. R. Schrieffer, Phys. Rev. Lett. **46**, 738 (1981).

[6]M. V. Berry, Proc. Roy. Soc. London, Ser. A **392**, 45–57 (1984).

[7]B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).

[8]F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).

[9]L. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), p. 290.

[10]Although ψ is a variational wave function, rather than the actual adiabatic wave function, the statistical properties of the quasiparticles are not expected to be sensitive to this inconsistency. We could regard ψ to be an exact excited-state wave function for a model Hamiltonian.

[11]B. I. Halperin, Phys. Rev. Lett. **52**, 1583, 2390(E) (1984).

[12]F. Wilczek and A. Zee, Institute of Theoretical Physics, University of California, Santa Barbara, Report No. NSF-ITP-84-25, 1984 (to be published).

"Fractional Statistics" in Arbitrary Dimensions: A Generalization of the Pauli Principle

F. D. M. Haldane[a]

Department of Physics, Princeton University, Princeton, New Jersey 08544

(Received 19 March 1991)

The concept of "fractional statistics" is reformulated as a generalization of the Pauli exclusion principle, and a definition independent of the dimension of space is obtained. When applied to the vortexlike quasiparticles of the fractional quantum Hall effect, it gives the same result as that based on the braid group. It is also used to classify spinons in gapless spin-$\frac{1}{2}$ antiferromagnetic chains as semions. An extensive one-particle Hilbert-space dimension is essential, limiting fractional statistics of this type to topological excitations confined to the interior of condensed matter. The new definition does not apply to "anyon gas" models as currently formulated: A possible resolution of this difficulty is proposed.

PACS numbers: 05.30.−d, 71.10.+x

The concept of "anyons" or particles with "fractional statistics" in two-dimensional (2D) systems [1] has been a subject of intense study in recent years, and has found application in the theory of the fractional quantum Hall effect [2] (FQHE). More recently, it has formed the basis of the theory of "anyon superconductivity" [3]. The anyon concept is essentially two dimensional; however, a recent study [4] of "spinon" excitations in *one*-dimensional antiferromagnets has led me to a variant notion of fractional statistics which, like the conventional Bose and Fermi statistics, is formulated without specific reference to spatial dimension.

This new definition can be considered as a generalization of the Pauli exclusion principle; when applied to the FQHE, it coincides with the now-standard 2D anyon definition [1] in terms of the braiding of particle trajectories; in general, however, the two definitions are *not* equivalent. The definition proposed here requires that single-particle Hilbert-space dimensions are *extensive,* which is a common property of, e.g., topological excitations of a condensed-matter state, but not of the flux-carrying Newtonian point particles of the "anyon gas" model [1]. The statistics as defined here is *not* affected by the attachment to the particles of gauge fields which conventionally lead to "statistical transmutation" of long-distance, low-energy properties: For example, "hard-core lattice bosons" would be classified as fermions (that carry a gauge field), despite their bosonlike low-energy properties, because of their fermionlike exclusion principle and band-filling property.

Consider the Hilbert space $\mathcal{H}_\alpha$ of states of a single particle of species α, confined to a finite region of matter, where this space is spanned by wave functions $\phi_\nu^\alpha(r)$, $\nu=1,\ldots,d_\alpha$. It will be crucial for this discussion that the dimension $\dim[\mathcal{H}_\alpha]=d_\alpha$ is *finite and extensive,* proportional to the size of the condensed-matter region in which the particle exists. This implies that the "particle" is an *elementary excitation that can only exist in the interior of a region of condensed matter,* and is not an elementary particle which can have arbitrarily large momentum, and exist in the vacuum outside the condensed-matter region. In general, the fractional-statistics particle will be a topological excitation of a condensed-matter state.

Now consider the wave function of an N-particle system of such particles with coordinates and species indices $\{r_i,\alpha_i;\ i=1,\ldots,N\}$ and let N_α be the number of particles of species α. If the coordinates of the $N-1$ particles with labels $j\neq i$ are held fixed, the wave function $\Psi(r_1,\ldots,r_N)$ can be expanded in a basis of wave functions of the ith particle as

$$\sum_\nu A_\nu(\{r_j,\alpha_j;j\neq i\})\phi_\nu^{\alpha_i}(r_i;\{r_j,\alpha_j;j\neq i\})\,. \tag{1}$$

The set of wave functions $\phi_\nu^\alpha(r;\{r_j,\alpha_j;j\neq i\})$ span a one-particle Hilbert space $\mathcal{H}_\alpha(\{r_j,\alpha_j;j\neq i\})$, with dimensions $d_\alpha(\{N_\alpha\})$. This dimension must be independent of the *coordinates* of the particles with labels $j\neq i$, and must be the same for all identical particles of the same species. It will depend only on the boundary conditions and size of the condensed-matter region and the numbers $\{N_\alpha\}$ of the different particle species.

In general, d_α will change as particles are added, *while keeping the boundary conditions and size of the condensed-matter region constant.* This will provide the basic notion of statistics developed here. I define the *statistical interaction* $g_{\alpha\beta}$ through the differential relation

$$\Delta d_\alpha=-\sum_\beta g_{\alpha\beta}\Delta N_\beta\,, \tag{2}$$

where $\{\Delta N_\beta\}$ is a set of allowed changes of the particle numbers at fixed size and boundary conditions. Conventional bosons have $g_{\alpha\beta}=0$, while the Pauli exclusion principle for fermions corresponds to $g_{\alpha\beta}=\delta_{\alpha\beta}$. The *relation (2) may be considered as a generalization of the Pauli principle.*

In order for a thermodynamic limit to exist with extensive single-particle Hilbert-space dimensions d_α, the statistical interactions $g_{\alpha\beta}$ must be independent of the numbers of particles. The existence of a thermodynamic limit also requires that the $g_{\alpha\beta}$ are *rational,* so that limit can be achieved through a sequence of proportional finite increments of the size of the system and the particle numbers.

For state-counting purposes *at fixed particle numbers* the particles can be regarded as bosons with a Fock-space

dimension $d_B = d_\alpha$ or fermions with a Fock-space dimension $d_F = d_\alpha + N_\alpha - 1$. In either interpretation, the total size of the full Hilbert space of many-particle states, at fixed $\{N_\alpha\}$, will be

$$\prod_\alpha \frac{(d_\alpha + N_\alpha - 1)!}{(N_\alpha)!(d_\alpha - 1)!} . \tag{3}$$

The construction of a fermionlike or bosonlike description of the states thus does *not* imply anything about the statistics: The particles are only true bosons (fermions) if their effective Fock-space dimension d_B (d_F) *remains constant as the number of particles is changed.*

Unless this condition is satisfied, the conventional techniques of second-quantized many-body theory cannot be applied. Arguably, the key step in Laughlin's [5] seminal treatment of the FQHE, so far the only established physical application of fractional statistics [2], was to abandon conventional second-quantized methods, which had proved fruitless, and return to a first-quantized description.

In the FQHE, the Pauli-like definition of statistics introduced here can be related to the braid-group notion of 2D statistics [1]. The basic model for 2D anyons is the charged flux tube [1]. I will use units where $e = \hbar = c = 1$, in which the London flux quantum Φ_0 is 2π. If an object carrying flux ϕ_1 and charge q_1 orbits around another object carrying flux ϕ_2 and charge q_2, the Bohm-Aharonov phase change of the wave function is

$$\exp(i\theta_{12}) = \exp[-i(q_1\phi_2 + q_2\phi_1)] . \tag{4}$$

If they are identical, the Bohm-Aharonov phase factor for interchange in $\exp(i\theta_1) \equiv \exp(i\theta_{11}/2)$. Anyons with an aribitrary statistical parameter θ_α are modeled by bosonic flux tubes with $q_\alpha\phi_\alpha = \theta_\alpha$.

In the FQHE, the quasiparticles of the Laughlin states have the character of vortices [5], with dynamics derived from quantizing the Eulerian dynamics of point vortices in an incompressible, inviscid fluid (which do not carry kinetic energy), rather than the Newtonian dynamics of a massive particle. The "guiding center" coordinates of the vortex do not commute:

$$[R_i^\mu, R_j^\mu] = il^2\delta_{ij}q_i\epsilon^{\mu\nu} . \tag{5}$$

Here $2\pi l^2$ is the area per particle of the underlying fluid; the integer q_i is the circulation of the ith vortex in units of the elementary circulation quantum of the fluid.

The commutation relations (5) imply [6] that the vortex cannot be localized in an area smaller than the mean area per fluid particle, and states representing the vortex centered at different point are *nonorthogonal*; for a fluid with open boundary conditions, the number of independent states of a vortex is $N+1$, where N is the number of particles in the underlying fluid. These particles act as quantized sources of "flux" ($\phi = 2\pi$ per particle); (5) implies that if an elementary quantum vortex carrying "charge" ± 1 moves around a closed path, its wave function picks up a phase factor that counts the mean number of particles enclosed by that path. (This complements the result that a fluid particle moving around a closed loop picks up a phase that counts the total vorticity inside the loop, forcing quantization of the circulation around a vortex [6].) The model for a fractional-statistics vortex is one which locally has a mean excess or deficit of fluid particles as compared to the mean fluid density in its absence. This is only well defined in an incompressible fluid [6].

The noncommutativity of guiding-center coordinates makes vortex wave functions equivalent to the *lowest-Landau level states of a charged particle in a magnetic field.* The FQHE quasiparticles can thus be modeled by flux-carrying charged bosons in the lowest Landau level. If there is a total flux Φ through the system, the number of independent single-particle states in the Landau level for bosonic particles carrying charge q_α is $d_\alpha = q_\alpha\Phi/2\pi$. If the particles carry flux ϕ_α, the total flux Φ, and hence the $\{d_\alpha\}$, change as particles are added, and $g_{\alpha\beta} = -q_\alpha\phi_\beta/2\pi$. The relative statistical phases $\theta_{\alpha\beta}$ for windings of particle trajectories are thus identified as $\exp(i\theta_\alpha) = \exp(\zeta i\pi g_{\alpha\alpha})$, and $\exp(i\theta_{\alpha\beta}) = \exp[\zeta i\pi(g_{\alpha\beta} + g_{\beta\alpha})]$. (An overall handedness $\zeta = \pm 1$ remains undefined.) Note that the statistical interactions $g_{\alpha\beta}$ convey more information than the statistical phases, which are ambiguous modulo 2π.

The Laughlin FQHE states at the primary Landau-level fillings $\nu = 1/m$ have two vortexlike excitations, quasiparticles and quasiholes. In this context, fixing the boundary conditions means fixing $BA/\Phi_0 = N_\Phi$, the total magnetic flux (in units of the flux quantum) passing through the system. Let there be N^+ quasiparticles and N^- quasiholes: The number of electrons N is then given by [7] $N_\Phi = m(N-1) + N^- - N^+$. The Hilbert-space dimension $d_\pm$ for both quasiparticles and quasiholes is [7] $N+1$. Changing N^+ and N^- by multiples of m at fixed N_Φ gives $g_{\alpha\pm} = \mp 1/m$, $\alpha = \pm$. The agreement between the Hilbert-space counting definition of statistics and the anyon definition in this case is no accident: Like the statistical phase, the Hilbert-space dimension can be obtained from the Bohm-Aharonov-like Berry's phase for adiabatic transport of a quasiparticle around a loop, in this case one encircling the entire fluid.

The dimension-independent definition (2) of fractional statistics opens up possibilities of non-2D applications. In principle, fractional statistics would be recognized by the presence of bands with an unusual (and variable) number of single-particle states, not given by simply counting the number of unit cells or atoms in the condensed-matter system.

As an example not restricted to two spatial dimensions, I consider spinon excitations in a spin-$\frac{1}{2}$ quantum antiferromagnet with a *nondegenerate* singlet "resonating valence bond" (RVB) ground state [8] without magnetic order: These may be thought of as isolated unpaired spins [9] in a RVB background of paired spins. The spin-

on is a spin-$\frac{1}{2}$ object coming in two species, labeled by $\sigma = \pm\frac{1}{2}$. If there are N spins and N_{sp} spinons, the number of unbroken bonds is $(N-N_{sp})/2$, which must be an integer. I identify the spinon Hilbert-space dimension d_σ as $1+(N-N_{sp})/2$, independent of σ.

This can be understood as follows: A given spinon can occupy the site it is initially on, or it can be moved to a site that is part of a pair. But if $|1(23)\rangle$ represents a three-site wave function where site 1 is unpaired and sites 2 and 3 are paired, nonorthogonality means that $|1(23)\rangle = (1/\sqrt{2})[|2(13)\rangle - |3(12)\rangle]$. Hence only the combination $[|2(13)\rangle + |3(12)\rangle]$ is independent of $|1(23)\rangle$. There is thus only *one* extra independent spinon state per bond. *This reduction of apparent Hilbert-space dimension by nonorthogonality of states describing localized topological defects at different points in space is also seen in the FQHE example, and seems to be the fundamental feature of fractional statistics as defined here.*

The statistical interaction between spinons is thus given by $g_{\sigma\sigma'} = \frac{1}{2}$, independent of spin. It can now be verified that the full Hilbert space of the spin system is spanned by the semionic many-spinon states, with no over or under counting. The total number of states is obtained using (3). At a given spinon number, the number of many-spinon states is equivalent to the number of ways to place $N_{sp} = N_{sp\uparrow} + N_{sp\downarrow}$ bosons in $2\times[1+(N-N_{sp})/2]$ orbitals; taking the background RVB state to be nondegenerate [10], the full Hilbert-space dimension is

$$\tfrac{1}{2}\sum_{N_{sp}}[1+(-1)^{(N-N_{sp})}]\frac{(N+1)!}{N_{sp}!(N-N_{sp}+1)!}. \qquad (6)$$

This is 2^N, as expected, showing completeness of the spinon description. Note that this discussion has not involved properties of the Hamiltonian, except perhaps through the assumption of a nondegenerate ground state. However, the spinon description will only be useful if the elementary excitations do indeed have spinon character.

Explicit examples of the above scenario are provided by spin-$\frac{1}{2}$ antiferromagnetic Heisenberg spin chains, in the gapless phase which at low energies is described by the level $k=1$ SU(2) Wess-Zumino-Witten conformal field theory (CFT) [11]. One example is the nearest-neighbor-exchange Heisenberg chain solved by Bethe [12]; however, the clearest and most explicit example is provided by the $S=\frac{1}{2}$ Heisenberg chain with inverse-square exchange [4,13,14] (ISE model), which also generates the $k=1$ CFT. The ISE model has the same state-counting rules as Bethe's model, but has a much simpler spectrum, allowing its thermodynamics to be constructed explicitly in closed form [4]. It is also deeply related to Laughlin's $\nu=\frac{1}{2}$ boson FQHE state [5], making a fractional-statistics interpretation very natural.

If N is the (even) number of sites, the spin-singlet ground-state wave function of the ISE model can be written in two forms [13] directly related to Laughlin FQHE wave functions [5]: The first form (expressed in terms of the $N/2$ sites $\{n_i\}$ "occupied" by reversed spins) is the Laughlin-Kalmeyer form [15], now defined on a 1D rather than a 2D lattice. This is a direct transcription of the $\nu=\frac{1}{2}$ bosonic FQHE wave function. If $z(n) = \exp(2\pi i n/N)$,

$$\Psi = \prod_{i<j}[z(n_i)-z(n_j)]^2\prod_i z(n_i). \qquad (7)$$

The other form is the $n=2$ case of the SU(n) singlet wave function where $\langle\sigma_1,\ldots,\sigma_N|\Psi\rangle$ vanishes unless $\sum\delta_{\sigma,\sigma_n} = N/n$, when it obtains from the antisymmetric Slater determinant function $\Psi(\{z_m,\sigma_m\})$ given by

$$\prod_{m<n}(z_m-z_n)^{\delta(\sigma_m,\sigma_n)}(i)^{\mathrm{sgn}(\sigma_m-\sigma_n)}, \qquad (8)$$

with "spatial coordinates" $\{z_m\}$ chosen to be a permutation of $\{z(m), m=1,\ldots,N\}$. Up to a spin-independent factor, (8) is the ground state of SU(n) fermions in 1D or filling the lowest Landau level in 2D [16].

The great simplifying feature of the ISE model is that there are *no* spin-dependent interactions between the spinons [4]. The spinon states form a band of $1+(N-N_{sp})/2$ states with a spin-independent statistical interaction $g_{\sigma\sigma'}=\frac{1}{2}$, just as predicted by the general argument given here. For a fixed N_{sp}, a spin-$\frac{1}{2}$ bosonic Fock space is the most appropriate description: The eigenstates are characterized by sets of occupation numbers, just as in the case of the ideal Fermi or Bose gases, except that the energy is a quadratic (instead of linear) function of the occupations. The full solution [4] for the thermodynamics of the ISE model shows that occupation-number distributions *in a Fock space that varies with temperature as the number of fractional-statistics excitations changes* have a role to play in the theory of fractional-statistics systems.

Because of spin exchange between spinons (which is marginally irrelevant at low energies) the Bethe-ansatz model has a much more complicated solution; however, from a study [17] of the adiabatic interpolation between the two models, I have established that the total number of complex rapidity strings in the Bethe-ansatz solution *(irrespective of their length)* has the simple interpretation as the number of unbroken valence bonds, $(N-N_{sp})/2$.

The ISE-model spinon band covers half the Brillouin zone (BZ), shrinks as spinons are added, and is gapless at its termination points. If an analogous RVB ground state can occur in some higher-dimensional model *without symmetry breaking that reduces the BZ volume, or leads to a degenerate ground state* [10], the result for d_σ obtained here requires the existence of a surface enclosing half the BZ volume which would mark the gapless boundary of the spin band. (The normalization of Bloch-state combinations of localized spinon states will diverge as this boundary is approached from the interior.) This scenario resembles that of the "pseudo-Fermi surface" which Anderson has predicted [8] to characterize a gapless 2D RVB state. The ideas developed here suggest that gap-

less Fermi-surface-like structures, and a generalized, *fractional* Luttinger-type relation between volume enclosed by such surfaces and quasiparticle number, may be a new type of collective behavior in two or more dimensions.

In the examples discussed so far, the fractional statistics emerges in the properties of elementary excitations of models built microscopically out of conventional objects such as electrons. In the anyon gas model [1], a Chern-Simons gauge field is added "by hand" to a model of conventional particles to produce Bohm-Aharonov phases as the particles orbit each other. A lattice-gas version of the model which satisfies the Hilbert-space extensivity condition consists of hard-core bosons (or spinless fermions) to which Chern-Simons flux is attached [18]. Since addition of a coupling to a gauge field does not affect Hilbert-space dimensions, such particles are classified as fermions by the definition proposed here.

The difference between the two definitions of statistics in 2D—when applied to the anyon gas model—is troubling: However, I note that the model has no true microscopic derivation. A possible resolution of the discrepancy may be conjectured: The appearance of a Chern-Simons field in the effective low-energy description of topological excitations of a 2D condensed-matter state of conventional particles may *inevitably* (as in the FQHE) be accompanied by nonorthogonality of localized states of the topological excitations, which reconciles the two definitions.

The ingredients of a modified "lattice anyon" model with an analog of Euler dynamics (appropriate to vortices defined on the plaquettes of a dual lattice) can be identified: If i is a "site" (plaquette) label, they are (a) a site-diagonal one-body Hamiltonian $H=\epsilon_0\sum_i|i\rangle\langle i|$ and (b) *nonorthogonality*, $\langle i|j\rangle=S_{ij}\neq\delta_{ij}$, which replaces kinetic energy as the generator of dynamics. The Hermitian overlap matrix S_{ij} has real non-negative eigenvalues; the number of states in the band is the number of *nonzero* eigenvalues S_ν, and the corresponding eigenvalues of the one-body Hamiltonian are $\epsilon_0 S_\nu$. If the rank of the overlap matrix is less than its dimension, the number of eigenstates of H is less than the number of sites. In the many-body case, the overlap matrix for a given particle (i.e., vortex) would depend on the number and positions of other particles, introducing both gauge interactions and fractional statistics as defined here.

In summary, I have introduced a variant definition (2) of fractional statistics that can be viewed as a generalization of the Pauli principle, and does not make reference to spatial dimension. It produces consistent results when applied to 2D Laughlin FQHE quasiparticles and spinons in nondegenerate RVB states of 1D $S=\frac{1}{2}$ quantum antiferromagnets. However, it does *not* apply to 2D models where Chern-Simons flux has been attached "by hand" to conventional particles. Such models lack what appears to be an essential element of fractional statistics as defined here: nonorthogonality of localized particle states.

Finally, I note that temperature appears to play no role in this approach, in contrast to a recent proposal [19] in which "statistics" is identified with "local fermionic charge."

This work was supported in part by NSF Grant No. DMR-8901985.

[a]On leave of absence from Department of Physics, University of California, San Diego, La Jolla, CA 92093.

[1] J. M. Leinaas and J. Myrheim, Nuovo Cimento **37B**, 1 (1977); F. Wilczek, Phys. Rev. Lett. **49**, 957 (1982).
[2] B. I. Halperin, Phys. Rev. Lett. **52**, 1583 (1984); **52**, 2390(E) (1984).
[3] R. B. Laughlin, Phys. Rev. Lett. **60**, 2677 (1988).
[4] F. D. M. Haldane, Phys. Rev. Lett. **66**, 1529 (1991).
[5] R. B. Laughlin, Phys. Rev. Lett. **50**, 1359 (1983).
[6] F. D. M. Haldane and Y.-S. Wu, Phys. Rev. Lett. **55**, 2887 (1985).
[7] F. D. M. Haldane, Phys. Rev. Lett. **51**, 605 (1983).
[8] P. W. Anderson, Science **235**, 1196 (1987).
[9] S. A. Kivelson, D. S. Rokhsar, and J. R. Sethna, Phys. Rev. B **35**, 8865 (1987).
[10] This applies to the ISE model of Refs. [13,14], but not to a dimerized valence bond state, nor to the chiral 2D state of Ref. [15] which (with periodic boundary conditions) is doubly degenerate in *both* chiral sectors.
[11] I. Affleck, Nucl. Phys. **B265 [FS15]**, 409 (1986).
[12] H. A. Bethe, Z. Phys. **71**, 205 (1931).
[13] F. D. M. Haldane, Phys. Rev. Lett. **60**, 635 (1988).
[14] B. S. Shastry, Phys. Rev. Lett. **60**, 639 (1988).
[15] V. Kalmeyer and R. B. Laughlin, Phys. Rev. Lett. **59**, 2095 (1987).
[16] *Note added.*—Since submission of this Letter, I have found that this construction also gives the exact ground state of the SU(n) generalization of the ISE model [17].
[17] F. D. M. Haldane (unpublished).
[18] G. S. Canwright, S. M. Girvin, and A. Brass, Phys. Rev. Lett. **63**, 2291 (1989).
[19] P. C. E. Stamp, Europhys. Lett. **14**, 569 (1991).

MIT-CTP/3997

New Kinds of Quantum Statistics

Frank Wilczek
Dept. of Physics
Massachusetts Institute of Technology
*Cambridge, MA 02139, USA**

I review the quantum kinematics of identical particles, which suggests new possibilities, beyond bosons and fermions, in 2+1 dimensions; and how simple flux-charge constructions embody the new possibilities, leading to both abelian and nonabelian anyons. I briefly allude to experimental realizations, and also advertise a spinor construction of nonabelian statistics, that has a 3+1 dimensional extension.

In quantum theory the notion of identity reaches a new level of precision and has profound dynamical significance. It becomes important that two particles can be *precisely* identical, i.e. indistinguishable, as opposed to merely similar. When passing from a classical description of indistinguishable particles to a quantum description one must supply additional rules, known as the quantum statistics of the particles.

For many years it was thought that there are only two possibilities for quantum statistics: bosons and fermions. But in 1977 Leinaas and Myrheim [1] demonstrated, at the level of particle quantum mechanics, that there were additional theoretically consistent possibilities. I'm told there were hints of this in earlier work in axiomatic field theory [2]. The work of Leinaas and Myrheim received little attention, and their insight was rediscovered, independently, by Goldin, Menikoff, and Sharp [3], who realized it in the context of a special formulation of quantum mechanics using currents and densities; and by me using conventional quantum field theory [4] (involving solitons, as below, and/or Chern-Simons terms [5]). This circle of ideas came to life as physics in 1984, when Arovas, Schrieffer and I demonstrated [6] – theoretically, but I think quite convincingly – that quasiparticles in the fractional quantum Hall effect obey forms of the new, "anyon" quantum statistics. (That possibility was foreseen by Halperin [7].). The anyonic behavior of quasiparticles (and quasiholes) in the fractional quantum Hall effect is so closely integrated into the overall theory of those states that it can be subtle to demonstrate as an independent phenomenon. A recent series of impressive experiments by V. Goldman and his collaborators [8] have been interpreted this way, and other experiments, requiring less interpretation, are in the works.

Rich mathematical possibilities arise when we consider *nonabelian* statistics. In the abelian case the operations characteristic of quantum statistics – roughly speaking: slow, distant exchange of particle positions – are implemented as multiplications of the wave-function by a complex number (phase). In the nonabelian case complex motions in large Hilbert spaces of degenerate states can come into play. The possibility of exploiting a robust mapping from operations in physical space (characterized topologically) to navigate through large Hilbert spaces has inspired visions of a possible route to quantum computing, known as topological quantum computing. Physical realization of topological quantum computing is still far off, if it can be achieved at all, but the program has inspired impressive work, both theoretical and experimental. An upcoming milestone may be demonstration of a proposal by Moore and Read [9] that quasiparticles in an observed $\nu = \frac{5}{2}$ quantum Hall state obey nonabelian statistics. Experimental programs to test this are well advanced, as well.

Here I will describe a few of the most fundamental concepts underlying these developments in what might appear, to a quantum field theorist, as their simplest natural context. (I will mention quantum Hall physics, experimental aspects, and quantum computing, but I will not even begin to do them justice.) In the course of this review a few intriguing new ideas will come up, too.

I. BRAIDS, PERMUTATIONS, AND IN BETWEEN

Traditionally, the world has been divided between bosons (Bose-Einstein statistics) and fermions (Fermi-Dirac statistics). Let's recall what these are, and why they appear to exhaust the possibilities.

If two identical particles start at positions (A, B) and transition to (A', B'), we must consider both $(A, B) \to (A', B')$ and $(A, B) \to (B', A')$ as possible accounts of what has happened. According the rules of quantum mechanics, we must add the amplitudes for these possibilities, with appropriate weights. The rules for the weights encode the dynamics of the particular particles involved, and a large part of what we do in fundamental physics is to determine such rules

*Electronic address: wilczek@mit.edu

and derive their consequences.

In general, discovering the rules involves creative guesswork, guided by experiment. One important guiding principle is correspondence with classical mechanics. If we have a classical Lagrangian $L_{\rm cl.}$, we can use it, following Feynman, to construct a path integral, with each path weighted by a factor

$$e^{i\int dt L_{\rm cl.}} \equiv e^{iS_{\rm cl.}} \tag{1}$$

where $S_{\rm cl.}$ is the classical action. This path integral provides – modulo several technicalities and qualifications – amplitudes that automatically implement the general rules of quantum mechanics. Specifically: it sums over alternative histories, takes products of amplitudes for successive events, and generates unitary time evolution.

The classical correspondence, however, does not instruct us regarding the relative weights for trajectories that are topologically distinct, i.e. that cannot be continuously deformed into one another. Since only small variations in trajectories are involved in determining the classical equations of motion, from the condition that $S_{\rm cl.}$ is stationary, the classical equations cannot tell us how to interpolate between topologically distinct trajectories. We need additional, essentially quantum-mechanical rules for that.

Now trajectories that transition $(A, B) \to (A', B')$ respectively $(A, B) \to (B', A')$ are obviously topologically distinct. The traditional additional rule is: for bosons, add the amplitudes for these two classes of trajectories[1]; for fermions, subtract.

These might appear to be the only two possibilities, according to the following (not-quite-right) argument. Let us focus on the case $A = A', B = B'$. If we run an "exchange" trajectory $(A, B) \to (B, A)$ twice in succession, the doubled trajectory is a direct trajectory. The the square of the factor we assign to the exchange trajectory must be the square of the (trivial) factor 1 we associate to the direct trajectory, i.e. it must be ± 1.

This argument is not conclusive, however, because there can be additional topological distinctions among trajectories, not visible in the mapping between endpoints. This distinction is especially important in 2 spatial dimensions, so let us start there. (I should recall that quantum-mechanical systems at low energy can effectively embody reduced dimensionality, if their dynamics is constrained below an energy gap to exclude excited states whose wave functions have structure in the transverse direction.) The topology of trajectory space is then specified by the *braid group*. Suppose that we have N identical particles. Define the elementary operation σ_j to be the act of taking particle j over particle $j+1$, so that their final positions are interchanged, while leaving the other particles in place. (See Figure 1.) We define products of the elementary operations by performing them sequentially. Then we have the obvious relation

$$\sigma_j \sigma_k \; = \; \sigma_k \sigma_j; \quad |j-k| \geq 2 \tag{2}$$

among operations that involve separate pairs of particles. We also have the less obvious *Yang-Baxter* relation

$$\sigma_j \sigma_{j+1} \sigma_j \; = \; \sigma_{j+1} \sigma_j \sigma_{j+1} \tag{3}$$

which is illustrated in Figure 1. The topologically distinct classes of trajectories are constructed by taking products of σ_js and their inverses, subject only to these relations.

If we add to the relations that define the braid group the additional relations

$$\sigma_j^2 \; = \; 1 \tag{4}$$

then we arrive at the symmetric (permutation) group S_N. In 3 spatial dimensions, there are more ways to untangle trajectories. Indeed, one can always untangle two world-lines by escaping into the transverse direction to avoid potential intersections, so the permutation of endpoints captures all the topology.

Yet in 3 dimensions, famously, rotations through 2π are not topologically trivial. This topological fact underlies the possibility of spin-$\frac{1}{2}$ (projective) representations of the rotation group. In such representations, the action of a 2π rotation is to multiply the wave function by -1. On the other hand, rotations through 4π are topologically trivial. This suggests that for particles with extended structure, that cannot be adequately represented as simple points (e.g., magnetic monopoles, or solitons with extended zero-modes) we should consider relaxing Eqn. (4) to

$$\sigma_j^4 \; = \; 1 \tag{5}$$

since σ_j^2 can be implemented by a 2π rotation moving the particles $(j, j+1)$ around one another, and σ_j^4 by a 4π rotation. The relations Eqn. (5), together with Eqns. (2, 3), define a group intermediate between the braid group and the symmetric group.

---

[1] As determined by the classical correspondence, or other knowledge of the interactions.

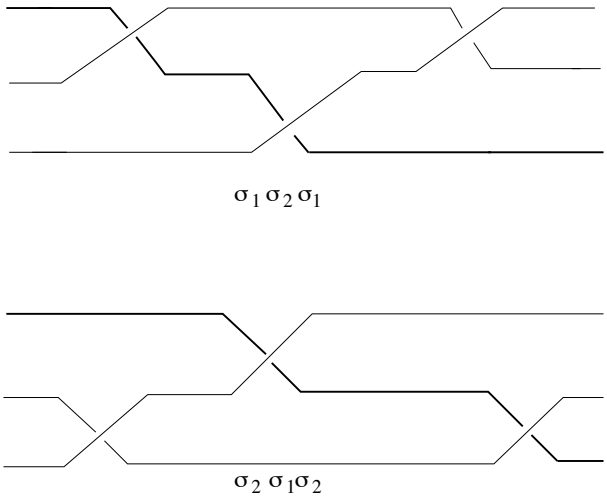


FIG. 1: The elementary acts of crossing one particle trajectory over another generate the braid group. The Yang-Baxter relation $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$, made visible here, is its characteristic constraint.

II. ABELIAN ANYONS

The substitution

$$\sigma_j \rightarrow e^{i\theta} \tag{6}$$

preserves the defining relations of the braid group for any phase factor $e^{i\theta}$, so it generates a unitary representation of the braid group. Thus, at the level of quantum kinematics, it is consistent to weight the amplitudes from topologically distinct classes of trajectories with the corresponding phase factors. (Of course, the additional constraint Eqn. (4) reduces the freedom to $e^{i\theta} = \pm 1$.) This possibility defines the classic, abelian anyons.

There is a simple dynamical realization of anyons, using flux and charge. Consider a $U(1)$ gauge theory that has particles of charge q, associated with a field η and is spontaneously broken by a condensate associated with a field of ρ of charge mq, with m an integer[2] Gauge transformations that multiply η by $e^{2\pi ik/m}$ will multiply ρ by $e^{2\pi ik}$. Thus for integer k they will leave the condensate invariant, but generally act nontrivially on η. We are left with an unbroken gauge group Z_m, the integers modulo m. No conventional long-range gauge interaction survives the symmetry breaking, but there is a topological interaction, as follows:

The theory supports vortices with flux quantized in units of

$$\Phi_0 = \frac{2\pi}{mq} \tag{7}$$

in units with $\hbar \equiv 1$. A particle or group of particles with charge bq moving around a flux Φ will acquire a phase

$$\exp ibq(\oint dt\vec{v}\cdot\vec{A}) = \exp ibq(\oint d\vec{x}\cdot\vec{A}) = e^{i\Phi bq} \tag{8}$$

If the flux is $a\Phi_0$, then the phase will be $e^{2\pi i\frac{ab}{m}}$.

Composites with $(\text{flux}, \text{charge}) = (\text{a}\Phi_0, \text{bq})$ will be generally be anyons: as we implement the interchange σ_j, each charge cluster feels the influence of the others flux. (Note that in two dimensions the familiar flux tubes of three-dimensional physics degenerate to points, so it is proper to regard them as particles.) There are also topological interactions, involving similar accumulations of phase, for non-identical particles. What matters are the quantum numbers, or more formally the superselection sector, not the detailed structure of the particles or excitations involved.

The phase factors that accompany winding have observable consequences. They lead to a characteristic "long range" contribution to the scattering cross-section[3], first computed by Aharonov and Böhm [10] in their classic paper

[2] If m is irrational the gauge group is not compact, i.e. it is the additive group $\mathbf{R}^+$ rather than $U(1)$.
[3] It diverges at small momentum transfer and in the forward direction.

on the significance of the vector potential in quantum mechanics. Unfortunately, that cross-section may not be easy to access experimentally for anyons that occur as excitations in exotic states of condensed matter.

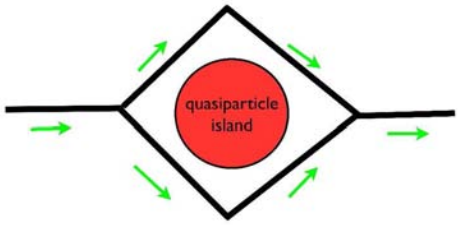


FIG. 2: A schematic interference experiment to reveal quantum statistics. We study how the combined current depends on the occupation of the quasiparticle island.

Interferometry appears more practical. The basic concept is simple and familiar, both from optics and (for instance) from SQUID magnetometers. One divides a coherent flow into two streams, which follow different paths before recombining. The relative phase between the paths determines the form of the interference, which can range from constructive to destructive recombination of the currents. We can vary the superselection sector of the area bounded by the paths, and look for corresponding, characteristic changes in the interference. (See Figure 2.) Though there are many additional refinements, this is the basic concept behind both the Goldman experiments and other planned anyon detection experiments [11].

Elementary excitations in the fractional quantum Hall effect are predicted to be anyons. By far the simplest states to analyze are the original Laughlin $1/m$ states, where the excitations are anyons with $\theta = \pi/m$. There is a rich theory covering more general cases.

III. NONABELIAN ANYONS

The preceding field-theoretic setting for abelian anyons immediately invites nonabelian generalization. We can have a nonabelian gauge theory broken down to a discrete nonabelian subgroup; vortex-charge composites will then exhibit long range, topological interactions of the same kind as we found in the abelian case, for the same reason.

The mathematics and physics of the nonabelian case is considerably more complicated than the abelian case, and includes several qualitatively new effects. First, and most profoundly, we will find ourselves dealing with irreducible *multidimensional* representations of the braiding operations. Thus by winding well-separated particles[4] around one another, in principle arbitrarily slowly, we can not only acquire phase, but even navigate around a multidimensional Hilbert space. For states involving several particles, the size of the Hilbert spaces can get quite large: roughly speaking, they grow exponentially in the number of particles.

As will appear, the states in question are related by locally trivial but globally non-trivial gauge transformations. Thus they should be very nearly degenerate. This situation is reminiscent of what one would have if the particles had an internal of freedom – a spin, say. However the degrees of freedom here are not localized on the particles, but more subtle and globally distributed.

The prospect of having very large Hilbert spaces that we can navigate in a controlled way using topologically defined (and thus forgiving!), gentle operations in physical space, and whose states differ in global properties not easily obscured by local perturbations, has inspired visions of *topological quantum computing*. (Preskill [12] has written an excellent introductory review.) The journey from this vision to the level of engineering practice will be challenging, to say the least, but thankfully there are fascinating prospects along the way.

---

[4] From here on I will refer to the excitations simply as particles, though they may be complex collective excitations in terms of the underlying electrons, or other degrees of freedom.

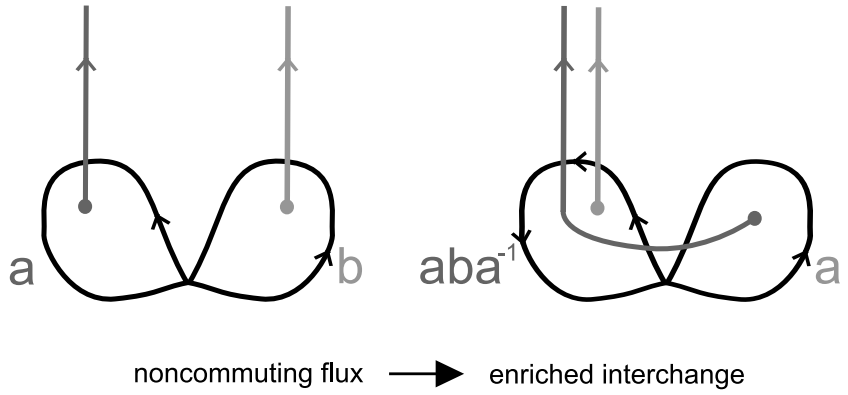


FIG. 3: By a gauge transformation, the vector potential emanating from a flux point can be bundled into a singular line. This aids in visualizing the effects of particle interchanges. Here we see how nonabelian fluxes, as measured by their action on standardized particle trajectories, are modified by particle interchange.

The tiny seed from which all this complexity grows is the phenomenon displayed in Figure 3. To keep track of the topological interactions, it is sufficient to know the total (ordered) line integral of the vector potential around simple circuits issuing from a fixed base point. This will tell us the group element a that will be applied to a charged particle as it traverses that loop. (The value of a generally depends on the base point and on the topology of how the loop winds around the regions where flux is concentrated, but not on other details. More formally, it gives a representation of the fundamental group of the plane with punctures.) If a charge that belongs to the representation R traverses the loop, it will be transformed according to $R(a)$. With these understandings, what Figure 3 makes clear is that when two flux points with flux (a, b) get interchanged by winding the second over the first, the new configuration is characterized as (aba^{-1}, a). Note here that we cannot simply pull the "Dirac strings" where flux is taken off through one another, since nonabelian gauge fields self-interact! So motion of flux tubes in physical space generates non-trivial motion in group space, and thus in the Hilbert space of states with group-theoretic labels.

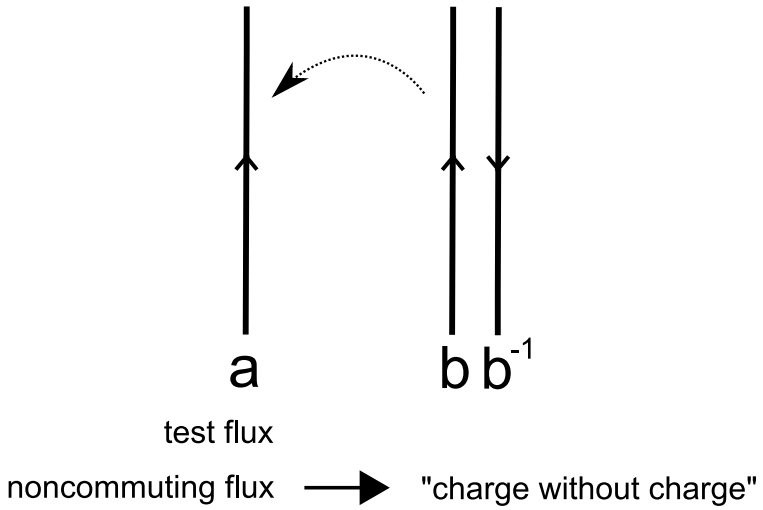


FIG. 4: Winding a flux-antiflux pair around a test flux, and seeing that it gets conjugated, we learn that the pair carries charge.

As a small taste of the interesting things that occur, consider the slightly more complicated situation displayed in Figure 4, with a pair of fluxes (b, b^{-1}) on the right. It's a fun exercise to apply the rule for looping repeatedly, to find out what happens when we take this pair all the way around a on the right. One finds

$$(a, (b, b^{-1})) \rightarrow (a, (aba^{-1}, ab^{-1}a^{-1})) \tag{9}$$

i.e., the pair generally has turned into a different (conjugated) pair. Iterating, we eventually close on a finite-dimensional space of different kinds of pairs. There is a non-trival transformation $\tilde{R}(a)$ in this space that implements

the effect of the flux a on pairs that wind around it. But this property – to be transformed by the group operation – is the defining property of charge! We conclude that flux pairs – flux and inverse flux – act as charges. We have constructed, as John Wheeler might have said, Charge Without Charge.

This flux construction makes it clear that nonabelian statistics is consistent with all the general principles of quantum field theory. Physical realization in condensed matter is a different issue – in that context, nonabelian gauge fields don't come readily to hand. Fortunately, and remarkably, there may be other ways to get there. At least one state of the quantum Hall effect, the so-called Moore-Read state at filling fraction $\frac{5}{2}$, has been identified as a likely candidate to support excitations with nonabelian statistics.

The nonabelian statistics of the Moore-Read state is closely tied up with spinors [13] [14]. I'll give a proper discussion of this, including an extension to 3 + 1 dimensions, elsewhere [15]. Here, I'll just skip to the chase. Taking N γ_j matrices satisfying the usual Clifford algebra relations

$$\{\gamma_j, \gamma_k\} \; = \; 2\delta_{jk} \tag{10}$$

the braiding σ_j are realized as

$$\sigma_j \; = \; e^{i\pi/4}\frac{1}{\sqrt{2}}(1 + \gamma_j\gamma_{j+1}) \tag{11}$$

It's an easy exercise to show that these obey Eqns. (2, 3), and $\sigma_j^4 = 1$ (Eqn. (5)) but not $\sigma_j^2 = 1$ (Eqn. (4)).

---

[1] J. M. Leinaas and J. Myrheim, Nuovo Cimento **B37** (1977), 1 .
[2] R. Doplicher, R. Haag, and J. Roberts, Comm. Math. Phys. **35** (1972), 49.
[3] G. Goldin, R. Menikoff, and D. Sharp J. Math. Phys. **22** (1981), 1664.
[4] F. Wilczek, Phys. Rev. Lett. **48** (1982), 1144; ibid. (1982), 1146; ibid. **49**(1982), 957.
[5] F. Wilczek and A. Zee, Phys. Rev. Lett. **51** (1983), 2250; D. Arovas, J. R. Schrieffer, F. Wilczek, and A Zee Nucl. Phys. **B251** [FS 13] (1985), 917.
[6] D. Arovas, J. R. Schrieffer and F. Wilczek, Phys. Rev. Lett. **53** (1984), 722.
[7] B. Halperin, Phys. Rev. Lett. **52** (1984), 1583.
[8] V. Goldman, J. Liu and A. Zaslavsky, Phys. Rev. **B71** (2005), 153303; F. Camino, F. Zhou, and V. Goldman Phys. Rev. Lett. **98** (2007), 076805.
[9] G. Moore and N. Read, Nucl. Phys. **B360** (1991), 362.
[10] Y. Aharonov and D. Bohm, Phys. Rev. **115** (1959), 485.
[11] See M. Dolev, M. Heiblum, V. Umansky, A. Stern, and D. Mahalu, Nature **452** (2008), 829; I. Radu, J. Miller, C. Marcus, M. Kastner, L. Pfeiffer, and K. West, Science **320** (2008), 899.
[12] J. Preskill, www.theory.caltech.edu/preskill/ph219 (2004-)
[13] C. Nayak and F. Wilczek Nucl. Phys. **B479** (1996), 529.
[14] D. Ivanov, Phys. Rev. Lett. **86** (2001), 268.
[15] F. Wilczek, paper in preparation.

R.W. BATTERMAN

Synthese
DOI 10.1007/s11229-008-9436-1

Idealization and modeling

Robert W. Batterman

Received: 5 December 2007 / Accepted: 17 October 2008

Abstract This paper examines the role of mathematical idealization in describing and explaining various features of the world. It examines two cases: first, briefly, the modeling of shock formation using the idealization of the continuum. Second, and in more detail, the breaking of droplets from the points of view of both analytic fluid mechanics and molecular dynamical simulations at the nano-level. It argues that the continuum idealizations are explanatorily ineliminable and that a full understanding of certain physical phenomena cannot be obtained through completely detailed, non-idealized representations.

Keywords Models · Idealizations · Simulations · Explanation · Fluid dynamics · Scaling · Molecular dynamics

1 Introduction

Physical applied mathematics is in the business of constructing and investigating models of physical phenomena. Typically these mathematical models take the form of an equation or set of equations which are then manipulated in various ways. Fowler (1997) discusses the nature of this art:

> Applied mathematicians have a procedure, almost a philosophy, that they apply when building models. First, there is a phenomenon of interest that one wants to describe or, more importantly, explain. Observations of the phenomenon lead, sometimes after a great deal of effort, to a hypothetical mechanism that can

R. W. Batterman (✉)
Department of Philosophy, University of Western Ontario,
Talbot College London, ON, Canada N6A 3K7
e-mail: rbatterm@uwo.ca

> explain the phenomenon. The purpose of a model is then to formulate a description of the mechanism in quantitative terms, and the analysis of the resulting model leads to results that can be tested against the observations. Ideally, the model also leads to predictions which, if verified, lend authenticity to the model. It is important to realize that all models are idealizations and are limited in their applicability. In fact, one usually aims to oversimplify; the idea is that if a model is basically right, then it can subsequently be made more complicated, but the analysis of it is facilitated by having treated a simpler version first. (Fowler 1997, p. 3)

I think that this is an accurate statement of a fairly widespread view about mathematical modeling and, while I agree with the overall sentiment, I also believe that in several ways it potentially misleads. For instance, I agree with Fowler that all models are or involve idealizations; although I disagree that this necessarily means that they are limited in their applicability. I agree that mathematical modelers usually aim to over-simplify; although I will argue that sometimes (often, in fact) if one tries to make the model more complicated, one fails to realize the stated goal of providing an *explanation* of the phenomenon. Finally (though I will not consider this here), I think that in many instances the search for a mechanism—at least if this is understood rather narrowly in causal terms—is not an important feature of the explanation provided by the mathematical model.

In what follows I would like to discuss these features of mathematical modeling. In particular, I will concentrate on the explanatory goals of modeling. In order to do so we must examine more closely the role of idealization and the proper understanding of that role in describing and explaining various features of the world. However, in order to do this we need to grasp what counts as the physical phenomenon to be modeled. I believe that most discussions of modeling simply take it for granted that we have an appropriate understanding of "the physical phenomenon". But, I think a proper investigation of this concept will help us (at least) to make some distinctions between different views about modeling. Thus, in the next section I try to say something about the nature of the phenomena that are often investigated, and how certain important features of those phenomena demand a particular way of thinking about the role of idealizations in the model—a way that is largely at odds with some of the things Fowler mentions. Following that in Sect. 3 I discuss, qualitatively, an example of the modeling of shocks. In Sects. 4 and 5 I consider in much more detail, first, the analytical modeling of the behavior of breaking droplets and, second, molecular dynamical simulations of the formation of droplets at the nano-level. These two problems are intimately related to one another and serve as good exemplars of the different roles played by idealizations in mathematical modeling. I conclude by arguing that some idealizations are explanatorily ineliminable. That is to say, I argue that the full understanding of certain phenomena cannot be obtained through a completely detailed, nonidealized representation.

2 Idealization and the phenomena

There are (at least) two views about the nature and role idealizations play in modeling and representing physical phenomena. There is what one might call a traditional view,

according to which one aims for the most exact and detailed representation of the phenomenon of interest. On this view, the use of idealizations is, in effect, justified pragmatically: We need to introduce idealizations into our equations in order to simplify them so as to make them tractable or solvable. (As the passage above indicates, Fowler appears to endorse something like this traditional view.) A second view finds virtue where the traditional view sees vice; namely, in the particular kinds of simplification that idealizations typically provide. This other view, which for lack of a better term I will call "nontraditional," maintains that in some cases (and actually in many cases) idealized "overly simple" model equations can *better* explain and characterize the *dominant* features of the physical phenomenon of interest. That is to say, these idealized models better explain than more detailed, less idealized models.

Let us consider the traditional view in a bit more detail. As noted this approach to modeling holds that one should try to find the most accurate and detailed mathematical representation of the problem at hand.[1] This fits nicely with Fowler's "philosophy" of modeling. If the model fails to capture adequately those features of the phenomenon one is interested in, then there are a couple of things one can do. For instance, one can try to add more detail to the mathematical representation,[2] or one might try to adjust the parameters already appearing in the model so as to better reflect what is going on. Most crucially, on this view, the aim is to try to effect a kind of convergence between model and reality. Ultimately, the goal is to arrive at a complete (or true) description of the phenomenon of interest. Thus, on this view, a model is better the more details of the real phenomenon it is actually able to represent mathematically. In effect, idealizations are introduced only to be removed later through further work on those details. This, too, fits nicely with Fowler's "philosophy" of modeling.

Before considering the contrasting approach, we need to get clear about the nature of the so-called "phenomenon of interest." As I noted, I think there is virtually no discussion of this in the literature on modeling and idealization. However, a proper understanding of the kinds of phenomena that are most often of interest will enable us to appreciate better the second, nontraditional, role of idealization in mathematical modeling.

It is an incontrovertible fact that nature presents us with patterns and regularities. And, much of scientific theorizing involves trying to understand how these regularities arise. This is not to say that every pattern we observe reflects a genuine lawful feature of the world. Humans are all too ready to see patterns in just about anything.[3] Neither is it to say that we are interested only in investigating "real" regularities and patterns. *Sui generis* phenomena are, of course, also worthy of investigation. As an example of the latter one might think of studying the nature of the transient behavior in a particular electrical circuit before it settles down to a steady state.

[1] I consider the work of Ronald Laymon as representative of this approach to idealization. See for instance, Laymon (1980).

[2] By this I mean, one might include mathematical representations of additional factors that may be relevant for the phenomenon under investigation.

[3] Fine, in his excellent discussion of computational complexity, randomness, and probability, puts the point as follows: "Too keen an eye for pattern will find it anywhere" (Fine 1973, p. 120).

Nevertheless, most often it seems that our attention is captured by regularities—by repeatable phenomena. It is, in part, the repeatability of phenomena that makes it dominant and captures our interest. That is to say, the repeatability itself is a salient feature that leads us to ask about what is responsible for that very repeatability. When we couple this feature—the salience of the phenomenon—with the fact that for all but the simplest empirical generalizations we need to idealize so as to find an adequate mathematical representation, we gain a fuller understanding of the meaning of "dominant feature."

One goal of mathematical modeling is, surely, to capture these salient features of the regularity in a mathematical formula. The repeatability of the phenomenon places a constraint on the nature of the mathematical model: The model must be sufficiently robust or stable under certain kinds of changes to reflect the fact that the phenomenon is repeatable in various situations where many details have changed. The world is constantly changing in myriads of ways; yet despite this, we see the same patterns over and over again in different situations. Idealizing is a means for focusing on exactly those features that are constitutive of the regularity—those features that we see repeated at different times and in different places. Equivalently, the process of idealization, understood in this way, is most broadly seen as a means for removing details that distract from such a focus—those details that can change without affecting the dominant, repeatable behavior of interest. The mathematical operation that represents the removal of such irrelevant details involves the taking of limits.

Let me now return to the discussion of what I have called the "nontraditional view" of the nature and role of mathematical modeling. Recall that the traditional view aims, ultimately, to "de-idealize" by adding more details so as to bring about a convergence to a complete and accurate descriptions. The nontraditional view, to the contrary, holds that a good model does not let these details get in the way. In many cases the full details will not be needed to characterize the phenomenon of interest, and those details may, in fact, actually detract from an understanding of that phenomenon. This nontraditional approach requires that one find a minimal model—a model "which most economically caricatures the essential physics" (Goldenfeld 1992, p. 33). The adding of details with the goal of "improving" the minimal model is self-defeating—such improvements are illusory.[4]

Once one arrives at a representative equation, there is, to some extent, a set of procedures the modeler typically follows in order to gain insight from the model. (These procedures are largely independent of ones view of the nature of modeling; though, as will become evident, I believe they best fit the nontraditional conception.) In effect, these procedures characterize the modeler's methods of simplification. Two features of this recipe stand out. First, one typically nondimensionalizes the equation or system of equations. This enables one to compare parameters appearing in the equation as to their importance or "size" even though they have been expressed in different units. Second, one takes limits thereby reducing the equation. Typically these limits involve letting a "small" nondimensionalized parameter approach the limiting value of zero or a "large" nondimensionalized parameter is taken to infinity. The aim is to simplify by

[4] See Batterman (2002) for a detailed discussion.

idealizing in this fashion. This is not by any means *solely* an exercise in pragmatics: It is not simply a means for finding exactly solvable solutions. In today's world of extraordinary computing capabilities, this analytical practice continues to play a major role in the investigation of physical phenomena. If all we cared about were correct and accurate numerical predictions, then we would not bother with these analytic investigations. (As Fowler puts it, sounding here as if he endorses the nontraditional conception of modeling, "computation can limit insight, because of an inability to pose questions properly" (Fowler 1997, p. 6).)

The hope is that if done correctly, one will end up with a model which exhibits the *dominant* features of the system. It will be a limiting model that displays the essential physics. As a qualitative example, consider the case of shocks. (A more detailed example is discussed in Sect. 4.)

3 Modeling shocks

Let us say we are interested in understanding the behavior of a gas as it moves through a tube. See Fig. 1. If a collection of the molecules are given a push (say by blowing into the tube at one end), then they will begin to catch up to those in front resulting in a more densely populated region separating two regions of relatively less molecular density. Across this region, molecules will exchange momentum with one another as if some kind of permeable membrane were present. The region occupied by this "membrane" is a shock. Of course it is very difficult to track the behavior of the individual molecules as they move through the tube and undergo the collisions in the shock region. (This is not to say that computational simulations cannot approximately track such behavior. I will have more to say about molecular dynamical simulation and this notion of approximation below.) But, often the applied mathematician will approach the problem by taking a continuum limit. This is a model in which the collection of molecules in the tube is treated as a continuous fluid. Such a limit will shrink the shock region onto a two-dimensional boundary. Upon either side of the boundary, the behavior of the fluid will be governed by the relevant (partial) differential equations of fluid mechanics. However, the behavior across the boundary is not governed by

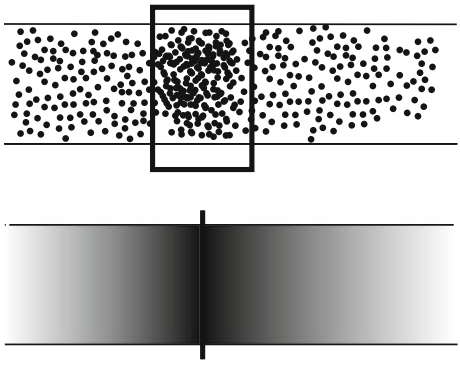


Fig. 1 Modeling shocks

any differential equation at all, but rather by algebraic "jump conditions"—singular behavior across the boundary.

One might think (if you held the more traditional approach to modeling) that the idealization of the collection of molecules to a continuous fluid would be to make the boundary region unimportant to the physics. After all, the boundary shrinks to two dimensions and is not "law governed." (All those ignored molecular details ought to be put back in!) In fact, traditional (covering law) accounts of explanation hold that laws do the essential explanatory work, and initial conditions and boundary conditions are given a sort of secondary status. Further, as the boundary is a place where the laws apparently break down, how can the boundary function in a covering *law* explanation?

Mark Wilson has argued that this view—the view that the boundary becomes unimportant to the physics—is mistaken. In fact, the boundary is the most important feature when it come to understanding the behavior of interest. As Wilson notes "the allegedly 'suppressed details' have become crushed into a singular (hence not law-governed) factor that still dominates the overall behavior through the way in which it constrains the manner in which the 'law governed regions' piece together" Wilson (2003, personal communication).[5] The idea is that such boundaries dominate the physics and that often the mathematical modeler's search focuses on those features to explain what is going on. The limits often yield boundaries that shape or constrain the phenomena. And, it is the elucidation of these shapes that is important for understanding.

Thus, the continuum limit provides a means for ignoring details about molecular interactions in the development of shocks. Most importantly, the taking of limits in this way often imposes mathematical constraints on the equations or formulas that represent the phenomenon of interest. In particular, it requires our models to exhibit the appropriate kind of stability under perturbation of various details—those details that are effectively eliminated by the taking of the limit. Our attempt to represent the dominant features of the phenomenon—genuine features of the world—dictates to some extent the nature of the appropriate mathematical representation. That representation, in turn, leads us to investigate in detail the nature of the imposed constraints. It turns out that in many instances such investigations lead to the discovery of singularities—places where the governing laws "breakdown." The example of shocks is just one such instance. In the next section I consider another example in considerably more detail.

4 Modeling drops and jets

As water drips from a faucet it undergoes a topological change—a single mass of water changes into two or more droplets. This is the most common example of a hydrodynamic discontinuity that arises in a finite period of time. In Victorian times Lord Rayleigh recognized that drops form as a result of a competition between gravitational force and surface tension. He was able to determine the typical size of a droplet and was able to set the time scale upon which a drop would form (Eggers 1997, p. 866).

[5] See Wilson (2006) for much more detailed discussions of these and related issues.

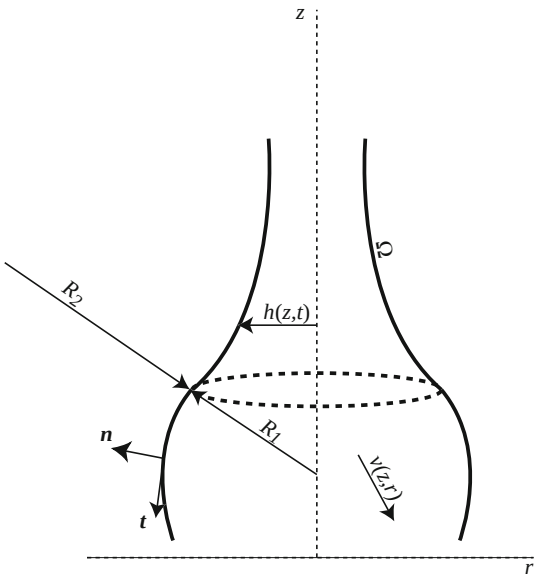


Fig. 2 Geometry of a falling drop

Recent work on the problem has focused on characterizing the shape of the fluid interface at and near the time of breakup. One needs to examine the nonlinear Navier–Stokes equations for free surface flows. These problems are considerably more difficult to solve than those where the fluid is constrained (say by the walls of a pipe).[6] The Navier–Stokes equations must develop a singularity in finite time that is characterized by divergences both in the fluid velocity and in the curvature of the interface at the point of snap-off.

To begin we assume that the typical geometry of a dripping drop is exhibits axial symmetry about the z-axis. Figure 2 provides the relevant details. Assuming axial symmetry, the velocity field inside the fluid is given by a function $v(z, r)$. One can define a time dependent radius function, $h(z, t)$, describing the shape of the drop at any given time. R_1 and R_2 are the principal radii of curvature of the axisymmetric surface Ω. In this geometry, using cylindrical coordinates, the Navier–Stokes equations are given by

$$\frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} + v_z \frac{\partial v_r}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial r} + \nu\left(\frac{\partial^2 v_r}{\partial r^2} + \frac{\partial^2 v_r}{\partial z^2} + \frac{1}{r}\frac{\partial v_r}{\partial r} - \frac{v_r}{r^2}\right), \quad (1)$$

$$\frac{\partial v_z}{\partial t} + v_r \frac{\partial v_z}{\partial r} + v_z \frac{\partial v_z}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial z} + \nu\left(\frac{\partial^2 v_z}{\partial r^2} + \frac{\partial^2 v_z}{\partial z^2} + \frac{1}{r}\frac{\partial v_z}{\partial r}\right) - g, \quad (2)$$

$$\frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z} + \frac{v_r}{r} = 0. \quad (3)$$

[6] In such cases (at least for laminar flows) one can conquer by dividing the problem into two asymptotically related regimes—one near the wall (the boundary layer where viscous effects will dominate), and the other, far from the wall, where such effects are subdominant.

The acceleration due to gravity (g) is in the negative z-direction; v_z and v_r are, respectively, the velocities in the axial and radial directions; p is the pressure; ρ is the fluid density; and ν is the kinematic viscosity. Equation 3 expresses the continuity of the fluid. Equations 1 and 2 express the force balance. The accelerations on the left-hand-sides are due to a pressure gradient (from surrounding the air), viscous stresses, and to gravity (in the z-direction).

These equations are subject to two boundary conditions. The first comes from a balance of normal forces

$$\mathbf{n}\sigma\mathbf{n} = -\gamma\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{4}$$

and the second from a balance of tangential forces

$$\mathbf{n}\sigma\mathbf{t} = 0. \tag{5}$$

Here σ is the stress tensor and γ is the surface tension and Eq. 4, called the "Young-Laplace equation," says that the stress within the fluid normal to the interface and near the surface must be balanced by a stress that acts normal to the surface due to surface tension. The formula "$(1/R_1 + 1/R_2)$" appearing here is equal to twice the mean curvature of the surface Ω at the point of evaluation. Equation 5 expresses the fact that sheer stresses vanishes at the interface. It is possible to express the mean curvature in terms of the radial "shape" function $h(z, t)$.[7] This allows us to write the equation of motion for $h(z, t)$ as follows:

$$\frac{\partial h}{\partial t} + v_z\frac{\partial h}{\partial z} = v_r\,|_{r=h}\,. \tag{6}$$

This says that the surface must move with the fluid at the boundary.

These equations define a difficult and complex moving boundary value problem. We are interested in what happens near the point at which the fluid breaks—at the singularity. Prima facie, that should make the problem even more difficult, as nonlinear effects will dominate. Nevertheless, by focusing on the behavior of the fluid near the singularity, it is possible to simplify the problem dramatically and provide exact solutions to these equations. (This is the modeling recipe mentioned above.) There are two aspects of the problem that allow this to happen.

The first (Eggers 1995, p. 942) derives from the fact that, near breakup, the axial extension of the fluid is much greater than its radial extension. This allows us to make the simplifying assumption that the singularity is line-like. In turn this allows us to find a one-dimensional solution to the full Navier–Stokes equations by introducing a characteristic axial length scale l_z that is related to a radial length scale l_r according to the following scheme:

$$l_r = \epsilon l_z, \tag{7}$$

[7] See Eggers (1995).

where ϵ is a small parameter. If, in addition, we introduce a characteristic time scale t_z we can nondimensionalize the quantities appearing in above equations. The characteristic scales l_z, l_r, and t_z are, of course, constants and so have zero time derivatives. Nevertheless, as the singularity forms, these characteristic scales will be different at different stages of the singularity formation (Eggers 1995, p. 942).

The second feature of the moving boundary problem that allows for simplification is the fact that near the singularity, surface tension, viscous forces, and inertial forces all become equally important (Eggers 1995, p. 942). Surface tension is related to the radius of curvature which diverges at the singularity, viscous forces are also important, and inertial forces must also be considered as the fluid velocity is increasing with greater pressure gradients due to the increasing curvature. Given this, the fluid acceleration diverges leaving the constant acceleration of gravity out of the picture near the singularity.

Furthermore, *and this is extremely important*, close to the singularity, all of the length scales become arbitrarily small in comparison with any external length scale such as the nozzle size of the faucet. This is an indication that one should expect the singular solutions of the one-dimensional Navier–Stokes problem to possess *similarity* or *scaling* properties. To a large extent and for a wide range of fluids, this turns out to be the case.

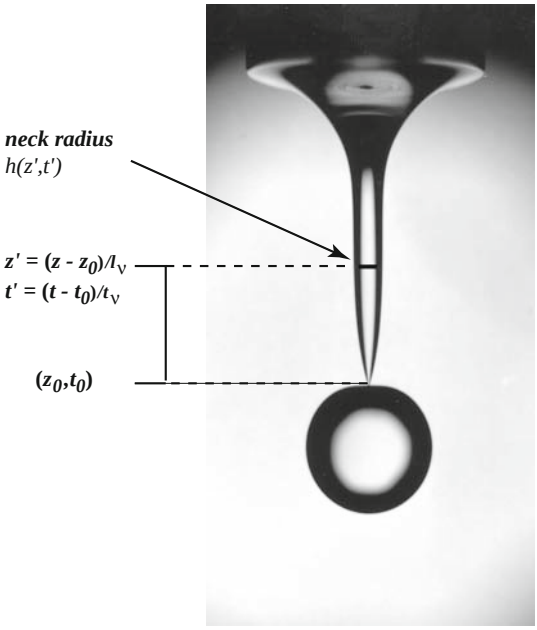
It is worth stressing the importance of discovering a similarity solution to a physical problem. This discovery will mean that one can expect essentially identical behavior in the system when "viewed" at different (appropriately chosen) scales. Such solutions are crucial in standard cases of modeling in which one builds a model, experiments with it, and then argues that the same observed properties will hold at different scales. For instance, consider the investigation of the aerodynamic properties of wings through experimentation on model wings in a wind tunnel.[8] In addition, however, the existence of similarity solutions and their corresponding scaling laws play essential roles in our understanding of why different systems exhibit identical or nearly identical behavior when described in the appropriate (dimensionless) variables. Another way of putting this is to say that the existence of a similarity solution is an indication of a kind of robustness or stability of the phenomenon under perturbation of various details. This, will become clear as the argument below progresses.

Returning to the process of drop formation, recall the following fact. "External" length and time scales that are determined by the initial conditions and the boundary conditions become irrelevant in the description of the singularity. This is critical for our understanding of the nature of the singularity. It means, for example, that it is possible to describe the flow near the breakup using only "internal" length and time scales, defined in terms of the fluid parameters. One introduces the so-called viscous length scale and the viscous time scale as follows:

$$l_v = \frac{\rho \nu^2}{\gamma} \tag{8}$$

[8] An excellent discussion of dimensional analysis, similarity solutions, scaling laws can be found in Barenblatt (2003).

Fig. 3 Water droplet at breakup



$$t_\nu = \frac{\rho^2 \nu^3}{\gamma^3} \tag{9}$$

These scales imply that when the viscosity ν is doubled, the breakup will look the same at length scales four times as large and at time scales eight times as large. This is an instance of scaling.

On the supposition that the breakup occurs at a single point z_0, and at an instant t_0, we can measure spatial and temporal distance from the singularity in terms of the dimensionless variables:

$$z' = \frac{z - z_0}{l_\nu} \tag{10}$$

$$t' = \frac{t - t_0}{t_\nu}. \tag{11}$$

See Fig. 3.[9]

In effect, the scales l_ν and t_ν characterize the width of the critical region around the singularity. For a specific fluid, they are fixed constants and do not change with time as do the characteristic scales mentioned above (l_z, l_r, l_t).

[9] The pictures of water drops in Figs. 3–6 are courtesy of Sidney R. Nagel and appear in Nagel (2001).

It is possible now to demonstrate that a scaling or similarity solution in the variables z', t' exists that describes the drop radius or shape function

$$h(z', t') = |(t')|^{\alpha} \Phi(\xi), \tag{12}$$

where the similarity variable ξ is defined as follows.

$$\xi = \frac{z'}{|t'|^{\beta}}. \tag{13}$$

One can determine the values of the scaling exponents α and β from dimensional analysis. Eggers then shows, both analytically and numerically, that the similarity solution (12) does hold for the problem. One finds the function Φ by inserting the similarity solution into a nondimensionalized version of the fundamental differential Eq. 6.[10] Furthermore, such a solution is in excellent agreement with the full solutions for the (one-dimensional) Navier–Stokes equations at low viscosities.[11]

The existence of such a similarity solution in the variable ξ indicates that the shape of breaking drops is universal. One can see evidence of this by examining the shapes in Figs. 4 and 5.

Notice the cone-to-sphere shape in Fig. 4 and note the *identical* shape at the top of the about-to-break satellite drop in Fig. 5. This demonstrates that how the drop is formed (whether, for instance, it drips solely under the influence of gravity or is sprayed in the air by a crashing wave) is irrelevant for the shape it takes on as it breaks.[12]

In fact, this similarity solution characterizes an entire *class*—a universality class—of fluids at breakup. This class is, in part, determined by the ratio of the viscosity of the fluid to the viscosity of the surrounding medium. For example, the shape of water drops dripping from a faucet surrounded by air (Figs. 4, 5) in which $\nu_{int} \gg \nu_{ext}$ is different than that of a drop forming in a fluid surrounded by another fluid of approximately the same viscosity (Fig. 6) where $\nu_{int} \approx \nu_{ext}$.[13]

That these shapes are to be expected is completely accounted for by the nature of the similarity solution (12) just prior to breakup. Furthermore, Eggers has shown that for scales sufficiently larger than the microscopic, it is actually possible to continue, uniquely, the similarity solution before breakup to one that holds beyond the singularity, after breakup. At breakup some molecular mechanism must come into play, but the uniqueness of this continuation is an indication of the self-consistency of the hydrodynamic description. *The striking conclusion is that the evolution of the fluid both before and after breakup is independent of the molecular microscopic details.*

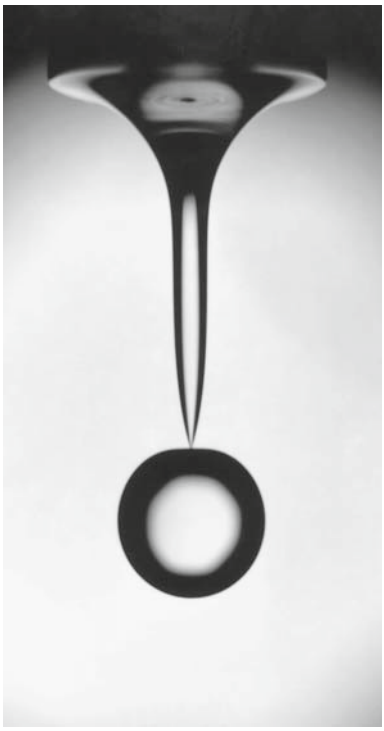
[10] This equation is nondimensionalized using Eqs. 10 and 11.

[11] Shi et al. (1994) argue that Eggers' and Dupont's solution needs to be corrected as there are perturbations (noise) that play an essential role in determining the character of the fluid shape near breakup.

[12] See Nagel (2001).

[13] Interestingly, Doshi et al. (2003) have recently demonstrated a third regime, characterized by $\nu_{int} \ll \nu_{ext}$ that fails to exhibit universal behavior. The breakup profiles in this latter regime are nonuniversal and depend upon initial and boundary conditions in a way that the other two regimes do not.

Fig. 4 Water droplet at breakup



So the existence of the scaling solutions to the one-dimensional Navier–Stokes equations provide evidence for the universality of the phenomenon. And, as a result, it is possible to explain why different fluids, of different viscosities, dripping from different nozzles, etc., will exhibit the same shape upon breakup.

5 Molecular dynamics and simulations

Let me now describe the drop breakup problem from the point of view of state-of-the-art simulations in molecular dynamics. (After all, as just noted, some molecular mechanism must be involved near breakup.) Moseler and Landman (2000) investigate the formation, stability, and breakup of jets at the nanolevel.They model propane as it is injected into a vacuum through a nozzle of diameter 6 nm. The simulation involves following approximately 200,000 propane molecules as they are pushed through a nozzle composed of gold molecules at various pressures. The molecules interact according to the Lennard-Jones 12-6 potential:

$$\phi_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{14}$$

where ϵ and σ are, respectively, energy and length scales appropriate to the materials. The term proportional to $(\frac{1}{r^{12}})$ dominates at short distances and represents the repulsion

Fig. 5 Water droplet after breakup

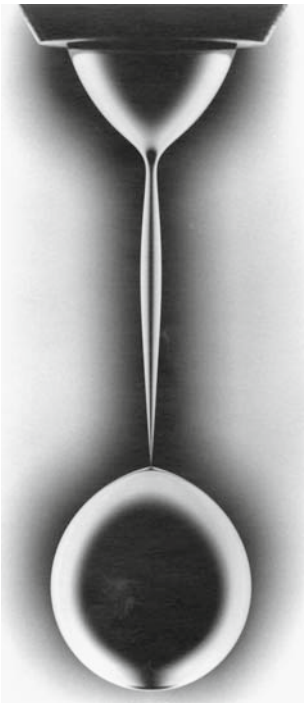
between molecules in very close proximity to one another. The $(\frac{1}{r^6})$ term dominates at large distances and represents the attractive forces between the molecules. Thus the potential has an attractive tail at large r, reaches a minimum near $r = 1.122\sigma$, and is strongly repulsive for $r < \sigma$.[14]

The nanojets in Fig. 7 were simulated by pressurizing the nozzle downstream at 500 MPa and with a controlled temperature at the nozzle of 150 K.[15] This results in a 200 m/s flow velocity for the jet. For $t < 1$ nanosecond following the initial exit of the fluid, the flow exhibits transient behavior. One can see the beginnings of the formation of fast moving droplets and molecular clusters in this initial period, and after that one sees the formation of necking instabilities resulting in breakup and the formation of drops. Moseler and Landman note that for $t \geq 1$ns, a steady state is

[14] The use of the Lennard-Jones potential is justified in investigation of this sort (interactions between closed-shell atoms) for the following reasons. It exhibits long-range van der Waals attraction, extremely strong short-range repulsion and has a potential well. Given these features, along with its relative ease of computational implementation, it is the potential of choice for investigations into generic properties of many molecular dynamical interactions. For a detailed discussion of molecular dynamical simulations see Ercolessi (1997).

[15] Figures 7 and 8 are courtesy of Uzi Landman and appear in Moseler and Landman (2000, p. 1166 and p. 1168, respectively).

Fig. 6 Two fluid breakup: $v_{\text{int}} \approx v_{\text{ext}}$



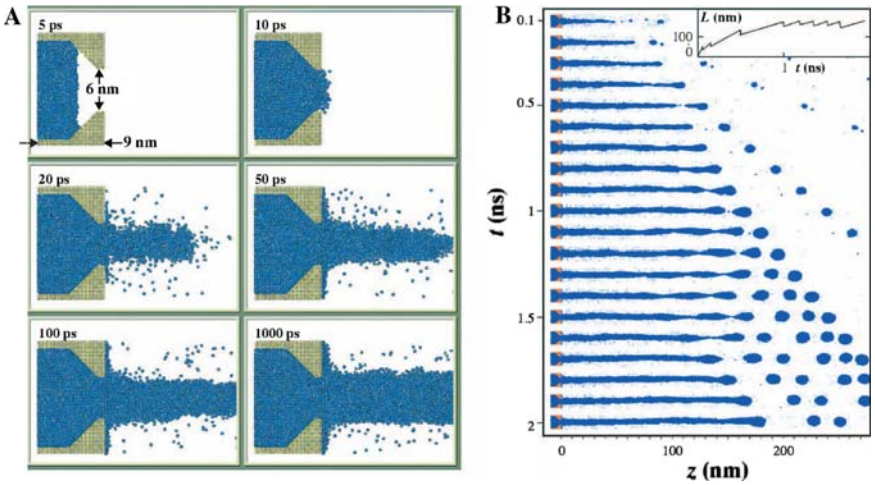


Fig. 7 Formation of nanojets

achieved with an average breakup length of 170 nm. They report that, upon repeated simulations, the typical shape at breakup resembles a double cone as shown in Fig. 8b and the upper image in Fig. 8d. Occasionally, however, they witness the formation of nonaxisymmetric necks as in Fig. 8a and an elongated neck configuration as in Fig. 8c

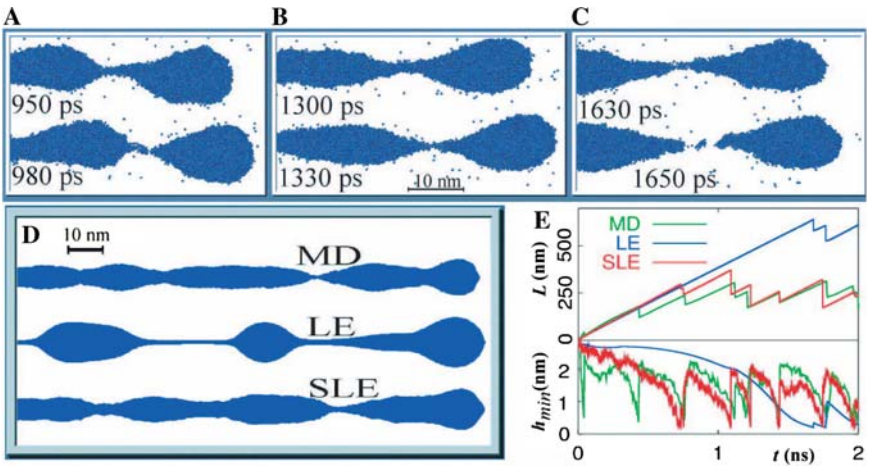


Fig. 8 Molecular dynamical configurations of nanojets

which was accompanied by the formation of small "split-off" molecular clusters or "satellite drops".

If we suppose that the hydrodynamic equations discussed in the last section could apply to the nanoscale drop formation problem, then we would expect the propane (at the nanoscale) to be quite viscous. (Even though, at larger scales, propane is surely not very viscous.)[16] Viscous fluids such as glycerol or honey exhibit long necks prior to breakup. (Just think about the honey that you drip into your cup of tea, or the maple syrup you pour over your pancakes.) In fact, Moseler and Landman apply the hydrodynamic equations (particularly, Eq. 6) and show that as expected for a viscous fluid, the propane jet should develop long necks prior to breakup. This is shown in Fig. 8d and is the simulation labelled "LE" for "lubrication equations."

The discrepancy between the double cone shape of the the molecular dynamical simulation and the hydrodynamic description of the same process is a direct indication that continuum deterministic hydrodynamics fails to apply at the nanoscale. Large hydrodynamic fluctuations become important at the nanolevel signaling a break down of the deterministic continuum description. As Moseler and Landman note,

> …the continuum description of such small systems requires the use of exceedingly small volumes, each containing a very limited number of particles, and consequently, continuum variables associated with such small volume elements, which represent (local) averages over properties of the microscopic constituents are expected to exhibit large fluctuations. (Moseler and Landman 2000, p. 1168)

[16] The reason for this depends upon the scale of observation. For "macroscopic" observation, the scale (l_{obs}) is on the order of one micron (10^{-6} m), and at this level of observation the ratio $l_{\text{obs}}/l_{\nu} \gg 1$. This ratio is what we expect for low viscosity fluids such as water that yield the asymmetric cone-to-cap shape at breakup. However, at the nanolevel—at the level of molecular dynamics—l_{obs} is on the order of a few *nanometers* (10^{-9} m). At this level, $l_{\text{obs}}/l_{\nu} \ll 1$. This ratio holds of viscous fluids such as glycerol and leads to an expectation of thin neck formation prior to breakup (Moseler and Landman 2000, p. 1167).

Moseler and Landman introduce a stochastic term (Gaussian noise) into the hydrodynamic equations and solve the stochastic continuum equations. They demonstrate remarkable agreement with the dominant double cone shape of the molecular dynamical simulations. This agreement is displayed in Fig. 8d. Compare the top molecular dynamical run with the stochastic continuum equations labelled SLE. This agreement "strongly suggests that in [nanojets] the very nature of the dynamical evolution is influenced strongly by hydrodynamic fluctuations, deviating in a substantial way from the behavior predicted through the analysis of the deterministic [continuum equations]" (Moseler and Landman 2000, p. 1168). Further analysis shows that it is possible to see the failure of the deterministic continuum equations as a consequence of a new length scale becoming important at the nanolevel. Moseler and Landman introduce this so-called "thermal capillary length" that for most materials is on the order of interatomic distances.

The fact that a new length scale becomes important at the nanolevel is, according to Moseler and Landman, further indication that the universality described above (provided by the scaling solutions to the Navier–Stokes equations) breaks down. As they say,

> The appearance of an additional length scale in the [stochastic continuum] simulations …is a direct consequence of the extension to include temperature-dependent stress fluctuations, and its magnitude determines the nature of the jet evolution, including the appearance of *solutions other than the universal ones predicted through the deterministic [continuum equations]*. (Moseler and Landman 2000, p. 1168, My emphasis.)

Let me make a few observations and pose a couple of questions concerning the molecular dynamical simulations and their potential for providing explanations for certain aspects of very small-scale drop phenomena. First of all, notice that every molecular dynamical simulation of nanojet formation is different.[17] The images in Fig. 8a–c attest to this. While Moseler and Landman assert that "[t]he most frequently observed breakup process [exhibits] close to pinch-off formation of an axisymmetric double cone shape of the neck …," this amounts to a statistical claim based solely upon generalizations from different simulation runs (Moseler and Landman 2000, p. 1168). And, while it is sometimes appropriate to say that the explanatory buck must stop somewhere, one might, in this situation, ask for an explanation of why *this* is the statistically dominant shape for nanojet breakup.

As we have seen, one important virtue of the scaling solutions to the Navier–Stokes equations discussed in Sect. 4 is that they allow for exactly such an answer to the analogous explanatory why-question on larger scales. We can explain and understand (for large scales) why a given drop shape at breakup occurs and why it is to be expected. The answer depends essentially upon an appeal to the existence of a genuine singularity developing in the equations of motion in a finite time. It is because of this

[17] One might think that this is merely an artifact of simulation and that it counts against treating the molecular dynamical simulations as genuinely providing *theoretical* information about the formation of nanojets. This would be a mistake. The differences in simulations can be attributed to difference in initial conditions, and, as a result, are to be expected.

singularity that there is a decoupling of the breakup behavior (characterized by the scaling solution) from the larger length scales such as those of the faucet diameter. Without a singularity, there is no scaling or similarity solution. *Thus, the virtue of the hydrodynamic singularity is that it allows for the explanation of such universal behavior*. The very break-down of the continuum equations enables us to provide an explanation of universality. This is completely analogous to the renormalization group explanation of the universality of critical phenomena.[18]

No such explanation—one that appeals to a singularity to explain the statistically universal double cone structure, is available from the "fundamental" theory employed in the molecular dynamical simulations. If one looks, for example, at any of the results presented in Fig. 8a, b, or c, one cannot locate the actual breakup location in either time or space. There is no well-defined singularity in the equations. And, of course, one would not expect there to be, since the Newtonian molecular dynamical equations do not develop singularities in finite times.

6 Analytical modeling versus simulation: a reconciliation?

So the question is whether it is possible to provide some kind of theoretical answer to the question of why the double cone structure is to be expected in nanojet breakup. Moseler and Landman show that if one introduces fluctuations into the continuum hydrodynamic equations, and solves those equations, the shape is similar to that typical of many molecular dynamical simulations. But the challenge is to understand the qualitative change in the breakup shape that occurs in the regime in which fluctuations apparently make a leading contribution to the shape function. To put this another way, we would like to have an account of the *statistical universality* of the double cone structure—one that provides the kind of understanding that the scaling solutions provide for the breakup profile at larger scales by demonstrating that most of the details of the evolution are by and large irrelevant.

In a paper entitled "Dynamics of Liquid Nanojets" Eggers (2002) provides the desired explanation. Eggers notes that Moseler's and Landman's stochastic continuum equations suggests that "hydrodynamics, at least when suitably generalized to include fluctuations, is fully capable of describing free surface flows down to the scale of nanometers" (Eggers 2002, p. 084502-1). There is a simple physical argument to understand what goes on at the nanolevel. One can think of the random noise introduced into the continuum equations as representing a kind of effective force that is generated by the fluctuations.

> [A] random fluctuation which increases the thread radius also increases its effective mass, slowing down the motion. Any fluctuation towards a smaller neck radius, on the other hand, accelerates the motion. On average, the fluctuations thus drive the thread towards breakup, in fact more effectively than surface tension …(Eggers 2002, p. 084502-2)

[18] See Batterman (2005) for a discussion.

As Eggers notes, however, conventional perturbative analysis around the deterministic continuum solution cannot describe this mechanism. This is because the fluctuations—the noise—makes the dominant contribution. The idea that one can average about a fixed time

> no longer makes sense for this problem, because there is a finite probability for pinchoff to have occurred, so the original formulation ceases to be valid. Thus a valid description has to be conditioned on the event of breakup to take place at a fixed time t_0. It is then natural to ask for the *most probable* sequence of profiles that brings one from the initial condition to a "typical" breakup event. (Eggers 2002, p. 084502-2)

Eggers develops an ingenious and difficult argument involving path integrals to determine probability of the "optimal" path to breakup. For our purposes here, the interesting feature is that to solve this problem he needs to *assume*, for a fixed breakup time t_0, that the solution is *self-similar*. He finds that the unique solution, on this assumption, is the symmetric profile of a double cone unlike the asymmetric long-neck similarity profile for the deterministic equations. The crucial feature is that the similarity solution is only possible on the assumption that there is a *singularity* at t_0 in the (stochastic) hydrodynamical equations. The result is an explanation for why such a symmetric profile seen in the *molecular dynamical simulations* is to be expected—one that is grounded in the "less fundamental" continuum theory of hydrodynamics.

A further consequence of this explanation is that we can understand why so few satellite drops are formed in nanojets and why there is a very narrow distribution in the size of the droplets that are formed. If one looks back at Fig. 5, one sees that a satellite drop is about to detach itself from the nozzle at the upper end of the picture. This is a consequence of the asymmetric, long-neck nature of the dripping process. That smaller satellite molecular clusters, such as that in Fig. 8c, are unlikely to form is a direct consequence of the universality of the double cone profile for nanojets. They occur only for large fluctuations in the neck region; and such fluctuations are statistically rare.

So, surface tension driven pinching at larger scales essentially determines the breakup time. Nevertheless, at times very close to that, a different process dominated by fluctuations takes over, speeding up the breakup at the nanoscale. The transition between these different scaling regimes can be understood in terms of the emergence, as one approaches the nanolevel, of a new length scale—the thermal capillary length. Most importantly, however, our understanding of this transition and of the universality of the different profiles, depends essentially upon the development of finite time singularities in the continuum hydrodynamical equations. These singularities entail that the breakup behavior at small length and time scales decouples from larger length and time scales. The details of the molecular dynamics drops out of our explanation of the origin of the different universality classes. Thus, the very breakdown of the continuum equations enables us to provide an explanation of the universal shapes.

However, from the point of view of pure molecular dynamical simulation, we can have no explanation of the universal shape of breaking drops and jets. The molecular dynamical equations do not exhibit any singularities—there are no blow-ups allowing for the scaling solutions that is required for this sort of understanding.

7 Conclusion

Fowler's characterization of the nature of idealization in mathematical modeling conforms well with what I have called the traditional conception. Recall that from that point of view, idealizations are pragmatically justified and (paradoxically) receive their ultimate warrant from the "fact" that they are to be (can in principle be) eliminated by further work on the details. In the context of our discussion of the nanojet simulations, one can think of the simulations as attempts to provide all of those details—to fully de-idealized a continuum description by tracking all of the molecular motions. Such simulations do surely provide significant and interesting information about the nature of those dynamical systems.

However, one lesson to be learned from this discussion is that, sometimes at least, such simulations do not tell us the whole story. The understanding of the process that they provide is only partial. They cannot, I have argued, provide an explanation for the universality of the shapes that appear in the jets at breakup. The gaps in the full story *can*, as I have tried to show, be filled in by employing (limiting) idealizations—idealizations that are ubiquitous in the mathematical analysts' approach to modeling. In particular, by appealing to the idealized continuum theory of hydrodynamics. Furthermore, it seems that these idealizations are in many instances explanatorily *ineliminable*. That is to say, they play an essential role in the proper explanation of the phenomenon of interest. They are not, as the traditional view of the use of idealization in modeling suggests, put in only to be subsequently removed by more detailed work.

Acknowledgements I would like to thank Roman Frigg, Stephan Hartmann, and Cyrille Imbert for the stimulating and informative *Models and Simulations* conference at which an earlier version of this paper was presented.

References

Barenblatt, G. I. (2003). *Scaling*. Cambridge: Cambridge University Press.

Batterman, R. W. (2002). Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science, 53*, 21–38.

Batterman, R. W. (2005). Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics, 36*, 225–244.

Doshi, P., Cohen, I., Zhang, W. W., Siegel, M., Howell, P., Basaran, O. A., & Nagel, S. R. (2003). Persistence of memory in drop breakup: The breakdown of universality. *Science, 2005*, 1185–1188.

Eggers, J. (1995). Theory of drop formation. *Physics of Fluids, 7*(5), 941–953.

Eggers, J. (1997). Nonlinear dynamics and breakup of free-surface flows. *Reviews of Modern Physics, 69*(3), 865–929.

Eggers, J. (2002). Dynamics of liquid nanojets. *Physical Review Letters, 89*(8), 084502–1–084502–4.

Ercolessi, F. (1997). *A molecular dynamics primer*. Trieste, Italy: Spring College in Computational Physics, ICTP/International School for Advanced Studies (SISSA-ISAS). Available at http://www.fisica.uniud.it/~ercolessi/md/.

Fine, T. L. (1973). *Theories of probability: An examination of foundations*. New York: Academic Press.

Fowler, A. C. (1997). *Mathematical models in the applied sciences* (Cambridge Texts in Applied Mathematics). Cambridge: Cambridge University Press.

Goldenfeld, N. (1992). *Lectures on phase transitions and the renormalization group*. Frontiers in physics (Vol. 85). Reading, Massachusetts: Addison-Wesley.

Laymon, R. (1980). Idealization, explanation, and confirmation. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Contributed Papers* (Vol. 1, pp. 336–350).

Moseler, M. & Landman, U. (2000). Formation, stability, and breakup of nanojets. *Science, 289*, 1165–1169.
Nagel, S. R. (2001). Shadows and ephemera. *Critical Inquiry, 28*(1), 23–39.
Shi, X. D., Brenner, M. P. & Nagel, S. R. (1994). A cascade structure in a drop falling from a faucet. *Science, 265*(5169), 219–222.
Wilson, M. (2006). *Wandering significance*. England: Oxford University Press.

S. HARTMANN

Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 32, No. 2, pp. 267–304, 2001

Printed in Great Britain
1355-2198/01/$ - see front matter

Effective Field Theories, Reductionism and Scientific Explanation

*Stephan Hartmann**

Effective field theories have been a very popular tool in quantum physics for almost two decades. And there are good reasons for this. I will argue that effective field theories share many of the advantages of both fundamental theories and phenomenological models, while avoiding their respective shortcomings. They are, for example, flexible enough to cover a wide range of phenomena, and concrete enough to provide a detailed story of the specific mechanisms at work at a given energy scale. So will all of physics eventually converge on effective field theories? This paper argues that good scientific research can be characterised by a fruitful interaction between fundamental theories, phenomenological models and effective field theories. All of them have their appropriate functions in the research process, and all of them are indispensable. They complement each other and hang together in a coherent way which I shall characterise in some detail. To illustrate all this I will present a case study from nuclear and particle physics. The resulting view about scientific theorising is inherently pluralistic, and has implications for the debates about reductionism and scientific explanation.

Keywords: Effective Field Theory; Quantum Field Theory; Renormalisation; Reductionism; Explanation; Pluralism.

1. Introduction

There is little doubt that effective field theories are nowadays a very popular tool in quantum physics. They are almost everywhere, and everything is

*Center for Philosophy of Science, University of Pittsburgh, 817 Cathedral of Learning, Pittsburgh, PA 15260, U.S.A. (*e-mail:* shart@pitt.edu) (correspondence address); and Sektion Physik, Universität München, Theresienstr. 37, 80333 München, Germany.

PII: S1355-2198(01)00005-3

considered to be an effective field theory (EFT). Particle physicists, for example, even take a supposed-to-be fundamental theory such as the celebrated Standard Model of the electromagnetic, weak, and strong interactions to be an EFT (Meissner, 1992). Nuclear physicists systematically derive low-energy EFTs from quantum chromodynamics (a theory which is part of the Standard Model) to account for the dynamics of protons and neutrons in atomic nuclei at low and intermediate energies (Van Kolck, 1999). And solid state theorists formulate age-old models such as the BCS theory of conventional superconductivity in the language of EFTs (Shankar, 1999). Even gravitational physicists seem to be infected by the EFT-virus: they consider the general theory of relativity to be the starting point of a power-series expansion, to which higher-order quantities that are still invariant under general coordinate transformations have to be added, to account for the physics at higher energies (Donoghue, 1994a,b). The resulting EFTs include quantum corrections to Einstein's theory which are considered to be footprints of a quantum theory of gravity, a theory we do not yet have, but which we might be able to find (or divine) by following the EFT programme.

EFTs account for the physics at a given energy scale by relying only on those entities which are relevant at that scale. These entities are, for example, quarks, leptons and the gauge bosons in the Standard Model, pions and nucleons in nuclear physics at not too high energies, and Cooper pairs in the theory of conventional superconductors. Using these effective degrees of freedom makes computations tractable and provides some intuitive understanding of what is physically going on at the energy scale under consideration. The resulting descriptions are very accurate. This indicates that the effects of the physics at higher energies do not really make a difference at lower scales in these cases: the physics at high energies is 'decoupled' from the physics at low energies. Its effects are contained in a few parameters of the low energy theory. The formalism of EFTs makes all this more precise. Besides, there is a systematic and controlled way to derive low energy EFTs from a more fundamental high energy theory.

For a long time, the criterion of renormalisability was considered to be a *sine qua non* for any acceptable physical theory. After all, we want our theories to give finite results and if higher orders in a perturbation expansion diverge, the theory is in trouble. Renormalisation is a way to 'get rid' of these infinities, but it turns out that many EFTs cannot be renormalised and are therefore, according to the old view, in trouble. Their appraisal requires that we reconceptualise what renormalisation amounts to. This reconceptualisation took place in the 1970s; it is a consequence of a realistic (as opposed to a formalistic) interpretation of the cut-off parameter in quantum field theories and of the insights of renormalisation group theory.

Besides their value in research, EFTs also played a role in a recent debate among scientists which was, however, in the end mainly about funding issues. In this debate, particle physicists (most prominently Steven Weinberg) advocated building a Superconducting Super Collider (SSC), an extraordina-

rily expensive particle accelerator, which should help theorists find the 'final theory' (Weinberg, 1993). In the end, the US Congress did not make this dream come true. Weinberg's opponents, such as the solid state physicists Philip W. Anderson and James Krumhansl, argued convincingly against this project. Since he could not point to technological spin-offs, Weinberg's main argument for the SSC was the very foundational character of particle physics: 'Particle physics is in some sense more fundamental than other areas of physics' (Weinberg, 1987, p. 434). It is more fundamental because it is 'on a level closer to the source of the arrows of explanation than other areas of physics' (*ibid.*, p. 437). Anti-reductionists, on the other hand, point to the autonomy of the different levels of organisation. All these levels have their own ontology and their own laws, so why not call them fundamental as well?[1] It is not an easy task to make more precise what it means exactly that different levels of organisation are autonomous. However, within the programme of EFTs, the notion of quasi-autonomy can be given a precise meaning and the relation of one level of organisation to a deeper level can be studied. We will come back to this issue below and discuss its consequences for the reductionism debate.

Despite the great importance of EFTs in actual scientific practice and in an important debate among scientists, philosophers of science have not paid much attention to EFTs. Following a seminal (though philosophically controversial) paper by Cao and Schweber (1993), some articles have been published which mainly focus on the issue of renormalisation and on the role of the renormalisation group.[2] In 1996, a remarkable conference on the conceptual foundations of quantum field theory took place at Boston University. Its participants included many of the main contributors to the development of quantum field theory and to the EFT programme.[3] At this conference a lot of attention was paid to EFTs. A full philosophical appraisal of EFTs and their consequences is still missing however. This is the aim of this article.

Philosophers of science have discussed theories and models a great deal. EFTs share similarities with both of them. My first goal will therefore be to locate EFTs in the 'conceptual space' defined by these tools. I will do this by looking at the *functions* of theories, models, and EFTs in the research process and conclude that EFTs share many of the functions of theories and models. Theories and models are, however, also an indispensible tool of scientific research and I will defend a pluralistic account of scientific theorising on the basis of a detailed case study. My second goal is then to draw some more general conclusions from my reconstruction of scientific practice, namely about the issues of reductionism and scientific explanation.

The remainder of this article is organised as follows. Section 2 provides some historical background and introduces the concept of EFTs. Section 3 points

[1] This debate is carefully reconstructed in Cat (1998).

[2] See the articles by Huggett and Weingard (1995), Robinson (1992), Cao (1993), Schweber (1993a), and the more general articles by Schweber (1993b, 1995).

[3] The proceedings of this conference are published in Cao (1999).

out the functions of, and relations between, theories, models, and EFTs on the basis of a case study from nuclear and particle physics. Philosophical conclusions concerning pluralism, reductionism and scientific explanation are then drawn in Section 4. Finally, Section 5 summarises my main points.

2. The Emergence of Effective Field Theories

Although the first paper on EFTs appeared only in 1979 (Weinberg, 1979, 1980b), the general idea behind it is much older. As early as in 1936, the German physicists Hans Euler and Werner Heisenberg calculated the process of photon-photon scattering at small photon energies within the framework of the quantum theory of fields developed by Paul Dirac a couple of years earlier. Euler and Heisenberg derived a non-linear modification of Maxwell's equations which could however be interpreted in an intuitive way. Another early example of an EFT is Fermi's theory of weak interactions. Both theories will be discussed in Section 2.1. For a long time, however, theories such as the ones by Euler, Heisenberg and Fermi were not taken seriously because they were not renormalisable. Only after a 'change in attitude' (Weinberg) among physicists—mainly due to the development of renormalisation group techniques—was it possible to consider non-renormalisable theories as full-blown scientific achievements. To arrive at the current conception of EFTs, one more step was required. In 1975, Appelquist and Carazzone derived a theorem (Appelquist and Carazzone, 1975) according to which under certain conditions the heavy particles in a theory decouple from the low-energy physics (modulo a renormalisation of the parameters of that theory). I will sketch these developments in Section 2.2. Finally, I present two ways of applying EFTs, viz. the bottom-up approach and the top-down approach, both of which have a variety of applications in physics.

2.1. Two early examples

This section introduces two early examples of an EFT, the Euler–Heisenberg theory of photon-photon scattering (Section 2.1.1) and the Fermi theory of weak interactions (Section 2.1.2). Both theories exhibit typical features of an EFT which are compiled in Section 2.1.3.

2.1.1. The Euler–Heisenberg theory

Soon after Dirac presented his first attempts towards a quantum theory of fields, Euler and Heisenberg applied this theory to the process of photon-photon scattering.[4] The authors did not worry much about the fact that Dirac's theory had various conceptual problems at that time. Quite to the contrary, by

[4] The first paper on this subject matter was published by Euler (1936); Heisenberg and Euler (1936) contains a considerable simplification and generalisation of Euler's calculation.

working out interesting applications of the theory and by exploring its consequences, Euler and Heisenberg hoped to get a hint in which direction one has to look in order to find a satisfactory quantum theory of fields. In a letter to Wolfgang Pauli, Heisenberg wrote about a similar situation a couple of months earlier:

> In respect to quantum electrodynamics, we are now in the same state as we were in 1922 in respect to quantum mechanics. We know that everything is wrong. But in order to find the direction in which we have to depart from the present state, we have to know much better the consequences of the present formalism.[5]

Following this line of thought, the motivation of Euler and Heisenberg's joint work was to get an understanding of the consequences of Dirac's provisional formalism in order to find ways to improve it.

Photon-photon scattering is a typical quantum electrodynamical process which has no classical analogue. It does not occur in classical physics because of the linearity of Maxwell's equations ('superposition principle'). In quantum electrodynamics, however, the superposition principle does not hold. Now, photons can interact and the elementary process, the discovery of which Euler attributes to Otto Halpern and Peter Debye, is this: the two photons scatter and create an electron-positron pair which then decays back into two photons, respecting energy and momentum conservation. This effect will lead to a modification of Maxwell's equations for the vacuum by adding non-linear terms to it.

Euler and Heisenberg did not tackle the full problem but considered a special case. While Breit and Wheeler calculated the cross section for this process for high photon energies, in which real electrons and positrons are created, Euler and Heisenberg's attention focused on photons with energies well below the production threshold of electrons and positrons.

But even this is not an easy problem. In the modern language of Feynman diagrams, the 'box-diagram' depicted in Fig. 1 has to be calculated.[6] Euler and Heisenberg, of course, did not know Feynman's efficient methods, but they calculated essentially this diagram and so we use it here to visualise the corresponding elementary process. Since the process is of the fourth order, it is clear that there are considerable mathematical difficulties which show up when calculating the transition amplitude. These difficulties even show up when Feynman diagrams are used explicitly. [7]

For the details of the calculation we now follow the modern reconstruction given by Itzykson and Zuber (1980, pp. 195f). This work is focused on Euler and Heisenberg (1936), which is a simplification and generalisation of Euler (1936). The modern covariant formulation goes back to Schwinger (1973,

[5] Letter to W. Pauli dated 25 April 1935; quoted from Cassidy (1995, p. 416), my translation.

[6] Besides this diagram, there are two other diagrams which contribute in the same (fourth) order in perturbation theory; they are obtained by permutating the external photon lines, cf. Jauch and Rohrlich (1976, Ch. 13).

[7] For an exact calculation, see Jauch and Rohrlich (1976, Ch. 13).

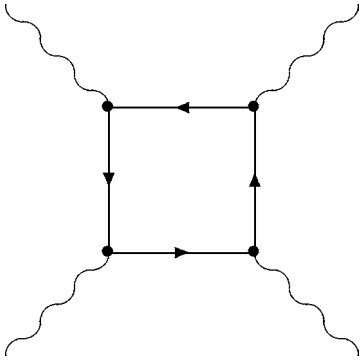


Fig. 1. *Feynman diagram for photon-photon scattering.*

pp. 123–134). If one requires (1) U(1) gauge invariance, (2) Lorentz invariance and (3) parity invariance, any Lagrangian density which should account for the process of photon-photon scattering must have the following structure:

$$\mathscr{L}_{eff} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{a}{m_e^4}(F^{\mu\nu}F_{\mu\nu})^2 + \frac{b}{m_e^4}F^{\mu\nu}F_{\nu\sigma}F^{\sigma\rho}F_{\rho\mu} + \mathcal{O}(F^6/m_e^8). \qquad (1)$$

Here m_e is the mass of the electron, and $F^{\mu\nu}$ is the field-strength tensor of the electromagnetic field. a and b are dimensionless constants which have to be determined.

Note that there are no electron degrees of freedom in Eq. (1). This is not necessary, however, since the considered process is purely photonic. Electrons do not show up explicitly. The first term in Eq. (1) is the well-known contribution of 'free' photons. All other terms are part of a systematic expansion in $1/m_e$, respecting the symmetries mentioned above. The non-linearity of these terms reflects the violation of the superposition principle. For low photon energies ($E_\gamma \ll m_e$), it suffices to consider only the first three terms in this expansion.

All information about this energy regime is therefore contained in the constants a and b. But how can these constants be determined? One possibility is to do the explicit expansion of the original Lagrangian density of quantum electrodynamics by 'integrating' out the electron degrees of freedom. There are efficient calculational tools available now which, however, presuppose the path-integral formulation which was not available when Euler and Heisenberg performed the calculations. Instead, Heisenberg and Euler (1936) applied elegant mathematical technique which essentially led to the same result. However, in the original publication, Euler (1936) chose another way. He calculated a special case of the process under consideration in two ways: exactly and by using the effective Lagrangian density given by $\mathscr{L}_{eff}$. Comparing both results, he obtained the following values for the two

constants:

$$a = -\frac{\alpha_0^2}{36}\,, \qquad b = \frac{7\alpha_0^2}{90}\,, \tag{2}$$

with the fine-structure constant $\alpha_0 = 1/137$. In terms of the electric and magnetic field strengths (E and B), the resulting effective Lagrangian density has the following form:

$$\mathscr{L}_{eff}^{EH} = \frac{1}{2}(E^2 - B^2) + \frac{2\alpha_0^2}{45m_e^4}[(E^2 - B^2)^2 + 7(E \cdot B)^2]. \tag{3}$$

This expression has been the basis of many subsequent calculations; it is still used today (see, for example, Becker, McIver, and Schlicher (1989) for a quantum optical application).

2.1.2. The Fermi theory

Another historical example is Enrico Fermi's theory (1933, 1934) of weak interactions which was developed soon after Wolfgang Pauli suggested the existence of the neutrino as a way to account for the continuous beta spectra discovered by James Chadwick in 1914. These spectra gave rise to various speculations, including Niels Bohr's famous suggestion of giving up energy conservation in order to account for them.

Following the model of quantum electrodynamics, Fermi developed a theory which uses Pauli's hypothesis and describes the elementary process $n \rightarrow p + e^- + \bar{\nu}_e$ quantum field theoretically. In this reaction, a neutron (n) decays in a proton (p), an electron (e^-) and an electron anti-neutrino ($\bar{\nu}_e$).[8] Since there was nothing known about the details of the interaction, Fermi had to start from scratch, with some, but not many experimental constraints. He assumed that the interaction is pointlike and that the interaction Hamiltonian is given by the product of the operators representing the relevant particles multiplied by a coupling constant which has to be derived from experiment. This coupling constant has the dimension $energy^{-2}$.

With these assumptions and the application of perturbation theory, Fermi was able to derive various mean lives of unstable nuclei as well as the shape of the electron spectra. Fermi's theory was highly successful and remained valid until experiments established that parity is violated in weak interactions. In order to account for this, Richard P. Feynman and Murray Gell-Mann (along with Robert E. Marshak and E.C.G. Sudarshan) suggested in 1958 a modification of Fermi's theory, the $V - A$ theory. This empirically very successful theory (apart from an explanation of CP violation) is still based on a point interaction and uses the same coupling constant Fermi used. Like Fermi's theory, the $V - A$ theory is not renormalisable.

[8] Fermi took it to be a neutrino. He did not yet know about the conservation of lepton number and other kinds of neutrinos.

2.1.3. Some conclusions

The theories of Fermi and Euler and Heisenberg have some interesting features. These features are typical for EFTs and can be summarised as follows:

1. Both theories take only the *relevant fields* into account. These fields, called effective fields, are the photon field (represented by the electrical and magnetical field strengths) in the Euler–Heisenberg example, and the proton, neutron, electron and neutrino fields in the Fermi theory of weak interactions. Other fields, such as the electron field in the Euler–Heisenberg case, do not show up explicitly at the respective energy scale. Their presence is hidden, reflected by the non-linear terms in the effective Lagrangian density.
2. Both theories are valid only at a given *energy scale*. The derivation of the Euler–Heisenberg theory presupposes that the photon energy is small compared to the rest mass of the electron. Applications to higher energies are not justified. The Fermi theory violates unitarity at high energies (above 300 GeV) and is therefore also valid only at a specific energy scale. For higher energies, alternative theories are needed.
3. Both theories are *non-renormalisable*. It can be shown on general grounds that the theories by Euler and Heisenberg and by Fermi are non-renormalisable. Divergent results show up once higher order contributions to the perturbation expansion are calculated. To eliminate them, a renormalisation scheme has to be specified. This renormalisation scheme is therefore part of the definition of the EFT if one is interested in higher order contributions.
4. Both theories are based on certain *symmetries*. Symmetry requirements are very important in the construction process of an EFT. This is demonstrated by our reconstruction of the development of the Euler–Heisenberg theory; to get the effective Lagrangian density of Eq. (1), all possible terms with the required symmetries have to be included—whether they are renormalisable or not. The hard job is then to determine the coefficients of the respective terms in the expansion.

   Symmetry considerations also played a role in the formulation of the Fermi theory. Since there was not much information about the structure of the weak interaction, simplicity suggested a scalar interaction term. After the discovery of the violation of parity conservation in the weak interactions, a combination of all other possible types of Lorentz-invariant pointlike interactions were tried. Fortunately, there are only five of them (scalar (S), pseudoscalar (P), vector (V), axialvector (A) and tensor (T) interactions) and a set of crucial experiments finally selected the $V - A$ of Feynman and Gell-Mann as the only one compatible with available experimental data (see Franklin, 1990).
5. Both theories produce *scientific understanding*. The work of Euler and Heisenberg had many motivations. Among those were the wish to apply, to test, and to find out the consequences of Dirac's provisional quantum theory

of fields. Another motive was to get a tractable mathematical formalism which allowed the calculation of the interaction of photons at low energies. This goal suggested the chosen approximation scheme. Yet another motive was to get some intuitive understanding of the respective processes. This is directly substantiated by a section title in Euler's original paper (1936, p. 400; my translation):

> § 1. Provisional statement of an intuitive expression for the interaction $\bar{U}_1$ of light with light [...]
>
> In the course of this work Euler mentions several times that his aim is to derive an intuitive (*anschaulich*) expression which describes the physically relevant processes. What does this mean? It was already well known at the time that certain materials react in a non-linear way to external fields. The guiding idea for Euler and Heisenberg now was that even the vacuum exhibits such non-linear behaviour. Starting from the quantum theory of electrons and photons, they succeeded in deriving non-linear corrections to Maxwell's equations for the vacuum based on this analogy. The corrections suggest the interpretation that even the vacuum can be polarised, an effect which is responsible for the non-vanishing photon-photon cross section. This analogy to an already well-understood effect guided their derivation and helped to interpret the final result. The resulting corrections to Maxwell's equations are also very easy to handle mathematically, once they are derived. It is a typical feature of EFTs that they are very easy to handle (compared to the full theory), and also produce (local) understanding (unlike the full theory, as I will argue below, see Section 3.1.1). Pragmatic and cognitive goals meet here in an interesting way.
>
> The Fermi theory produces understanding in so far as it is the simplest modification of quantum electrodynamics which accounts for the phenomena of weak interactions.

In the years between the development of Euler, Heisenberg and Fermi's theories and the late nineteen-forties, theoretical research in quantum field theory focused mainly on formulating a theory which avoids the divergences in the perturbative expansion from which Dirac's theory suffered. Satisfactory covariant renormalisation schemes were finally introduced by Dyson, Feynman, Schwinger, and Tomonaga.[9] Motivated by the astonishing success of QED, as manifested most convincingly in the precise calculation of the Lamb shift and the anomalous magnetic moment of the electron, renormalisability soon became the key criterion for the selection of quantum field theories for other phenomena (such as the weak and strong interactions). Henceforth, non-renormalisable theories had at best a provisional status: useful, perhaps, for various calculations, but of no deeper significance. This view began to crumble with the development of renormalisation group techniques in the 1970s which finally led to the rehabilitation of non-renormalisable theories and the

[9] See Schweber (1994) for an historical account of these exciting developments.

establishment of the research programme of EFTs. The next section will sketch this development in more detail.

2.2. Renormalisation and the renormalisation group

The modern development of EFTs is closely related to a new conceptualisation of renormalisation. This section will give a concise reconstruction of this development. It all began with quantum electrodynamics (QED) and the supposed need to find a way to eliminate the notorious infinities in the perturbation expansion. Let's first look at this expansion.

Let H_I be the interaction Hamiltonian of a system and let $\Psi(t)$ be a field operator in the interaction picture. $\Psi(t)$ satisfies the Schrödinger equation

$$H_I \Psi(t) = \mathrm{i}\, \frac{\partial \Psi(t)}{\partial t}, \tag{4}$$

which has the formal solution

$$\Psi(t) = \Psi(-\infty) - \mathrm{i} \int_{-\infty}^{t} dt_1 \; H_I(t_1)\Psi(t_1)\,. \tag{5}$$

In particle physics experiments, one is typically interested in the calculation of scattering processes. This only requires information about the asymptotic state $\Psi(+\infty)$ of the system under consideration. This state can be formally obtained from the initial state by applying the so-called S-matrix:

$$\Psi(+\infty) = S \; \Psi(-\infty). \tag{6}$$

If S is known, all relevant observables, such as scattering cross sections, can be obtained easily. More explicitly, S is given by

$$S = \sum_{n=0}^{\infty} \frac{(-\mathrm{i})^n}{n!} \int d^4x_1 \ldots d^4x_n \; P\{H_I(x_1) \ldots H_I(x_n)\} \;, \tag{7}$$

with the time ordering operator P. Based on the last equation, a perturbation expansion can be derived which in turn can be translated into the language of Feynman diagrams. These diagrams provide an intuitive identification of all contributions to S with a representation of elementary spacetime processes which lead from the initial state to the final state while respecting all relevant conservation laws.

In quantum electrodynamics (QED), the theory we will focus on for a while, S can be expanded in powers of the fine structure constant α_0: $S = S^{(0)} + S^{(1)} + S^{(2)} + \ldots$, with $S^{(n)}$ being proportional to α_0^n. Due to the smallness of $\alpha_0 \approx 1/137$, a small number of terms suffices to determine S with good accuracy.

But complications arise since divergences show up. Before pointing this out in some detail, a change of notation will be useful: since $S = 1$ describes the trivial reaction that the finite state is identical to the initial state, it is useful to introduce the so-called T-matrix which only captures non-trivial reactions: $S = 1 + \mathrm{i}\; T$.

2.2.1. Renormalisation in QED

To present the idea of renormalisation, let us focus on one specific process and let us use the language of Feynman diagrams.[10] The process we will focus on is the scattering of an electron by an external potential A_{ext} (cf. Fig. 2). The first divergent term in the perturbation expansion of the S-matrix is depicted in Fig. 3. In this diagram, a photon is emitted by the 'incoming' electron and reabsorbed by the 'outgoing' electron. This photon is called *virtual* because it does not show up in the Feynman diagram as an external line with one loose end.

It is now interesting to ask what the energy and the momentum of this virtual photon are. Although energy and momentum conservation hold at the vertices, this does not fix these values uniquely. A whole spectrum of values is possible, ranging from zero to infinity. In the mathematical formalism this means that all those contributions have to be added up, and here lies the source for the divergences.

In more technical terms, the story so far amounts to this: let k be the 4-momentum of the virtual photon, and let p be the momentum of the incoming electron with bare mass m_0 and bare charge e_0. 4-momentum

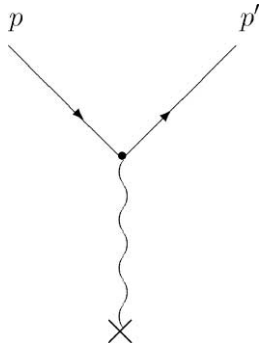


Fig. 2. *The main contribution to the potential scattering of an electron.*

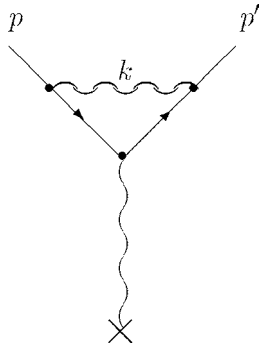


Fig. 3. *Vertex-correction to the potential scattering of an electron.*

[10] I am following the clear presentation given by Lepage (1989). See also Mills (1993).

conservation then requires that the momentum of the electron after the emission of the virtual photon is $p - k$. Similarly, if p' denotes the final electron 4-momentum, the 4-momentum of the electron immediately after the interaction with the external potential is then $p' - k$. 4-momentum conservation does not give us any more information. k is completely undetermined and all positive real numbers are possible for its absolute value. All resulting contributions then have to be integrated up to get the T-matrix. Applying the Feynman rules of QED, one obtains

$$T^{(a)} = \\ -e_0^3 \int_0^\infty \frac{d^4k}{(2\pi)^4} \frac{1}{k^2} \cdot \bar{u}(p')\gamma \frac{1}{(p'-k)\cdot\gamma - m_0} A_{ext}(p'-p)\cdot\gamma \frac{1}{(p-k)\cdot\gamma - m_0}\gamma u(p), \tag{8}$$

with spinors $u(p)$ and $\bar{u}(p')$ for the incoming and outgoing electron and the 4-vector of Dirac's γ-matrices denoted by γ. This integral is divergent since the numerator is proportional to k^3dk, while the denominator is proportional to k^4 for large k: $T^{(a)}$ diverges logarithmically. Many other divergent terms like this show up in higher orders of the perturbation expansion, and many physicists came to the conclusion that this is an indication of a serious inconsistency in the very foundations of the theory.

Pragmatically oriented as many physicists are, the empirical success of the renormalisation procedure developed around 1949 overruled these negative feelings. The suggested renormalisation schemes were not only a way to eliminate the infinities; they also led to new predictions and explanations of tiny corrections to observables. These quantities (such as the anomalous magnetic moment of the electron or the Lamb shift) have been measured in the laboratories with a remarkable accuracy. This is generally considered as an impressive confirmation of QED and the renormalisation scheme it applies. Let us see in some more detail how this works.

First, the divergent contributions have to be eliminated in all orders of perturbation theory. It turned out that in QED these terms have the same structure as terms which showed up already in the original Lagrangian density. Hence, they can be eliminated by a suitable redefinition of the parameters of the original Lagrangian density. These parameters are the mass and the charge of the electron:

$$m_0 \to m_R = m_0 + \delta m, \quad e_0 \to e_R = e_0 + \delta e. \tag{9}$$

Here, m_R is the renormalised mass and δm is the modification of the mass due to higher order contributions to the perturbation theory. Simlilarly, e_R is the renormalised charge and δe includes the radiative corrections.

Dyson showed that a reparametrisation of this kind can be carried through in all orders of perturbation theory in QED.[11] Now, while this procedure is so

[11] Later, Salam and Weinberg completed Dyson's original proof; cf. Cao (1993, pp. 42f).

far purely mathematical and might look like a trick, a physical interpretation is required to justify this procedure. The basic idea behind this justification is to identify the renormalised mass and charge of the electron with its 'physical' (i.e. real) mass and charge. To get the observed finite values for these parameters, it has to be assumed that the bare mass and charge of an electron are also infinite, cancelling the infinite radiative corrections δm and δe. In the absence of the radiation field, the mass as well as the charge of the electron would be infinite. It is the 'switching-on' of the radiation field which accounts for the finite mass and charge of the electron.

At first sight it might sound problematic to attribute an infinite bare mass and charge to the electron. Indeed, Richard Feynman meant exactly this when he claimed that in QED the real problems are swept under the rug. For Feynman it does not help to simply state that there is no real problem because a finite result can be obtained by subtracting one infinite number (the calculated radiative correction) from another infinite number (the—fortunately!—unobservable bare mass or charge).

But the situation is more subtle than this suggests. Let us go back to the example of the electron interacting with an external potential. Equation (8) represented the contribution of the Feynman diagram from Fig. 3 to the scattering amplitude. Instead of integrating over all k up to infinity, let us first introduce an upper limit Λ_0 to this integral. Doing so can be understood as a purely formal trick to keep the integral well-defined and the numbers finite. The value of the original integral can then be obtained by performing the limit $\Lambda_0 \to \infty$ at the end of the calculation.

Next, let us go one step further. We consider another theory which has a cut-off $\Lambda < k < \Lambda_0$ and ask how the original theory with cut-off Λ_0 has to be modified in order to produce the same results as the theory with cut-off Λ. It turns out that the following Lagrangian density has to be subtracted:

$$T^{(a)}(k > \Lambda) = \\ -e_0^3 \int_{\Lambda}^{\Lambda_0} \frac{d^4k}{(2\pi)^4} \frac{1}{k^2} \cdot \bar{u}(p') \; \gamma \frac{1}{(p'-k)\cdot\gamma - m_0} \; A_{ext}(p'-p)\cdot\gamma \; \frac{1}{(p-k)\cdot\gamma - m_0}\gamma \; u(p). \tag{10}$$

To proceed, let us assume that all masses and external momenta (p and p') are much smaller than Λ, so that the quantities m_0, p and p' can be neglected in the integrand. One obtains:

$$\begin{aligned} T^{(a)}(k > \Lambda) &\approx -e_0^3 \int_{\Lambda}^{\Lambda_0} \frac{d^4k}{(2\pi)^4} \frac{1}{k^2} \bar{u}(p') \; \gamma \frac{k\cdot\gamma}{k^2} \; A_{ext}(p'-p)\cdot\gamma \; \frac{k\cdot\gamma}{k^2}\gamma u(p) \\ &\approx -e_0^3 \; \bar{u}(p') \; A_{ext}(p'-p)\cdot\gamma \; u(p) \int_{\Lambda}^{\Lambda_0} \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2)^2}. \end{aligned} \tag{11}$$

In order to get the full T-matrix in this order of perturbation theory, other diagrams have to be included as well. Treating them in the same way as described above, the following result obtains for that part of the scattering amplitude which can be neglected due to the introduction of the new cut-off Λ:

$$T(k > \Lambda) \approx -\mathrm{i}\ e_0\ c_0(\lambda/\Lambda_0)\bar{u}(p')\ A_{ext}(p' - p) \cdot \gamma\ u(p), \tag{12}$$

with a dimensionless quantity c_0, which only depends on the ratio Λ/Λ_0:

$$c_0(\Lambda/\Lambda_0) = -\frac{\alpha_0}{6\pi}\ \log(\Lambda/\Lambda_0). \tag{13}$$

Note that Λ and Λ_0 are the only energy scales at high energies for this problem since p, p' and m_0 have been neglected.

Now $T(k > \Lambda)$ is certainly an important contribution to the scattering amplitude which cannot be ignored. However, it is possible to 'simulate' the contribution of this term in the Lagrangian density with cut-off Λ by adding a *counter term* of the form:

$$\delta\mathscr{L}_0 = -e_0\ c_0(\Lambda/\Lambda_0)\bar{\psi}\ A \cdot \gamma\ \psi. \tag{14}$$

Astonishingly, $\delta\mathscr{L}_0$ has the same structure as the current-field coupling term in the original Lagrangian density of QED. The effect of the additional term can therefore be included by a redefinition of the charge parameter: $e_0 \to e_0[1 - c_0(\Lambda/\Lambda_0)]$.

To sum up: in a renormalisable quantum field theory such as QED, the contributions of the high-energy sector of the theory can be effectively taken into account by a reparametrisation of the original theory. As the above derivation shows, this procedure does not presuppose that the original cut-off Λ_0 goes to infinity. However, if one calculates this limit one is bound to assume that the bare mass and charge are indeed infinite in order to account for the finite values of various quantities which we obtain in experiments.

A theory is called *renormalisable* if such an absorption of the divergent contributions of the high-energy sector of the theory in the mass and charge parameter(s) can be accomplished at all orders in perturbation theory. Doing so is, however, not possible for all quantum field theories. In fact, only a very small subclass of all quantum field theories is renormalisable. Motivated by the enormous empirical success of QED, renormalisability soon became *the* selection criterion for the construction of new quantum field theories. In his speech during the award of the Nobel price for 1979, Steven Weinberg argued this point:

> To a remarkable degree, our present detailed theories of elementary particle interactions can be understood deductively, as consequences of symmetry principles and of the principle of renormalizablity which is invoked to deal with the infinities (Weinberg, 1980a, p. 515).

Somewhat later in this speech, Weinberg addressed the issue of the role of renormalisability for his own work:

> I learned about renormalization as a graduate student, mostly by reading Dyson's papers. From the beginning it seemed to me to be a wonderful thing that very few quantum field theories are renormalizable. Limitations of this sort are, after all what we most *want*; not mathematical methods which can make sense out of an infinite variety of physically irrelevant theories, but methods which carry constraints, because constraints can point the way towards the one true theory [.... At the time] I thought that renormalizability might be the key criterion, which also in a more general context would impose a precise kind of simplicity on our theories and help us pick out the one true physical theory out of the infinite variety of conceivable quantum theories [...]. I would say this a bit differently today, but I am more convinced than ever that the use of renormalizability as a constraint on our theories of the observed interactions is a good strategy (Weinberg, 1980a, p. 517).

Renormalisation started as a pragmatic scheme which allowed efficient and precise predictions but which was theoretically considered to be unsatisfying and perhaps only provisional. It soon became the selection criterion (besides symmetry principles) for future quantum field theories. The theories found in this way (such as the Standard Model) turned out to be highly successful. The subsequent developments, summarised in the next section, led to a rehabilitation of non-renormalisable theories. Renormalisation is not any longer considered to be a decisive criterion for theory choice, but one should not forget to address the issue why renormalisable theories such as the Standard Model are so successful.

2.2.2. A new conceptualisation of renormalisation

Let us summarise the renormalisation story told so far: in a renormalisable theory, the divergent terms in the perturbation can be eliminated by redefining the parameters of the original theory. The standard reasoning for this is that the bare masses and charges of the particles involved are not observable because of the ineliminable presence of radiation fields. These parameters can therefore be chosen freely, e.g. one can choose them to be infinite. The observed values of the masses and charges are then due to compensation effects of the radiative corrections.

This story did not convince everyone. One of the most severe critics was Paul Dirac, the founder of the old quantum theory of fields in which the problem with infinities first occurred. Several years later he wrote:

> This [i.e. the renormalisation programme] is quite nonsense physically, and I have always been opposed to it. It is just a rule of thumb that gives results (Dirac, 1983, p. 55).[12]

[12] See also Kragh (1990, pp. 165f).

However amazed by the empirical success of the renormalisation machinery and the possibilities it opened up of constructing new theories such as the Standard Model, most physicists did not follow Dirac's skepticism. Those who did, such as the advocates of the programme of Axiomatic QFT, did not succeed in finding the supposed inconsistency in the foundations of the theory.[13] The historical development took a different path. The key to make sense of renormalisation came from within physics and the guiding idea was to interpret realistically the cut-off parameter which showed up in renormalisation schemes. It represents the energy scale up to which the theory in question is applicable.

To see this, let us go back to our example. It turned out to be helpful to replace the upper limit of the integral by some cut-off Λ. The standard procedure, first suggested by Feynman, is to take the limit and let Λ go to infinity at the end of the calculation. For Feynman, the cut-off was a purely formal calculational device ('formal interpretation'), and indeed there was a good reason for this, since a realistic interpretation of the cut-off leads immediately to a new problem.

In this case, photons with all energies contribute, for example, to the diagram of Fig. 3. Now, photons and electrons are not the only particles present at high energies. And these other particles (such as protons, muons and pions) also couple to the electron and the photon. If one interprets the cut-off parameter realistically, there is no reason why these particles and their interactions should not be included in the calculation. But in that case, maybe it is just these contributions that make the theory finite. According to the realistic interpretation of the cut-off (if taken in the limit $\Lambda \to \infty$), only the final theory, if there is any, should be finite. So there is no reason anymore why QED with photons and electrons should be renormalisable.

But there is another way to interpret this situation. One can say that there is no reason anymore to consider theories with a finite cut-off as second rank if the cut-off parameter is interpreted realistically. This parameter then just reflects the production threshold of new particles (muons, pions, etc.) and the theory is henceforth only applicable up to this energy. A realistic interpretation of the cut-off leads to a *rehabilitation of theories with a finite cut-off*.

The formalism sketched so far can be easily extended to solve the following problem: given a theory with cut-off Λ_1, how can we get from there to a theory with a higher cut-off Λ_2? If no new particles show up in the energy regime between Λ_1 and Λ_2, it turns out that only the parameters of the theory (masses, charges) have to be changed ('renormalised'). Consequently, the masses and charges of particles depend on the energy scale under consideration; they have no absolute values. To get the values of these parameters on a higher or lower scale, the so-called *renormalisation group equations* have to be solved.[14]

[13] Cf. Wightman (1986) for a discussion of renormalisation in this programme. See also Cao (1997, pp. 217–219).

[14] See Fischer (1999) for a general introduction.

Let us again illustrate this with our example from QED. As we saw, the elimination of states above some energy $\Lambda_0 < \Lambda$ (with Λ being the cut-off of the original theory) can be obtained by adding additional *local* terms of the form

$$\delta\mathscr{L} = -e_0 c_0\left(\frac{\Lambda}{\Lambda_0}\right)\ \bar{\psi}\ \gamma_\mu A^\mu\ \psi - m_0\tilde{c}_0\left(\frac{\Lambda}{\Lambda_0}\right)\ \bar{\psi}\ \psi, \tag{15}$$

with dimensionless parameters c_0 und $\tilde{c}_0$ proportional to $\ln \Lambda/\Lambda_0$. The new Lagrangian density is then

$$\mathscr{L}' = \mathscr{L}_{QED} + \delta\mathscr{L} = \bar{\psi}(\mathrm{i}\gamma_\mu\partial^\mu - e_\Lambda\gamma_\mu A^\mu - m_\Lambda)\psi, \tag{16}$$

with the *renormalised* charge and mass given by

$$e_\Lambda = e_0\left[1 + c_0\left(\frac{\Lambda}{\Lambda_0}\right)\right], \quad m_\Lambda = m_0\left[1 + \tilde{c}_0\left(\frac{\Lambda}{\Lambda_0}\right)\right]. \tag{17}$$

Changing the cut-off can therefore be compensated by changing the parameters of the theory. It can be shown that the parameters obey the *renormalisation group equations*:

$$\Lambda\frac{de_\Lambda}{d\Lambda} = \beta(e_\Lambda),\ \Lambda\frac{dm_\Lambda}{d\Lambda} = m_\Lambda\gamma(e_\Lambda), \tag{18}$$

with appropriate functions β and γ.

These equations grew out of the research of Michael Fisher, Leo Kadanoff and most importantly Kenneth Wilson in the context of solid state physics. It happened to have fruitful applications in particle physics as well. Here is the upshot of all this: (1) interpreting the cut-off realistically leads to a reappraisal of theories with a finite cut-off. (2) There is a systematic algorithm as to how to change the parameters of a theory when the energy scale is changed. (3) This algorithm, introduced here only for the case of a specific renormalisable theory, can be extended to other quantum field theories, including non-renormalisable quantum field theories, and to applications where new particles show up in the energy regime between the two cut-offs.

2.2.3. The decoupling theorem

Only one more idea is missing to make EFTs a powerful tool: the decoupling theorem, proved by Appelquist and Carazzone in 1975. In its simplest case, this theorem demonstrates that for two coupled systems with different energy scales m_1 and m_2 (with $m_2 > m_1$) and described by a renormalisable theory, there is always a renormalisation condition according to which the effects of the physics at scale m_2 can be effectively included in the theory with the smaller scale m_1 by changing the parameters of the corresponding theory. The decoupling theorem implies the existence of an EFT at scale m_1 which will, however, cease to be applicable once the energy gets close to m_2.

One might think that this is not such a spectacular result since the whole edifice of physics is grounded on the assumption that empirical reality is layered so that for the physics at a given energy scale the details of the physics at much higher energies do not really matter. The idea of eliminating the

physics at higher energies to get an effective account that is valid only at lower energies is also quite popular in other parts of physics. In his beautiful book *Qualitative Methods in Quantum Theory*, Migdal (1977) discusses an instructive example from quantum mechanics. Let S be a system which is composed of a fast subsystem S_f and a slow subsystem S_s, characterised by two frequencies ω_f and ω_s. It can be shown that the effects of S_f on S_s can be taken into account effectively by adding a potential energy term to the Hamiltonian operator of S_s. In this case, as well as in many other cases, one ends up with an *effective Hamiltonian operator* for the subsystem characterised by the smaller frequency (or energy). It is interesting, however, that the decoupling theorem holds, given certain assumptions, also in quantum field theory. This is far from trivial if one recalls all those complicated radiative corrections which have to be taken into account here.

The decoupling theorem gives further legitimacy to non-renormalisable theories. If that theorem holds, the physics at higher energies can be effectively included in the parameters of a non-renormalisable EFT. Higher energy scales decouple and empirical reality seems to be divided into a set of 'quasi-autonomous domains', each theoretically captured by an EFT which employs only those particles and their interactions that are relevant at that scale. The domains are only quasi-autonomous since the effects of the physics at higher energy scales get more important once the energy reaches the cut-off energy of the EFT under consideration. It should be noted that EFTs can also be divined or obtained if one is not in the possession of a fundamental and renormalisable theory such as QED. It might, however, be more difficult to 'anchor' these theories, as the example of Fermi's theory showed. In these cases, finding a suitable EFT is more like guessing.

Does the decoupling theorem imply that empirical reality is, as a whole, layered into quasi-autonomous domains, as suggested by some authors? No, since the decoupling theorem is based on assumptions which may not always be fulfilled. Most importantly, the decoupling theorem (as proved by Appelquist and Carazzone) presupposes that there is a renormalisable theory of the composite system which is the starting point of the decoupling procedure. Without such a theory, which is supposed to be valid on all energy scales, the decoupling theorem cannot be applied. Furthermore, the decoupling theorem presupposes that different mass scales exist in the underlying renormalisable theory. But sometimes mass scales do not separate neatly, as examples from the theory of complex systems (such as turbulence) demonstrate. In these cases the physics at high energies cannot simply be absorbed in the parameters of a low energy theory and the picture of empirical reality as layered into a hierachy of quasi-autonomous domains turns out to be too wild an extrapolation.

2.3. Two ways to apply EFTs

There are two ways to use EFTs in physics, the *bottom-up approach* and the *top-down approach*. I will describe both of them in some detail.

2.3.1. The bottom-up approach

This strategy is closely related to observable phenomena and some think that this is the way physics has to proceed. We will look at these arguments below (Section 4.2).

To apply this strategy, two scenarios have to be distinguished. First, there might be no relevant theory at all. In this case one has to start from scratch and construct a Lagrangian density from the particles, symmetries and interactions assumed to be relevant at the energy scale under consideration. Second, there is already some EFT T_1 which represents the physics at some energy scale, characterised by a cut-off parameter Λ_1. This theory might be, for example, QED or the Standard Model, both of which are—despite being renormalisable—considered to be EFTs. They might be applicable only up to some maximal energy Λ_1. At higher energies, new phenomena might happen to show up, and T_1 does not account for them. In order to obtain a new theory T_2 (valid up to some energy $\Lambda_2 > \Lambda_1$) from the old theory T_1, two more cases have to be distinguished:

1. There are no new particles between Λ_1 and Λ_2. In this case all the parameters of T_1 (i.e. charges and masses) have to be modified according to the renormalisation group equations. If the energy is less than Λ_1, both theories will give the same results for observable phenomena. But T_2 can also be applied for the energy range between Λ_1 and Λ_2. It should be noted that the relation between T_1 and T_2 for energies up to Λ_1 is very interesting. On the one hand, the theories differ from each other because their respective mass and charge parameters have different numerical values, while on the other hand, both are empirically equivalent.
   Even if there are no new particles in the energy regime between Λ_1 and Λ_2, new interactions between the old particles might become important. T_2 is then constructed by including these new interactions in the Lagrangian density of T_1. In order to save the phe-nomena accounted for already by T_1, some of the parameters of T_1 might have to be changed and new parameters have to be adjusted appropriately.
2. There are new particles between Λ_1 and Λ_2. This case is, of course, the more complicated one. T_2 is now constructed in several steps. First, the masses and charges of T_1 have to be adapted to the new energy scale; again, this is done by solving the renormalisation group equations. Second, all new particles which show up in the energy regime between Λ_1 and Λ_2 have to be identified. Are they fermions or bosons? What is their mass and charge (on the scale Λ_2)? How do they couple to the other particles? The formalism of quantum field theory presents a tool box to systematically construct the new terms in the Lagrangian density of T_2. In many cases, the relevant coupling constants have to be adjusted to experimental results. This procedure therefore has a *theoretical* (or *a priori*, if you like) and an experimental (or *a posteriori*) component. The structure of the new terms follows from the general formalism of quantum field theory. The masses, charges and

coupling constants have to be determined on the basis of experiments. This procedure is, of course, not completely theory-free. The determination of the relevant parameters takes place on a given energy scale, and auxilliary (or measurement) theories have to be used to determine their numerical values. These measurement theories also work on a given energy scale, and consistency must be achieved in this whole process.

There are several examples of the bottom-up approach in physics. The Fermi theory of weak interactions, discussed in Section 2.1.2, is a good example of a theory which had to start from scratch. Doing so, one might be mistaken, for example, when it comes to specifying the correct interaction between the relevant particles, as the discovery of parity violation for weak interactions shows. The work based on taking the general theory of relativity as an EFT is a good example of case 1 mentioned above. Here new interactions are included which correct Einstein's theory at higher energies. The new theories obtained by following this approach have been interpreted tentatively as low-energy limits of a quantum theory of gravity. We do not know this theory yet, but following the EFT approach might eventually lead to new ideas as to what such a theory might look like (Donoghue, 1994a,b). The problem here is of course that there are almost no experimental data available which can be used to fix the parameters in the new theory (as in the case of Fermi's theory). It might, however, be possible to derive these parameters from candidate theories for a quantum theory of gravity. Supersymmetric extensions of the Standard Model are an example of case 2. Here the Standard Model is essentially duplicated by including the supersymmetric partners of all particles which are already present in the Standard Model (Meissner, 1992). Again, the problem here is that there are no experimental data yet that can be used to fix the new parameters (such as the masses of the supersymmetric partners of the leptons and quarks).

2.3.2. The top-down approach

This strategy starts with a more fundamental theory which is valid on a given energy scale Λ_1. The aim is now to construct an EFT for lower energies $\Lambda_2 < \Lambda_1$. There is a systematic procedure for getting these low-energy theories. Once the original theory is renormalisable and the decoupling theorem holds, a tower of EFTs can be uncovered in this way. A typical example of this strategy is the theory of Euler and Heisenberg, discussed in Section 2.1.1. Here, a purely photonic theory was obtained from QED by eliminating all electronic degrees of freedom. The resulting EFT is then valid for photons whose energy is much smaller than the rest mass of the electron. Other examples of this strategy are the various attempts to justify some kind of superstring theory. These theories cannot be tested experimentally at a typical energy of a superstring. Instead, systematic low energy expansions are carried out in order to obtain low energy footprints of the high energy regime of this theory. Yet another example of the

application of the top-down strategy is provided by nuclear and particle physics. I will look at this case in some more detail in the next section.

3. Theorising in Nuclear and Particle Physics: A Case Study

The present situation in theoretical nuclear and particle physics is rather involved. On the one hand, there is the Standard Model, a renormalisable and well-confirmed theory which should, in principle, account for all phenomena which are not gravitational. On the other hand, there are all sorts of models and EFTs which are often used in practical applications. Especially in the sector of the Standard Model which deals with strong interactions, a plurality of theoretical accounts can be identified. They all seem to coexist peacefully, and they all seem to complement each other in a way which I will investigate in more detail below. In this section, I shall focus on the physics of the strong interaction only and first introduce quantum chromodynamics, the underlying theory, as well as some typical models and EFTs in this part of physics (Section 3.1). By focusing on their respective functions in the research process, Section 3.2 argues that all of them are indispensable and Section 3.3 points out various interrelations between them. This case study will be the basis of my argument for some variant of pluralism in the next section.

3.1. Theories, models and EFTs

For a long time, the physics of strong interactions lacked a fundamental theory. There were phenomenological models, all of which could be applied for some purposes, but all of which had their well-known limits. Among the models used extensively in nuclear physics are the liquid drop model and the nuclear shell model. The liquid drop model helps to understand nuclear fission, as Niels Bohr and John A. Wheeler pointed out, but fails to explain why certain configurations of protons and neutrons are particularly stable. These 'magic numbers' of protons and neutrons can be naturally explained with the nuclear shell model which has, in turn, other deficiencies. Among the models of the constituents of the nuclei (protons, neutrons and pions) are various bag models, chiral quark models, the purely bosonic Skyrme model, and approaches which utilise sum rules derived in the spirit of the *S*-matrix tradition.

All these models were considered provisional at best, and applied and studied because of a lack of a more satisfactory alternative. Fortunately, this alternative was found in the early nineteen-seventies.[15] Quantum chromodynamics (QCD) was born, but it soon turned out that tractable applications of this theory could only be obtained in the high-energy regime. Nuclear physics and the theory of hadron structure remained almost completely

[15] For a reconstruction of this development and for an analysis of the role models played in this context see Hartmann (1995a,b).

unaffected and developed quite independently for a long time. There have been some attempts to derive low energy results from QCD, but these endeavours turned out to be technically extremely hard and rather uninteresting and unilluminating. Other QCD-inspired research in hadron physics involved the qualitative modelling of features of QCD such as confinement and dynamical chiral symmetry breaking. These models (see Section 3.1.2) have been quite successful, but their formal relation to QCD is far from clear. This is where EFTs come in handy. EFTs allow a systematic low energy expansion of QCD, and many of the old models could be given a more solid foundation.

3.1.1. The theory: quantum chromodynamics

QCD is generally considered to be the fundamental theory of strong interactions. It is a renormalisable gauge theory, and its fundamental entities are the fermionic quarks (spin-1/2) and the bosonic gluons (spin-1). There are six different kinds of quarks ('flavours'): *up, down, strange, charm, bottom* and the recently discovered *top*. Besides spin and flavour, quarks have an additional degree of freedom which is called 'colour'. Gluons, the exchange particles of the strong interactions, show up in eight different kinds and, unlike photons in QED, directly interact with each other. This fact, which is a consequence of the internal colour structure of the gluons, along with the large value for the coupling constant of QCD, makes actual calculations very complicated and involved. The self-interaction of the gluons follows mathematically from the non-commutativity of the generators of the corresponding gauge group, colour-SU(3). Here is the Lagrangian density of QCD:

$$\mathscr{L}_{QCD} = \bar{\psi}(\mathrm{i}\gamma_\mu D^\mu - \hat{m}_0)\psi - \frac{1}{4}F_{k\mu\nu}F_k^{\mu\nu}. \tag{19}$$

ψ represents the quark field and $F_k^{\mu\nu}$ (with $k = 1, \ldots, 8$ for the eight gauge degrees of freedom) is the field strength tensor associated with the gluons. The operator D^μ fixes the gauge invariant coupling of the quarks and the gluons and $\hat{m}_0$ is the mass matrix of the quarks; the quarks which show up in the Lagrangian density of QCD are also called 'current quarks', as opposed to the much heavier 'constituent quarks' of non-relativistic quark models. This matrix cannot be deduced from first principles and has to be adjusted to experimental data.

For most low-energy applications, exact consequences of Eq. (19) can only be obtained numerically with a method called lattice gauge theory (see Montvay and Münster, 1994). Here quark and gluon fields are defined on a lattice with a finite spacing a; exact results can be obtained by running extensive computer simulations with finite a and extrapolating the results to the continuum limit $a \to 0$. Although this method also suffers from technical problems, my main point about lattice gauge theory is that it effectively works like a black box. Technical problems aside, lattice gauge theory produces the exact results of QCD and hence makes tests of this theory possible. However, it does not reveal more about the concrete mechanisms which account for the

calculated result. Like a black box-theory, lattice gauge theory yields consequences of a theory, but it does not produce insight and understanding.[16]

This account of 'black-boxism' does not square with traditional black box theories such as behaviourism. These theories were criticised for not providing a detailed mechanism for the dynamics of a system. According to Bunge (1964), these mechanisms have to be provided by a fundamental theory.[17] In the case of strong interactions, QCD does indeed specify the overall dynamics of the system; there are quarks and gluons, and these entities interact in a very complicated way with each other according to the Lagrangian density of QCD. But not much more can be said: the rest has to be done numerically with the help of high-powered computers (cf. Lepage, 1994). And computers function like a black box. All possible Feynman diagrams are summarised, although, perhaps, only a few of them (or a certain subclass of them) produce almost the whole effect under investigation. A knowledge of these actually relevant processes would produce insight and understanding. Lattice gauge theory does not produce this insight, and QCD is, therefore, effectively a black-box theory.

In order to learn something about the actually relevant processes, models and EFTs are applied. While EFTs can be directly obtained from QCD by following well-defined procedures, models usually extract one or more of the general features and consequences of the theory and explore their implications. Some of these general features and consequences are well known. Among them are the following three (Donoghue *et al.*, 1992).

First, QCD is *asymptotically free*. This means that quarks move freely at very high energies. At low energies (energies of the order of the rest mass of the proton), the reverse effect shows up and quarks and gluons interact very strongly with each other. This is why perturbation theory, which works so well in QED, cannot be applied here.

Second, QCD exhibits *quark confinement*. First introduced to account for the fact that no one ever observed a free quark, it now seems clear that quark confinement is a strict consequence of QCD. But what does confinement really amount to? This is not so clear and there are several options to be found in the literature. Some argue that the interaction between the quarks increases with their distance, others favour a model according to which quarks are bound inside some solid sphere which prevents the existence of free quarks. And there are other, more technical proposals (for details see Hartmann (1999)).

Third, low energy QCD is (almost) *chirally invariant* and exhibits *dynamical chiral symmetry breaking*. Unlike confinement, this phenomenon is well understood. Here is the basic idea. The masses of the quarks in the Lagrangian density of QCD (see Eq. (19)) are very small (about 10 MeV) compared to the typical energy scale of strong interactions (about 1 GeV). Let us therefore assume that quarks are massless. The Lagrangian density of QCD then exhibits

[16] This point is elaborated in Hartmann (1999).

[17] For an interesting discussion of the role of phenomenological theories in physics see also Heisenberg (1966).

another symmetry, called chiral symmetry. As a consequence of this symmetry, so-called left-handed and right-handed eigenstates of the QCD Hamiltonian cannot be distinguished energetically. Every hadron which is an eigenstate of the QCD Hamiltonian should have a chiral partner with the same mass, but with opposite chirality. Now, these chiral partners do not seem to exist. There is, for example, no other particle with the mass and charge of the proton, but with opposite chirality.[18] A way out of this difficulty is to assume that chiral symmetry is dynamically broken. This means that the interaction itself breaks the symmetry so that a large mass gap between the chiral partners emerges. The pion emerges as the corresponding Goldstone boson of the broken symmetry. It should be noted, however, that chiral symmetry is also explicitly broken due to the non-vanishing of the values of the current quark masses in the Lagrangian density of QCD. This effect has some interesting consequences, such as the finite pion mass.

3.1.2. Models

I take a model to be a set of assumptions (augmented, perhaps, by diagrams, sketches, and other visualisations), where some of these assumptions might be inspired by a theory. All other assumptions specify the concrete object or system under consideration. Phenomenological models, like the ones in hadron physics, use theories like a tool box; they pick some of the relevant features of (at least) one theory, fit these into a larger theoretical framework (which might be different from the one employed by the theory), and explore the consequences of the assumptions made. This procedure allows for models to be used as probes for the features of the underlying theory (Hartmann, 1999). The deductive relation between a model and an underlying theory is, however, not at all clear.

Among the relevant features of QCD at low energies are quark confinement and the dynamical breaking of chiral symmetry. While *bag models* concentrate on the first feature, *chiral quark models* explore the consequences of QCD's second main feature. The first and conceptually easiest bag model is the MIT-Bag Model (see Mosel, 1999, Ch. 16). Here quark confinement is included in the model assumptions by restricting the motion of quarks to a finite region in space, the 'bag'. Mathematically this is done by imposing an appropriate boundary condition to the quark wavefunctions which are assumed to be a solution of the (free) Dirac equation for relativistic particles. Bag models like the MIT-Bag Model therefore do not operate in the framework of quantum field theory (such as QCD), but in the framework of relativistic quantum mechanics which is mathematically easier to handle.

Chiral models explore the consequences of chiral symmetry and its dynamical breaking. Some of these models take only quark degrees of freedom into account (such as the Nambu–Jona–Lasinio model), others (such as the Skyrme model) neglect quark degrees of freedom completely and describe

[18] The chirality is given by the projection of the spin on the momentum of the particle.

hadrons in terms of scalar and pseudoscalar meson fields, while yet other models favour an hybrid account of quark and meson degrees of freedom (such as the soliton models of the Friedberg–Lee type) (see Mosel, 1999, Ch. 17). These models have a long history which started in the days of the *S*-matrix programme and the work on current algebra in the nineteen-fifties and sixties. In this context, plenty of experimentally well-confirmed relations between hadron masses have been derived from the assumption of chiral symmetry only. A famous example is the Gell-Mann–Oakes–Renner relation which relates properties of the pion (its mass m_π and its decay constant f_π) to quark properties:

$$m_\pi^2 f_\pi^2 = -\frac{m_u + m_d}{2}\langle \bar{q}q \rangle, \tag{20}$$

where $\langle \bar{q}q \rangle \approx (-250\ \text{MeV})^3$ is the so-called quark condensate.

3.1.3. Effective field theories

The most popular EFT based on QCD is *Chiral Perturbation Theory*, developed by Steven Weinberg, Heinrich Leutwyler and others in the nineteen-eighties. I will give a short outline of this approach along the lines of Leutwyler (1994).

The main idea of Chiral Perturbation Theory is to expand the Lagrangian density of QCD in terms of a typical momentum for the process under consideration. For the sake of simplicity let us assume for a moment that the current quark masses vanish ($m_u = m_d = 0$) and that there are only two quark flavours. This last assumption is reasonable in the low energy regime of about 1 GeV.

In order to get the desired momentum expansion, one first replaces the quark and gluon fields of QCD by a set of pion fields which are, as pointed out above, the Goldstone bosons of the theory due to the dynamical breaking of chiral symmetry. These fields can be conveniently represented by a 2×2 matrix $U(x) \in \text{SU}(2)$. Next, the Lagrangian density of QCD is expressed exclusively in terms of $U(x)$. After this decisive step in the derivation, quark and gluon degrees of freedom do not show up in the Lagrangian density anymore. One obtains:

$$\mathscr{L}_{QCD} = \mathscr{L}_{eff}(U, \partial U, \partial^2 U, \ldots). \tag{21}$$

It turns out that a low energy expansion of this Lagrangian density can be obtained by expanding $\mathscr{L}_{QCD}$ in terms of the derivatives of $U(x)$.[19] The Lorentz invariance of the whole Lagrangian density implies that only terms with an even number of derivatives show up in the truncated expression. One obtains:

$$\mathscr{L} = \mathscr{L}_{eff}^2 + \mathscr{L}_{eff}^4 + \mathscr{L}_{eff}^6 + \ldots \tag{22}$$

[19] This follows from the observation that the momentum operator is given by $p_\mu = -\mathrm{i}\partial_\mu$.

The requirement of chiral symmetry very much constrains the form of the terms in this expansion. The second-order contribution is given by:

$$\mathscr{L}^2_{eff} = \frac{1}{4} f_\pi^2 \; tr[\partial_\mu U^\dagger \partial^\mu U]. \tag{23}$$

This term is essentially determined by the decay constant of the pion (f_π). The next term of order p^4 is already a bit more complicated:

$$\mathscr{L}^4_{eff} = \frac{1}{4} l_1 (tr[\partial_\mu U^\dagger \partial^\mu U])^2 + \frac{1}{4} l_2 \; tr[\partial_\mu U^\dagger \partial_\nu U] \; tr[\partial^\mu U^\dagger \partial^\nu U]. \tag{24}$$

It turns out that these first two terms suffice already for many practical applications.

Why is this procedure called 'Chiral Perturbation Theory'? The reason is this: we have, so far, assumed that the quark masses vanish and that chiral symmetry is hence an exact symmetry of QCD. This is an approximation since chiral symmetry is explicitly broken due to the finite (though small) current masses of the quarks. These effects are taken into account in another perturbation expansion in the quark masses.

The whole procedure of Chiral Perturbation Theory therefore consists of two power series expansions, one in some typical momentum, and the other in the mass matrix $m = \mathrm{diag}(m_u, m_d)$ of the quarks. In order to be consistent, one contribution of a quark mass term in the expansion must correspond to two powers of the momentum. In leading (i.e. second) order one obtains:

$$\mathscr{L}^2_{eff} = \frac{1}{4} f_\pi^2 \; tr[\partial_\mu U^\dagger \partial^\mu U] + \frac{1}{2} f_\pi^2 \; tr[m(U + U^\dagger)]. \tag{25}$$

It is interesting to note that many of the phenomenological models and current algebra relations (low-energy theorems) derived in the nineteen-sixties and seventies can be strictly deduced from QCD and summarised in a compact fashion (cf. Ecker, 1995).

While the original programme of Chiral Perturbation Theory only aimed at applications in hadron physics, the more complicated task of deriving EFTs for nuclear physics from QCD has also been undertaken. This remarkably popular and successful research programme is reviewed by Van Kolck (1999). It demonstrates how the relation between theories of several domains (nuclei—protons, neutrons, and pions—and quarks) can be studied in a mathematically controlled way. This fact has implications for the reductionism debate, which will be discussed in Section 4.2.

3.2. Functions of theories, models and EFTs

Theories, models, and EFTs have various functions in actual scientific practice. They are more or less efficient tools in the process of theorising, helping scientists to reach certain cognitive goals. It turns out that none of these tools serves all the functions scientists are interested in. Consequently, a suitable combination of them has to be applied. We will come back to this in Section 4.1.

3.2.1. The functions of theories

Theories have a wide scope of applicability. QCD, for example, should apply to all phenomena governed by strong interactions. It should not only account for the properties of protons and neutrons, but also for whole nuclei and their interactions as well as for astrophysical objects such as neutron stars. Theories provide a coherent account of a large class of phenomena: they unify phenomena which, at first sight, do not have much to do with each other. When it comes to calculations, theories (such as QCD) give the most precise values for the quantities in question. They are therefore also good tools to predict new effects. Theories constrain the assumptions made in models, and may also suggest models (such as the hadron models mentioned above).

However, the price for universal scope and predictive accuracy is that the theory does not provide *local understanding* of the relevant physical processes. Usually, the theory can only be solved numerically and the entities employed by it, such as quarks and gluons, are 'too far away' from the phenomena in question. An understanding of why neutron stars eventually collapse, for example, is hardly achieved by referring to the dynamics of the myriads of quarks and gluons which supposedly constitute these astrophysical objects. Theories produce *global understanding* by fitting an object or system under consideration into a bigger framework, but tend to fall short in their efforts to produce local understanding (see also Section 4.3).

3.2.2. The functions of models

Models, on the other hand, produce local understanding. They often go with a causal-mechanistic story and aim at capturing the essential physics of a phenomenon in a few assumptions (with only a few parameters), just as a caricature represents a person with a few brush lines. As a consequence of this, models are easy to handle mathematically (compared to a more fundamental theory), and deductive consequences of the model can be obtained in an efficient manner. This pragmatic (or computational) superiority of models to theories can hardly be over-estimated (see Humphreys, 1994, 1995). Models are also heuristically very important; they often play a decisive role in the construction of more fundamental theories (such as QCD, as I showed in detail in Hartmann (1995a,b)) or suggest strategies to derive EFTs from a more fundamental theory.[20]

Among the drawbacks of models are the following. The assumptions made by models often lack a deeper foundation; sometimes they are just *ad hoc* in order to save a phenomenon. As the case study has shown, there might not be a 'controlled' deductive relation between the model and an underlying theory, and if a derivation of the model from the theory is actually carried through, further assumptions have to be made to obtain the model, and these assumptions (which might turn out to be more dubious than the assumptions made by the original model) again require a justification, and so on, *ad infinitum*. The

[20] A more complete list of the various functions of models can be found in Hartmann (1999).

parameters which enter a model are sometimes derived from the theory; often, however, they are simply adjusted to experimental results. So they often require a deeper theoretical underpinning. Assumptions made by models might contradict assumptions made by other models, or by a theory for that matter. The MIT-Bag Model, for example, violates chiral symmetry, and many chiral quark models lack confinement. And yet, as I have argued in Hartmann (1999), the models and the underlying theory are bound together in some sense. There often is a story which connects the vocabulary of the model to the vocabulary of the underlying theory even if there are no formalised ontological bridges in the sense of Rohrlich (2000). We shall come back to this in Section 4.1.

3.2.3. The functions of EFTs

EFTs share many of the functions of theories and models. Like models, they provide a local, intuitive account of a given phenomenon in terms of the degrees of freedom which are relevant at the energy scale under consideration. They are relatively easy to solve and to apply, and they are heuristically useful. This is demonstrated by the Fermi theory and the $V - A$ theory which eventually led to the Standard Model, as well as by the EFTs which are used to test the low-energy regime of a future quantum theory of gravity. Like theories, EFTs are part of a bigger picture or framework, from which they can be derived in a controlled way. They help to make predictions and to test the theory they relate to. EFTs avoid the disadvantage of theories of being 'too far away' from the phenomena.

In practice, however, EFTs often contain more adjustable parameters than a model of the same system. Besides, EFTs are only applicable if the energy scales of a system separate well. That is why EFTs work well in particle physics, but do not work so well in the physics of complex systems. Here models and perhaps more fundamental theories are required. It should also be noted that EFTs are closely related to the general framework of QFT. If this framework theory breaks down at some energy and, say, a superstring theory takes over, the whole idea of EFTs might also be vitiated. Then it would have to be clarified in which sense the old EFTs can be recovered or obtained within a certain limit from the new theory.

3.3. Relations between theories, models and EFTs

The theories, models, and EFTs discussed in this section are intimately related to each other. There are various interactions and dependencies as well as conceptual and cognitive relations between these tools which I will now point out by going through the material presented in the case study above.

Theories and models are often not related deductively. However, theories may inspire models which pick out a feature of a more fundamental theory, such as confinement or chiral symmetry in the case of QCD, and embed it in a less complex theoretical framework (such as non-relativistic quantum mechanics). By doing so, models function as a probe to explore the consequences of

just one aspect of the theory. This leads to a better understanding of the physics represented by the theory. Theories may also be used to fix the numerical value of the parameters of a model. Many quark models, for example, employ the strong coupling constant which is taken from QCD. Other models have adjustable parameters which can be calculated directly from the underlying theory (although this is often not done in order to have more freedom to adjust the model to experimental data). Models, on the other hand, often play a role in the construction process of a theory (as the QCD example demonstrates).

Models and EFTs are not always easy to distinguish. Sometimes, a newly developed EFT or some consequence of it turns out to be identical to a model developed many years earlier. An example of this is the work on current algebra. Many of the results which were derived in this framework since the nineteen-sixties turn out to be consequences of chiral perturbation theory. The EFT then provides additional support for the model. Some EFTs are treated like models because no attempt has been made to calculate, for example, the coupling constants and renormalised masses from first principles. They are simply fitted to experimental data. In some cases, models are employed to obtain these parameters.

EFTs and theories may be related in a deductive sense, provided that there is a theory. EFTs then serve to apply and test the theory because they are easier to handle mathematically. They point out the relevant mechanisms at a given energy scale, which helps to better understand the physics covered by the theory. If there is no theory, following the bottom-up strategy of constructing a tower of EFTs might eventually give scientists a hint as to where to look for a more fundamental theory. But of course, there is no direct way from, say, QED to quarks; just using the tricks of the renormalisation group will not take you from here to there. This is where creativity and imagination comes in.

Most important are the cases where theories, models and EFTs complement each other. The establishment of dynamical chiral symmetry breaking as a feature of QCD, for example, resulted from the interaction of all three approaches. Lattice gauge calculations suggested models, consequences of models were used to derive an EFT, which in turn inspired other models and allowed for analytical results. It is this interaction between various tools that makes scientific research so exciting.

4. Some Philosophical Lessons

I will now draw some more general conclusions from this case study. The main point I would like to make is that theories, models, and EFTs are indispensible tools in scientific research. They complement each other in a way which will be analysed in some more detail in Section 4.1. Sections 4.2 and 4.3 focus on the consequences for the reductionism debate and the controversy about scientific explanation.

4.1. Pluralism and coherence

Generalising the results of the case study, the following picture of theorising in physics emerges. Scientists use a variety of theoretical tools; among these tools are theories, models and EFTs. All of them have specific functions, and all of these functions are required.

This has consequences for the notorious debate about the final theory. Setting aside worries that we will anyway never be in the position to write down this theory, a final theory faces at least three problems. First, it will be 'too far away' from the phenomena we experience directly or in a laboratory. A final theory is therefore unlikely to provide a local understanding of these phenomena. Models and EFTs are still needed for this. Second, the final theory cannot be applied without various additional (model-) assumptions about the concrete system under investigation. Besides, low energy expansions have to be carried out in a systematic way. This is where EFTs will come in handy. And third, we will probably never arrive at a final theory without a supporting scaffolding of various models and EFTs. This has been the case for all theories of physics so far. So there is no reason to believe that things will be different for a final theory.

Let us now consider whether models are likely to dominate theoretical science as some philosophers of science think (see Cartwright, 1999). In our case study, none of the models we looked at was taken seriously by scientists if there was not at least a qualitative story which connected the model to a more fundamental theory. And even before the formulation of QCD, physicists did not consider the plurality of nuclear and hadron models to be a satisfactory state of affairs. Theories are needed to inspire the development of models and to present a framework for the various models.

EFTs are also not likely to make the other two tools obsolete. Although EFTs share many of their functions with theories and models, they will not be able to fulfil all of these functions. To illustrate this point, let us distinguish two cases. First, there is a fundamental theory from which specific EFTs can be deduced. Then, of course, the EFT depends on the fundamental theory which is then still an essential part of our theorical account of the world. Second, there is no fundamental theory. In this case the 'recipe' which is part of the bottom-up EFT ideology is to try and construct a tower of EFTs. More and more new particles will be added to the theory, all of them being 'elementary' in a certain sense, and all of them might couple to all the other particles known so far. But the resulting theory will not be of much value; it is simply too complicated, and its predictive power will go down just as the predictive power of the Ptolemaic system went down when more and more epicycles were added (see Forster, 2000). Instead, theorists will search for a more fundamental theory which will reduce the contingency that goes along with the tower-construction strategy.

The upshot of all this is that theories, models and EFTs are each indispensable tools of theoretical research in physics. But how are they

related? Obviously there are logical contradictions between various models, and also between QCD and EFTs. Nevertheless, they all hang together in some sense. Sometimes there is an approximate deductive relation between, for example, QCD with an EFT at a given energy scale. In the framework of Chiral Perturbation Theory, both are related through an *ontological bridge* (Rohrlich, 2000), i.e. the identification of the terms involving quark degrees of freedom in the Lagrangian density of QCD with the bosonic field $U(x)$. But the connections are not always so tight. Sometimes there is only a plausible story which relates the vocabularies of a model and QCD and which sets the model in the bigger framework provided by a theory. This story can be interpreted as a *semantic bridge* between the model and the theory. There is a whole continuum of relations between theories, models and EFTs which range from strict reductive relations through ontological bridges to rather vague associations through semantical bridges. It is these bridges which integrate all these approaches in a coherent whole (note that coherence comes in degrees!). It is tempting to reconstruct this picture in terms of the probabilistic model of the coherence of a belief set suggested in Bovens and Hartmann (2000) and developed in Bovens and Hartmann (forthcoming). The various bridges between a model M and a theory T, for example, could then be modelled by the conditional probability $P(M|T)$.[21] While this discussion refers to the statics of scientific theorising (how do the theories and models at a given time hang together?), a discussion of the role of considerations concerning coherence in the dynamics of successive scientific theories can be found in Hartmann (forthcoming).

4.2. Reduction and emergence

The issue of reduction is probably the one which so far has got most attention in the philosophical literature about EFTs. Much of this debate relates to the radical conclusions Cao and Schweber draw in their article from 1993. The authors claim to have 'found that the recent developments support a pluralism in theoretical ontology, an antifoundationalism in epistemology and an antireductionism in methodology' (p. 69). I will evaluate these claims on the basis of the case study presented in the last section.

First, *ontological pluralism*. This thesis is based on the observation that empirical reality seems to be organised in a multitude (infinity?) of quasi-autonomous layers. Each layer has its own ontology, and this ontology is to a considerable extent independent of the physics at higher energies. Only when the probing energy approaches the cut-off of a given layer, do effects of the higher layers turn out to have some influence. This influence might lead to a renormalisation of the mass and charge parameters, but it might also lead to the insight that the ontology used so far is not really fundamental. Nuclei

[21] Given the prior probabilities for M and T, $P(M)$ and $P(T)$, $P(T|M)$ can be obtained by Bayes' Theorem.

turned out to be composed of nucleons and pions, and nucleons and pions turned out to be composed of quarks and gluons. Cao and Schweber's claim of ontological pluralism rests on the assumptions, as Robinson (1992) has pointed out, 'that we can build up our ontological commitments in QFT only by the method of first identifying the referring terms of the theory we accept. But to accept this form of realism—theoretical realism—is to deny that we can build up our ontological commitments through experiment in the absence of theory' (p. 403). I would not go as far as Robinson and subscribe to some variant of entity realism. Rather, I would point to the role which other theories and their interrelations play when it comes to establishing the ontology of the world. And given these other theories and their interrelations it seems clear that ontological pluralism can at best be defended as a pragmatic thesis.[22] We are not trapped in the language game of one theory. Given these other theories it might be interesting to take EFTs as a case study in emergence. How do the properties of nucleons, for example, emerge from the complicated interplay of quarks and gluons? This is a physics problem, and the role of symmetry breaking mechanisms has to be addressed in detail. It should also be noted that Cao and Schweber's talk of quasi-autonomous domains rests on the validity of the decoupling theorem. As I have explained in Section 2.2.3, this theorem can only be proven if there is a underlying renormalisable theory and if the energy scales of the particles separate. While the second assumption might not be fulfilled empirically, the first assumption renders many of Cao and Schweber's more radical conclusions implausible since they are based on the assumption that there is no underlying theory.

Second, *epistemological antifoundationalism*. According to this thesis, quasi-autonomous layers do not only have their own ontology, but also have their own 'fundamental' theory. Since none of the layers is distinguished, none of the theories is *the* fundamental one. There is no foundation for all other theories, and there is no (and there will never be a) final theory which entails all other theories. Cao and Schweber know that this is a metaphysical thesis, but they think that it is well supported by the practice of science. And indeed, a final theory seems to be as far away as it was twenty years ago. But this, of course, does not imply that there will never be a final theory. Leaving metaphysical questions aside, it seems to be philosophically more interesting to examine the formal relations between the theories, models and EFTs we have already. As I argued in Section 4.1, the relation between some underlying theory and an EFT can be reconstructed along the lines of Rohrlich's (2000) two-step model: first, a deductive relation between the two formalisms has to be obtained. This typically involves approximations and a limiting process. Second, ontological bridges between the incommensurable concepts of both theories have to be established. This model, designed to account for the relation between *mature* theories, does not help to analyse the relation between a model and a theory, which is typically much more involved. This calls for detailed case studies.

[22] For another account of ontological pluralism see Rohrlich (1988).

Third, *methodological antireductionism*. This position advocates the bottom-up EFT research strategy which is also favoured by many pragmatically-minded physicists. Howard Georgi, for example, writes:

> The philosophical question underlying old-fashioned renormalizability is this: How does this process end? It is possible, I suppose, that at some very large energy scale, all nonrenormalizable interactions disappear, and the theory is simply renormalizable in the old sense. This seems unlikely, given the difficulty with gravity. It is possible that the rules change dramatically, as in string theory. It may even be possible that there is no end, simply more and more scales as one goes to higher and higher energy. Who knows? Who cares? In addition to being a great convenience, effective field theory allows us to ask all the really scientific questions that we want to ask without committing ourselves to a picture of what happens at arbitrarily high energy (Georgi, 1993, p. 215).

Georgi recommends the bottom-up EFT strategy for pragmatic reasons. Unlike a final theory, EFTs can be systematically tested experimentally, and this is taken to be a feature any acceptable scientific theory should have:

> My personal suspicion is that Nature is much more imaginative than we are. If we theorists approach her study with the proper respect, if we recognize that we *are* parasites who must live on the hard work of our experimental friends, then our field will remain healthy and prosper. But if we allow ourselves to be beguiled by the siren call of the 'ultimate' unification at distances so small that our experimental friends cannot help us, then we are in trouble, because we will lose that crucial process of pruning of irrelevant ideas which distinguishes physics from so many other less interesting human activities (Georgi, 1989, p. 457).

Georgi does not commit himself to a view concerning the possible existence of a final theory. Whether it exists or not is not a question which can be settled in the laboratory. Maybe the tower of EFTs never ends. It would, however, be a mistake to stop taking the possibility of a final theory into account. Michael Redhead also argues this point:

> [F]rom a point of view of methodology of science a recurring theme has been the search for an *ultimate* underlying order characterized by simplicity and symmetry that lies behind and explains the confusing complexity of the phenomenal world. To subscribe to the new EFT programme is to give up on this endeavour and retreat to a position that is admittedly more cautious and pragmatic and closer to experimental practice, but is somehow less intellectually exciting. Perhaps one should not allow such considerations to enter into one's evaluation of a scientific programme, but to my own taste, the regulative ideal of an ultimate theory of everything remains a powerful aesthetic ingredient motivating the existence of the greatest intellectual ingenuity in the pursuit of what may admittedly, in itself, be an illusory goal. But that after all is the proper function of regulative ideals (Redhead, 1999, p. 40)

4.3. Explanation and understanding

One of the major aims of science is to explain phenomena. Although the concept of explanation is pretty vague, an acceptable explanation should show (1) how the phenomenon under consideration reached its present state and (2) how it fits into a larger theoretical framework. Although these two requirements do not exclude each other, it remains to be seen if both can be fulfilled by the same scientific theory or model. This is not clear to start with, and philosophical theories of explanation therefore usually concentrate on one of these requirements—a task which turns out to be hard enough as the controversial nature of the debate over the last four decades or so impressively shows (Salmon, 1989).

According to the causal/mechanical account, pioneered by Salmon and others (see Salmon, 1998), a phenomenon is explained by providing a mechanism which produces the effect under consideration. This mechanism is often given by a model (or an EFT for that matter), as the MIT-Bag Model illustrates. Here quarks are confined to a hard sphere in which they can move freely apart from occasional bounces off the inner side of the bag, a situation which can be easily visualised classically (see Section 3.1.2). The mass of the proton, for example, can then be determined by summing up the kinetic energies of the quarks and the potential energy of the bag. Explanations of this kind produce local understanding, but lack global understanding because no general principles are required to specify the mechanism.

According to the unification account, developed by Friedman and elaborated by Kitcher (1989), a successful explanation fits the explanandum in a general framework. This view, which is a distant descendant of the original Hempel–Oppenheim account, supports the intuition that something is explained if it is integrated in a larger theoretical context. Explanations of this kind are provided by theories such as QCD. An explanation of the mass of the proton, for example, goes like this: there are quarks and gluons coupled to a state with the quantum numbers of the proton and interacting in a very complicated way. Deductions from the Lagrangian density of QCD, facilitated by high-powered computers, then yield the result of 938 MeV. As I have argued in Section 3.1.1, this account does not produce local understanding. By integrating the proton in a bigger framework it produces, however, global understanding.

I think that the question which of the two accounts of explanation is the right one is misguided. The case study presented in Section 3 has shown that both accounts complement and interact with each other. The account of scientific explanation I consider to be in accordance with scientific practice is therefore a pluralist one: science studies a given phenomenon from various theoretical perspectives, all of which reveal some explanatory information about the phenomenon in question. Putting all of them together should result in a coherent explanatory account of the phenomenon. It should be noted that Salmon (1998, pp. 73f) also acknowledges a rapprochement of the causal/

mechanical account and the unification account of scientific explanation in his recent book.

5. Conclusions

Science is a complex and involved activity. All simple reconstructions of it will probably fail. Generalisations based on the work on theory unification in particle physics, for example, are as hasty as the philosophical conclusions some now draw from the current interest in EFTs among physicists. Science usually does not address issues such as unity, reductionism, and what *the* characteristics of a good explanation are. Rather, scientists use a plurality of interrelated conceptual tools, and explanations are obtained by attacking a phenomenon from a variety of theoretical perspectives. It is this pluralism of tools which is good for science and which makes science flourish.

Acknowledgements—This article has a long history. Prior versions of it were presented at the Centre for Philosophy of Natural and Social Sciences at the London School of Economics in 1994, the '2nd Congress of the German Society for Analytical Philosophy' in Leipzig, Germany (1994), and at the 'Congress for Logic, Methodology and Philosophy of Science' in Florence, Italy (1995). Thanks to the audiences, especially Jeremy Butterfield, Nancy Cartwright, Margaret Morrison, Michael Redhead and Manfred Stöckler for many suggestions and stimulating discussions. Luc Bovens, Alfred Nordmann, Fritz Rohrlich, and the editors commented on a draft of this article and suggested various stylistic and substantial improvements. Thanks also to Jim Lennox and the Center for Philosophy of Science at the University of Pittsburgh in whose stimulating atmosphere this article was completed, and the Alexander von Humboldt Foundation (Bonn, Germany) for financial support.

References

Appelquist, T. and Carazzone, J. (1975) 'Infrared Singularities and Massive Fields', *Physical Review* **D11,** 2856–2861.

Becker, W., McIver, J. and Schlicher, R. (1989) 'Testing the Photon-Photon-Sector of Quantum Electrodynamics With Free-Electron Lasers', *Journal of the Optical Society of America* **6,** 1083–1092.

Bovens, L. and Hartmann, S. (2000) 'Coherence, Belief Expansion and Bayesian Networks', in C. Baral and M. Truszczynski (eds), *Proceedings of the 8th International Workshop on Non-Monotonic Reasoning, NMR 2000* (http://www.cs.engr.uky.edu/nmr2000/proceedings.html).

Bovens, L. and Hartmann, S. (forthcoming) 'Solving the Riddle of Coherence', Preprint Boulder and Konstanz.

Brown, L (ed.) (1993) *Renormalization: From Lorentz to Landau (and Beyond)* (New York: Springer).

Bunge, M. (1964) 'Phenomenological Theories', in M. Bunge (ed.), *The Critical Approach to Science and Philosophy. In Honor of Karl R. Popper* (London: Collier-Macmillan), pp. 234–254.

Cao, T. Y. (1993) 'New Philosophy of Renormalization: From the Renormalization Group Equations to Effective Field Theories', in Brown (ed.) (1993), pp. 87–134.

Cao, T. Y. (1997) *Conceptual Developments of 20th Century Field Theories* (Cambridge: Cambridge University Press).
Cao, T. Y. (ed.) (1999) *Conceptual Foundations of Quantum Field Theory* (Cambridge: Cambridge University Press).
Cao, T. Y. and Schweber, S. (1993) 'The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory', *Synthese* **97,** 33–108.
Cassidy, D. (1995) *Werner Heisenberg. Leben und Werk* (Heidelberg: Spektrum Akademischer Verlag).
Cat, J. (1998) 'The Physicist's Debate on Unification in Physics at the End of the 20th Century', *Historical Studies in the Physical and Biological Sciences* **28,** 253–299.
Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science* (Cambridge: Cambridge University Press).
Dirac, P. (1983) 'The Origin of Quantum Field Theory', in L. Brown and L. Hoddeson (eds), *The Birth of Particle Physics* (Cambridge: Cambridge University Press), pp. 39–55.
Donoghue, J., Golowich, E. and Holstein, B. (1992) *Dynamics of the Standard Model* (Cambridge: Cambridge University Press).
Donoghue, J. (1994a) 'Leading Quantum Corrections to the Newtonian Potential', *Physical Review Letters* **72,** 2996–2999.
Donoghue, J. (1994b) 'General Relativity as an Effective Field Theory: The Leading Quantum Corrections', *Physical Review* **D59,** 3874–3888.
Ecker, G. (1995) 'What is a Low-Energy-Theorem?', *Comments on Nuclear and Particle Physics* **21,** 347–367.
Euler, H. (1936) 'Über die Streuung von Licht an Licht nach der Diracschen Theorie', *Annalen der Physik* **26,** 398–448.
Fermi, E. (1933) 'Tentativo di una teoria del l'emissione dei raggi β', *Ric. Sci.* **4,** 491–495.
Fermi, E. (1934) 'Versuch einer Theorie der β-Strahlen I', *Zeitschrift für Physik* **88,** 161–171.
Fischer, M. (1999) 'Renormalization Group Theory: Its Basis and Formulation in Statistical Physics', in Cao (ed.) (1999), pp. 89–134.
Forster, M. (2000) 'Key Concepts in Model Selection: Performance and Generality', *Journal of Mathematical Psychology* **44,** 205–231.
Franklin, A. (1990) *Experiment, Right or Wrong* (Cambridge: Cambridge University Press).
Georgi, H. (1989) 'Effective Quantum Field Theories', in P. Davies (ed.), *The New Physics* (Cambridge: Cambridge University Press), pp. 446–457.
Georgi, H. (1993) 'Effective Field Theory', *Annual Review of Nuclear and Particle Science* **43,** 209–252.
Hartmann, S. (1995a) *Metaphysik und Methode. Strategien der zeitgenössischen Physik in wissenschaftsphilosophischer Perspektive* (Konstanz: Hartung-Gorre Verlag).
Hartmann, S. (1995b) 'Models as a Tool for Theory Construction. Some Strategies of Preliminary Physics', in W. Herfel *et al.* (eds), *Theories and Models in Scientific Processes* (Amsterdam: Rodopi), pp. 49–67.
Hartmann, S. (1999) 'Models and Stories in Hadron Physics', in M. Morgan and M. Morrison (eds), *Models as Mediators* (Cambridge: Cambridge University Press), pp. 326–346.
Hartmann, S. (forthcoming) 'Review of S. French *et al.*: *Correspondence, Invariance and Heuristics. Essays in Honour of Heinz Post'*, to appear in *Studies in History and Philosophy of Modern Physics*.
Heisenberg, W. (1966) 'Die Rolle der phänomenologischen Theorien im System der theoretischen Physik', in A. De-Shalit, H. Feshbach and L. Van Hove (eds), *Preludes*

in Theoretical Physics. In Honor of V.F. Weisskopf (Amsterdam: North-Holland), pp. 166–169.
Heisenberg, W. and Euler, H. (1936) 'Folgerungen aus der Diracschen Theorie des Positrons', *Zeitschrift für Physik* **98,** 714–732.
Huggett, N. and Weingard, R. (1995) 'The Renormalization Group and Effective Field Theories', *Synthese* **102,** 171–194.
Humphreys, P. (1994) 'Numerical Experimentation', in P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher*, Vol. 2 (Dordrecht: Kluwer Academic Publishers), pp. 103–121.
Humphreys, P. (1995) 'Computational Science and Scientific Method', *Minds and Machines* **5,** 499–512.
Itzykson, C. and Zuber, J. B. (1980) *Quantum Field Theory* (New York: McGraw-Hill).
Jauch, J. and Rohrlich, F. (1976) *The Theory of Photons and Electrons* (Berlin: Springer).
Kitcher, P. (1989) 'Explanatory Unification and the Causal Structure of the World', in P. Kitcher and W. Salmon (eds), *Scientific Explanation* (Minneapolis: University of Minnesota Press), pp. 410–505.
Kolck, U. van (1999) 'Effective Field Theory of Nuclear Forces', *Progress in Particle and Nuclear Physics* **43,** 337–418.
Kragh, H. (1990) *Dirac. A Scientific Biography* (Cambridge: Cambridge University Press).
Lepage, P. (1989) 'What is Renormalization?', in T. DeGrand and D. Toussaint (eds), *From Actions to Answers, Proceedings of the 1989 Theoretical Study Institute in Elementary Particle Physics* (Singapore: World Scientific), pp. 483–509.
Lepage, P. (1994) 'Lattice QCD for Small Computers', lectures presented at TASI 93 (Boulder, June 1993), and at the UK Summer Institute (St. Andrews, August 1993), hep-lat/9403018.
Leutwyler, H. (1994) 'Foundations and Scope of Chiral Perturbation Theory', talk given at the conference 'Chiral Dynamics: Theory and Experiment' (MIT, July 1994), Preprint BUTP-94/18.
Meissner, U.-G. (ed.) (1992) *Effective Field Theories of the Standard Model* (Singapore: World Scientific).
Migdal, A. (1977) *Qualitative Methods in Quantum Theory* (London: Benjamin).
Mills, R. (1993) 'Tutorial on Infinities in QED', in Brown (ed.) (1993), pp. 59–86.
Montvay, I. and Münster, G. (1994) *Quantum Fields on a Lattice* (Cambridge: Cambridge University Press).
Mosel, U. (1999) *Fields, Symmetries, and Quarks* (Springer: Berlin).
Redhead, M. (1999) 'Quantum Field Theory and the Philosopher', in Cao (ed.) (1999), pp. 34–40.
Robinson, D. (1992) 'Renormalization and the Effective Field Theory Programme', in D. Hull, M. Forbes, and K. Okruhlik (eds), *PSA 1992, Vol. 1* (East Lansing: The Philosophy of Science Association), pp. 393–403.
Rohrlich, F. (1988) 'Pluralistic Ontology and Theory Reduction in Physical Sciences', *British Journal for the Philosophy of Science* **39,** 295–312.
Rohrlich, F. (2000) 'Cognitive Scientific Realism', *Philosophy of Science* **67** (forthcoming).
Salmon, W. (1989) *Four Decades of Scientific Explanation* (Minneapolis: University of Minnesota Press). Reprinted in P. Kitcher and W. Salmon (eds) (1989), *Scientific Explanation* (Minneapolis: University of Minnesota Press), pp. 3–219.
Salmon, W. (1998) *Causality and Explanation* (Oxford: Oxford University Press).
Schweber, S. (1993a) 'Changing Concepualization of Renormalization Theory', in Brown (ed.) (1993), pp. 135–166.

Schweber, S. (1993b) 'Physics, Community and the Crisis in Physical Theory', *Physics Today* **46,** 34–40.
Schweber, S. (1994) *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton: Princeton University Press).
Schweber, S. (1995) 'Physics, Community and the Crisis in Physical Theory', in K. Gavroglu (ed.), *Physics, Philosophy and the Scientific Community* (Dordrecht: Kluwer), pp. 125–152.
Schwinger, J. (1973) *Particles, Sources, and Fields* (Reading, MA: Addison-Wesley).
Shankar, R. (1999) 'Effective Field Theory in Condensed Matter Physics', in Cao (ed.) (1999), pp. 47–55.
Weinberg, S. (1979) 'Phenomenological Lagrangians', *Physica* **96A,** 327–340.
Weinberg, S. (1980a) 'Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions', *Reviews of Modern Physics* **52,** 515–523.
Weinberg, S. (1980b) 'Effective Gauge Theories', *Physics Letters* **B91,** 51–55.
Weinberg, S. (1987) 'Newtonianism, Reductionism, and the Art of Congressinal Testimony', *Nature* **330,** 433–437.
Weinberg, S. (1993) *Dreams of a Final Theory* (London: Hutchinson Radius).
Wightman, A. (1986) 'Some Lessons of Renormalization Theory', in J. de Boer, E. Dal and O. Ulfbeck (eds), *The Lessons of Quantum Theory* (Amsterdam, Elsevier), pp. 201–226.

R. JACKIW

EFFECTS OF DIRAC'S NEGATIVE ENERGY SEA ON QUANTUM NUMBERS

R. Jackiw

Center for Theoretical Physics
Laboratory for Nuclear Science
and Department of Physics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139 U.S.A.

Dirac Prize Lecture
Trieste, Italy – March 1999

The Dirac Prize Committee cited my work on fractional charges and on chiral anomalies; therefore, I shall discuss these two topics. As with all true and deep physical effects, there are many ways of arriving at the results. It is particularly appropriate here today that one route towards understanding both fractional charges and chiral anomalies delves into Dirac's negative energy sea. This is especially provocative, because usually we think of Dirac's negative energy sea as an unphysical construct, invented to render quantum field theory physically acceptable by hiding – that is, by renormalizing – negative energy solutions. But I am suggesting that in fact physical consequences can be drawn from Dirac's construction.

First let me set the stage for the discussion.

Quantum physics has taught us that a physical observable need not be a quantity with arbitrary magnitude. Because it is the eigenvalue of a linear Hermitian operator, it will in general be quantized. This is not the case in classical physics where most observables, like angular momentum and energy, are continuously varying and can take on any value. On the other hand, even in classical physics, there are concepts that are intrinsically integral – for example particle number of conserved particles – and one expects that the integrality will be preserved in the quantum theory, that is, eigenvalues of the relevant operator – the number operator in our example – are expected to be integers. Quantization of eigenvalues is most easily attained when the operator is a generator of a compact, non-Abelian group, like angular momentum generating SO(3) rotations. However, the number operator frequently generates only Abelian transformations with no group-theoretic quantization.

Closer examination of the number operator in a theory with second quantized fermions raises doubts that it will in fact possess only integer eigenvalues. The problem derives from Dirac's negative energy sea, which must be filled to define the vacuum. This involves an infinite number of "particles". Since the number of any further particles must be measured relative to this infinity, there may very well emerge a non-integral answer. Nevertheless, it had been believed that various renormalization procedures, like normal ordering, can unambiguously ensure integrality of the eigenvalue. Therefore, it was a suprise when Rebbi and I established about twenty years ago that fermions moving in the field of a topologically nontrivial soliton (kink in one spatial dimension, vortex in two, monopole in three) possess non-integer eigenvalues for their number operator.[1] It is perhaps even more surprising that this peculiar effect has a physical realization in properties of actual condensed matter systems – polyacetylene being the standard example.[2] Here, I shall describe this to you, first in a general, formal way, and then in a physically intuitive language appropriate to polyacetylene.

We wish to second quantize fermions moving in a static background that generically is described by φ. Fermion dynamics is governed by a Dirac Hamiltonian $H(\varphi)$. Two different backgrounds are envisioned: one is appropriate for the vac-

uum sector of the theory φ_v, the other for the soliton φ_s. For example, φ may be a condensate field that takes a homogenous value in the vacuum sector and a topologically nontrivial profile in the soliton sector.

Second quantization proceeds by computing the energy eigenvalues and eigenfunctions of $H(\varphi)$, which possesses both positive and negative energy eigenstates, and "filling" the negative energy sea. The number density of the soliton ground state is given by

$$\rho(\mathbf{x}) = \sum\!\!\!\!\!\!\int_{-\infty}^{0^-} dE \, \left(|\Psi_E(\mathbf{x})|^2 - |\psi_E(\mathbf{x})|^2 \right) \tag{1}$$

where the integration, which also includes summutation over discrete levels, extends over all negative energy states, since they are filled in the vacuum. Here Ψ_E is the energy eigenfunction in the presence of the soliton and ψ_E is the eigenfunction in the vacuum sector:

$$H(\varphi_s)\Psi_E = E\Psi_E, \quad H(\varphi_v)\psi_E = E\psi_E \quad . \tag{2}$$

In (1) the contribution from the vacuum sector is subtracted; the soliton charge density is renormalized, so that it is measured relative to the vacuum. The fermion number of the soliton ground state is the spatial integral of ρ

$$N_F = \int d\mathbf{x} \, \rho(\mathbf{x}) \quad . \tag{3}$$

A very beautiful aspect of the theory is that one can evaluate (1) and (3) by general methods, which bypass solving the eigenvalue problem (2) explicitly. Rather, one uses spectral sum rules whose form is dictated by general features of the Hamiltonian, in particular by the topological properties of the background φ and of the space $\{\mathbf{x}\}$. While these methods are powerful, they are also technical, requiring much mathematical knowledge, so I shall not present them here. However, when the Hamiltonian posseses one further property, which I shall now describe, the sum rules become trivial, and the result for N_F is immediate.

We assume further that $H(\varphi)$ possesses a conjugation symmetry, taking positive energy states into negative energy states and *vice versa*; that is, we assume there exists an operator $\mathcal{C}$ that anticommutes with $H(\varphi)$: $\mathcal{C}^{-1}H\mathcal{C} = -H$. One consequence of this is that the number density at E is an even function of E: $|\Psi_E|^2 = |\Psi_{-E}|^2, \quad |\psi_E|^2 = |\psi_{-E}|^2$. A less obvious consequence is that in the soliton sector, there are always normalizable, discrete zero-energy modes

$$H(\varphi_s)u_0 = 0, \quad \int d\mathbf{x} \, |u_0(\mathbf{x})|^2 = 1 \quad . \tag{4}$$

This fact may be seen by explicit solution of the eigenvalue problem, but it also follows from a general mathematical argument, called *index theory*, which ensures that the Dirac operator has normalizable zero-energy modes in the presence of a topologically nontrivial background.

We are now in a position to evaluate (1) and (3). First, we use completeness of the eigenfunctions in the soliton and vacuum sectors

$$\sum\!\!\!\!\!\!\int_{-\infty}^{0^-} dE\,|\Psi_E(\mathbf{x})|^2 + \sum\!\!\!\!\!\!\int_{0^+}^{\infty} dE\,|\Psi_E(\mathbf{x})|^2 + |u_0(\mathbf{x})|^2 - \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} dE\,|\psi_E(\mathbf{x})|^2 = 0 \quad . \tag{5}$$

The zero-energy mode in the soliton sector has been explicitly separated; we assume there is just one. (In the vacuum sector there are none.) Then, use of the conjugation symmetry allows equating the positive energy integrals with the negative energy ones, and converts (5) into an evaluation of (1)

$$\sum\!\!\!\!\!\!\int_{-\infty}^{0^-} dE\,\left(|\Psi_E(\mathbf{x})|^2 - |\psi_E(\mathbf{x})|^2\right) = -\tfrac{1}{2}|u_0(\mathbf{x})|^2 \quad . \tag{6}$$

The spatial integration that determines N_F is trivial since the zero mode is normalized:

$$N_F = -\tfrac{1}{2} \quad . \tag{7}$$

The conclusion is that the soliton vacuum, defined with the zero mode empty, carries fermion number $-\frac{1}{2}$; of course, when the zero mode is filled, this fermion number is $+\frac{1}{2}$! The fermion number assignment of $\pm\frac{1}{2}$ for two states degenerate in energy is the only possible one consistent with a conjugation-odd fermion number operator.

Two comments should be made in connection with this very elementary derivation of our surprising result.

(i) The above evaluation concerns the *expected* value of the second-quantized, field theoretic number operator, $\hat{N}_F$. However, one can show, by expanding the second quantized field in terms of creation and annihilation operators in the presence of the soliton, that in fact the *eigenvalues* are $\pm\frac{1}{2}$.

(ii) We have viewed the soliton as an external field. In a complete description, one must take the soliton's quantum dynamics into account. Necessarily there will occur spontaneous symmetry breaking in the vacuum sector. Calculations in the full theory can be carried out by Monte-Carlo methods, or by analytic techniques of the Born-Oppenheimer variety.

The three ingredients necessary for fermion number fractionization – spontaneous symmetry breaking, solitons and fermions – come together in a description of a physical system, polyacetylene. This is a one-dimensional array of carbon atoms that can form one of two degenerate ground states. The degeneracy arises from a spontaneous breaking of the right-left symmetry (Peierls instability) and manifests itself in an alteration of the bonding pattern, as illustrated in Fig. 1.

A microscopic Hamiltonian for the system has been proposed by Su, Schrieffer and Heeger (SSH).[2] In the continuum and infinite volume limit, electron transport is governed by a Dirac-type Hamiltonian in one spatial dimension:

$$\begin{aligned} H(\varphi) &= \sigma_3\hat{p} + \sigma_1\varphi(x), \quad \hat{p} = \frac{1}{i}\,\frac{d}{dx} \\ \sigma_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad . \end{aligned} \tag{8}$$

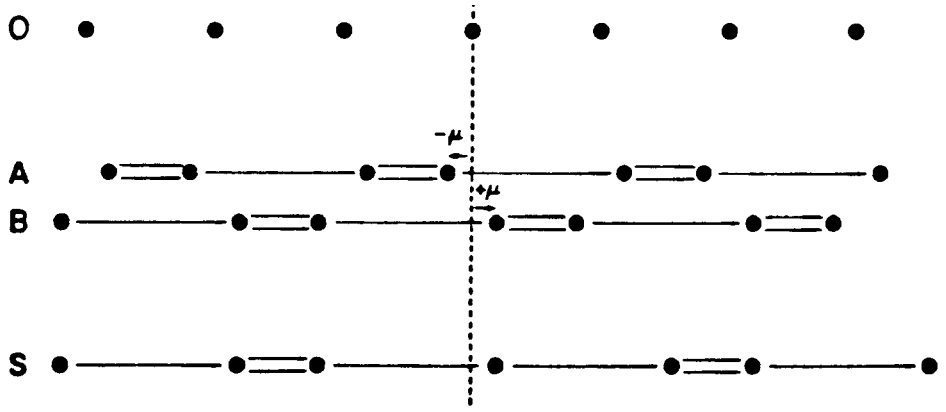


Fig. 1: Polyacetylene consists of a linear chain of carbon atoms (dots). The equally spaced configuration (O) possesses a left-right symmetry, which however is energetically unstable. Rather in the ground states the carbon atoms shift a distance μ to the left or right, breaking the symmetry and producing two degenerate ground states (A,B). (The drawing is not to scale; the shift is only a few percent of the total bond length.) A soliton (S) is a defect in the alteration patters; it provides a domain wall between configurations (A) and (B).

Here, $\varphi(x)$ is the phonon field; it measures the displacement of the carbon atom from its equalibrium position. The matrix structure of the above Hamiltonian does not arise from spin. In the SSH description, electron-electron interactions are ignored and spin is a passive label; the Hamiltonian in (8) describes separately spin up and spin down electrons. Rather, the Dirac-like matrix form for H arises through a linearized approximation and the two-component wavefunctions that are eigenmodes of H refer to the right-moving and left-moving electrons with momentum $\pm|p|$. The filled negative energy states are the valence electrons, while the conduction electrons populate the positive energy states.

In the SSH model, $\varphi(x)$ is determined self-consistently by the phonon's dynamics, and in the lowest (vacuum) states $\varphi(x)$ takes the uniform values $\pm\mu$, as illustrated in Fig. 1. The corresponding spectrum of (8) exhibits a gap.

In addition to the two ground states, where the phonon field takes a constant value, there exist stable excited states where $\varphi(x)$ assumes a kink shape, which interpolates as x passes from $-\infty$ to $+\infty$, between the vacuum configurations $-\mu$ and $+\mu$. This is the soliton, and it describes a defect in the alteration pattern, as is also exhibited in Fig. 1.

The Hamiltonian in (8) admits a conjugation symmetry: $\mathcal{C} = \sigma_2 = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$ anticommutes with H – this is ordinary charge conjugation invariance in the absence of Coulomb interactions. The zero eigenvalue problem is easily solved with a kink background: there is one normalizable zero-energy solution. Thus, our general analysis predicts that fermion number, here coinciding with charge, fractionizes to $\pm\frac{1}{2}$ in the one-soliton state.

This result may also be seen pictorially. When two solitons are inserted into the ground state (B), the bonding pattern is depicted in Fig. 2. Note that the number of bonds in the two-soliton state is one fewer than in the ground state. If the two solitons are now separated far apart, so that they act independently, the quantum numbers of the missing bond must be shared between the two states, and that is how the fraction $\frac{1}{2}$ arises.

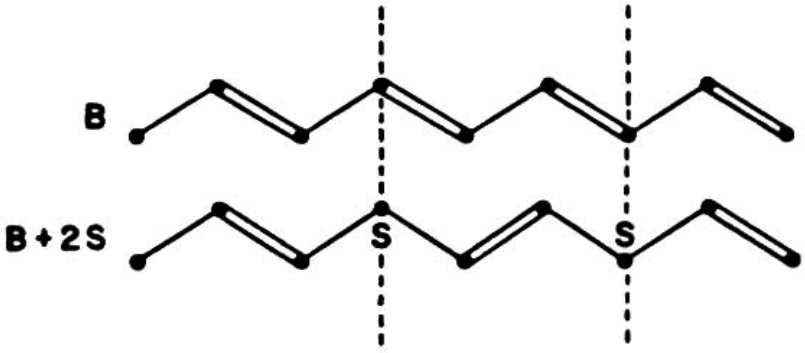


Fig. 2: With two solitons (SS) inserted in vacuum (B), the number of bonds $\approx$ electrons between the sites of the defects decreases from five to four.

Of course, for actual physical samples, where both the volume and the separation between defects are finite, the non-integer fermion number is only an expectation value for the operator $\hat{N}_F$, and there are corrections that vanish in the infinite limit. The important point is that the variance $\langle \hat{N}_F^2 \rangle - \langle \hat{N}_F \rangle^2$ also vanishes in the limit. This is to be contrasted to the uninteresting situation of, say, an electron circulating about two nuclei. When the nuclei are far apart, the expected value for electron number near each nucleus is $\frac{1}{2}$, plus small corrections. However, the variance remains $\frac{1}{4}$ for infinite separation, which shows that this fraction never becomes an eigenvalue.

The concept of fractional quantum numbers has now extended beyond soliton systems – for example the theory of the quantum Hall effect makes use of the idea, even though dynamical details are quite different from the above example.

While fractional quantum numbers were first seen in relativistic field theory, thus far they have not played any experimentally verified role in particle physics. Nevertheless, it is curious that an effect which in principle is physical, and has been observed in condensed matter systems, should arise from distortions in the negative energy sea, which for particle physicists is an unphysical construct, in contrast to the condensed matter application, where negative energy states correspond to physical quantities – the valence electrons.

In particle physics there is another, physically realized circumstance where the Dirac negative energy sea modifies symmetry behavior of fermions. This is the chiral anomaly phenomenon whereby the axial vector currect $i\bar{\psi}\gamma^\mu\gamma^5\psi$, which is conserved for free massless Dirac fermions, ceases to be conserved when the massless fermions are quantized and made to interact with a gauge field, even though the interaction appears to be chirally invariant.[3] Indeed, for unquantized Dirac fields

the chiral invariance ensures conservation of the unquantized axial vector current. This disappearance after quantization of chiral symmetry is usually associated with infinities that plague relativistic quantum field theory: the infinities must be regulated and renormalized, but there is no chirally invariant regulator procedure. However, a more directly physical discussion of the anomaly phenomenom may be given, which shows that in fact it is the filling of the negative energy sea that breaks the chiral symmetry.

Let me first state the essential puzzle of the chiral anomaly. In spite of profound differences between the classical and quantal description of physical phenomena, it was generally believed that symmetry properties and conserved constants of motion transcend the classical/quantal dichotomy: when a classical model possesses symmetries and supports constants of motion, one expects that quantization preserves the symmetries, so that conserved quantities – now propmoted to quantum operators – remain time-independent, that is, they commute with the Hamiltonian operator. But as observed thirty years ago by Bell and me,[3] and also independently by Adler,[3] this need not be so.

A simple instance of quantum mechanical violation of symmetry is encountered by considering a massless Dirac fermion moving in a background gauge field A_μ. The dynamics is governed by a Lagrangian, which splits into separate right and left parts – this separation is a manifestation of chiral symmetry:

$$\begin{aligned} \mathcal{L} &= \bar{\psi}(i\not{\partial} - e\not{A})\psi \\ &= \bar{\psi}_+(i\not{\partial} - e\not{A})\psi_+ + \bar{\psi}_-(i\not{\partial} - e\not{A})\psi_- \\ \psi &= \psi_+ + \psi_-, \quad \psi_\pm = \tfrac{1}{2}(1 \pm i\gamma_5)\psi \quad . \end{aligned} \tag{10}$$

In the first-quantized theory, where ψ is a wavefunction and $\bar{\psi}\gamma^\mu\psi$ is a probability current, the separate right and left currents are conserved, and the separate probabilities $\int d\mathbf{x}\,\psi_\pm^\dagger\psi_\pm$ are time-independent. In the second quantized theory, where ψ becomes an operator, the anomaly phenomenon renders the separate right and left currents no longer conserved, and the right and left charges are not time-independent. Nevertheless, the sum of right and left currents – the vector current – is conserved, while the divergence of the difference between the right and left currents – the axial vector current – is nonzero owing to the anomaly. Our task then is to understand what causes the separate nonconservation of left and right currents even though there is no coupling between the two apparent in (10).

Evidently, the problem derives from the second quantization procedure, hence we implement it. We set A^0 to zero, find the eigenmodes of the Hamiltonian in the background field $\mathbf{A}$, and define the second quantized vacuum by filling the negative energy modes, leaving the positive energy modes empty. The background $\mathbf{A}$ is chosen in a specific functional form so that the anomaly is nonvanishing. This requires $\mathbf{A}$ be time-dependent, but we chose a potential constant in time and space and model the time variation by an adiabatic change $\mathbf{A} \to \mathbf{A} + \delta\mathbf{A}$.

The simplest model to study is two-dimensional and Abelian – two-dimensional massles quantum electrodynamics.[4] The Dirac matrices are 2×2 and ψ is a two-

component spinor.

$$\gamma^0 = \sigma^1, \quad \gamma^1 = i\sigma^2, \quad \gamma_5 = i\gamma^0\gamma^1 = -i\sigma^3$$
$$\psi_+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\psi, \quad \psi_- = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\psi \ . \tag{11}$$

The axial current possess an anomalous divergence proportional to $\epsilon^{\mu\nu}F_{\mu\nu} \propto \partial_t A$ (since $A_0 = 0$ and $\mathbf{A}$ has the single spatial component $A^1 \equiv A$),[5] which we wish to understand. The eigenmodes to be second quantized satisfy a one-dimensional Dirac equation,

$$H\psi_E = \alpha(\hat{p} - eA)\psi_E = E\psi_E, \quad \alpha = -\sigma^3 \tag{12}$$

where A is constant. They are given by

$$\psi_+ = \begin{pmatrix} e^{ipx} \\ 0 \end{pmatrix} \text{ with } E = -p + eA$$
$$. \tag{13}$$
$$\psi_- = \begin{pmatrix} 0 \\ e^{ipx} \end{pmatrix} \text{ with } E = p - eA$$

Second quantization is performed by filling the negative energy sea. For $A = 0$, the energy-momentum dispersion is depicted in Fig. 3, where the right-hand branch corresponds to fermions of one chirality, and the left-hand branch to those of the other chirality. The negative energy states are filled, as indicated by the filled circles; the positive energy states are empty, as indicated by the empty circles. As A increases from 0 to δA, empty states in the right-hand branch acquire negative energy, while filled states of the left-hand branch become positive energy states; that is, there is a net production of right-handed antiparticles and left-handed particles; see Fig. 4. So the separate right and left charges are not conserved, but their sum is. Put in another way, the separation between positive and negative energy states of definite chirality cannot be achieved gauge-invariantly, since changing A from 0 to a constant δA is a gauge transformation, yet particles are produced.

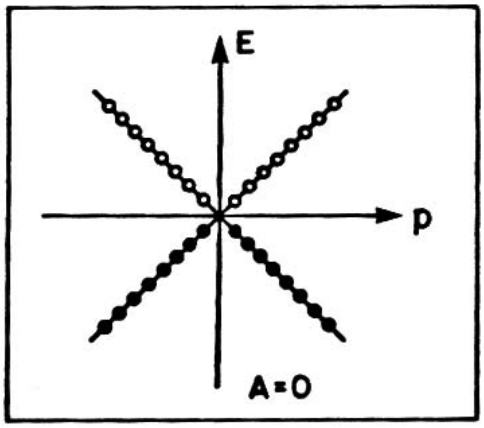


Fig. 3: Energy-momentum dispersion at $A = 0$. Empty circles are empty states; filled circles are filled states.

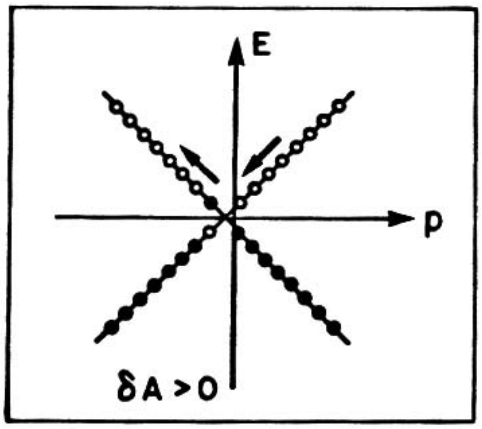


Fig. 4: Energy-momentum dispersion at $A = \delta A > 0$. The energy shift produces negative energy empty states and positive energy filled states.

We thus see that the negative energy sea is responsible for nonconservation of chirality even though the dynamics is chirally invariant. This effect was called *anomalous* because its discovery was a surprise.[3] However, a better name might be *quantum mechanical symmetry breaking* – a symmetry-breaking mechanism, which like Heisenberg's *spontaneous symmetry breaking* attributes physical asymmetry to the vacuum state and not to the dynamics. Here, however, unlike in Heisenberg's case, it is not vacuum degeneracy but the very definition of the vacuum that is responsible. Once again we must assign physical reality to Dirac's negative energy sea, because it produces the chiral anomaly, whose effects are experimentally observed, principally in the decay of the neutral pion to two photons, but there are other physical consequences as well.[6]

The two phenomena that I have described show that Nature seems to know and make use of what at first appears to be a defect of a quantum field theory, and which is usually "renormalized away". Remarkably, the infinities in the formalism give rise to finite and physical effects. One may quite appropriately call this an example of "The Unreasonable Effectiveness of Quantum Field Theory in Physics".

In my presentation of fractional charge and of the chiral anomaly I have used concepts that are primitively physical, employing little analysis beyond drawing pictures and counting. But to the same end a most sophisticated mathematical discussion can also be given, wherein zero modes are controlled by various index theorems (Atiyah-Singer, Callias), fractional charges are related to Atiyah, Patodi, Singer spectral flows, and chiral anomalies are identified with Chern-Pontryagin densities. This confluence between physics and mathematics, which was brought about by physicists' research on the topics that I described, seeded an interaction between these two disciplines, which still flourishes and today fuels the current string/M-theory program.

But in sharp contrast with the above, contemporary mathematical discussions within physics do not as yet have experimentally observed correlatives. These days research follows very closely Dirac's dictum:

> The most powerful method of advance [in physics] . . . is to employ all the resources of pure mathematics in attempts to perfect and generalize the mathematical formalism that forms the existing basis of theoretical physics, and . . . to try to interpret the new mathematical features in terms of physical entities. — *P.A.M. Dirac*

Early in my career I heard him say this at a seminar devoted to his research history, and I was very inspired. However, my subsequent work with physics and mathematics makes me feel that Dirac's suggestion is too radical, while a quote by Yang more accurately describes my own experience of mathematical physics:

> Physics is not mathematics, just as mathematics is not physics. Somehow nature chooses only a subset of the very beautiful and complex and intricate mathematics that mathematicians develop, and that precise subset is what the theoretical physicist is trying to look for. — *C.N. Yang*

REFERENCES

1. R. Jackiw and C. Rebbi, *Phys. Rev. D* **13**, 3398 (1976); for a review, see A. Niemi and G. Semenoff, *Phys. Reports* **135**, 99 (1986).

2. W.-P. Su, J.R. Schrieffer and A. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979) and *Phys. Rev. B* **22**, 2099 (1980).

3. J.S. Bell and R. Jackiw, *Nuovo Cim.* **60**, 47 (1969); S. Adler, *Phys. Rev.* **177**, 2426 (1969).

4. J. Schwinger, *Phys. Rev.* **128**, 2425 (1962).

5. K. Johnson, *Phys. Lett.* **5**, 253 (1963).

6. For extensive discussion of the anomaly phenomena see S. Treiman, R. Jackiw, B. Zumino and E. Witten, *Current Algebra and Anomalies*, (Princeton University Press/World Scientific, Princeton, NJ/Singapore, 1985) or S. Adler in *Lectures on Elementary Particles and Quantum Field Theory*, S. Deser, M. Grisaru, and H. Pendleton, eds. (MIT Press, Cambridge, MA, 1970).

TOPOLOGY IN PHYSICS*

R. JACKIW†
Center for Theoretical Physics
MIT, Cambridge, MA 02139-4307
E-mail: jackiw@lns.mit.edu

The phenomenon of quantum number fractionalization is explained. The relevance of non-trivial phonon field topology is emphasized.

1. Introduction

Discussions of the spatial forms of physical materials use in a natural way geometrical and topological concepts. It is to be expected that arrangements of matter should form patterns that are described by pre-existing mathematical structures drawn from geometry and topology. But theoretical physicists also deal with abstract entities, which do not have an actual material presence. Still geometrical and topological considerations are relevant to these ephemeral theoretical constructs. I have in mind fields, both classical and quantum, which enter into our theories of fundamental processes. These fields $\phi(x)$ provide a mapping from a "base" space or space-time on which they are defined into the field "target" manifold on which they range. The base and target spaces, as well as the mapping, may possess some non-trivial topological features, which affect the fixed time description and the temporal evolution of the fields, thereby influencing the physical reality that these fields describe. Quantum fields of a quantum field theory are operator valued distributions whose relevant topological properties are obscure. Nevertheless, topological features of the corresponding classical fields are important in the quantum theory for a variety of reasons: (i) Quantized fields can undergo local (space-time dependent) transformations (gauge transformations, coordinate diffeomorphisms) that involve classical

*TOP2005 symposium, sapporo, japan, march 2005.
†This work is supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative research agreement DE-FC02-94ER40818.

functions whose topological properties determine the allowed quantum field theoretic structures. (ii) One formulation of quantum field theory uses a functional integral over classical fields, and classical topological features become relevant. (iii) Semi-classical (WKB) approximations to the quantum theory rely on classical dynamics, and again classical topology plays a role in the analysis.

Topological effects in quantum electrodynamics were first appreciated by Dirac in his study of the quantum mechanics for (hypothetical) magnetic monopoles. This analysis leads directly to contemporary analysis of Yang-Mill theory – the contemporary generalization of Maxwell's electrodynamics – and has yielded several significant results: the discovery of the θ-vacuum angle; the recognition that c-number parameters in the theory may require quantization for topological reasons (like Dirac's monopole strength); the realization that the chiral anomaly equation is just the local version of the celebrated Atiyah-Singer index theorem.

Here I shall not describe the Yang-Mills investigations; they are too technical and too specialized for this general audience. Rather I shall show you how a topological effect in a condensed matter situation leads to charge fractionalization. This phenomenon has a physical realization in 1-dimensional (lineal) polymers, like polyacetylene, and in 2-dimensional (planar) systems, like the Hall effect.

The polyacetylene story is especially appealing, because it can be told in several ways: in pictorial terms which only involves counting, or in the first quantized formalism for quantum mechanical equations, or in the second quantized formalism of a quantum field theory [1].

2. The Polyacetylene Story (Counting Argument)

Polyacetylene is a material consisting of parallel chains of carbon atoms, with electrons moving primarily along the chains, while hopping between chains is strongly suppressed. Consequently, the system is effectively 1-dimensional. The distance between carbon atoms is about 1Å.

If the atoms are considered to be completely stationary, *i.e.* rigidly attached to their equilibrium lattice sites, electron hopping along the chain is a structureless phenomenon. However, the atoms can oscillate around their rigid lattice positions for a variety of reasons, like zero-point motion, thermal excitation, etc. It might be thought that these effects merely give rise to a slight fuzzing of the undistorted-lattice situation.

In fact this is not correct; something more dramatic takes place. Rather

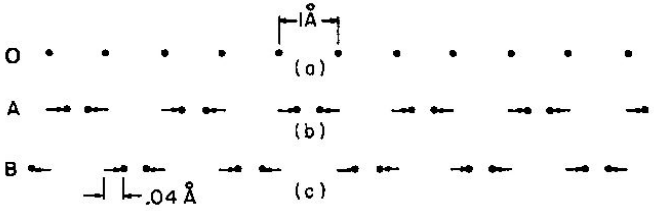


Figure 1. (a) The rigid lattice of polyacetylene; (O) the carbon atoms are equally spaced 1 Å apart. (b), (c) The effect of Peierls' instability is to shift the carbon atoms .04Åto the right (A) or to the left (B), thus giving rise to a double degeneracy.

than oscillating about the rigid-lattice sites, the atoms first shift a distance of about .04 Å and then proceed to oscillate around the new, slightly distorted location. That this should happen was predicted by Peierls, and is called the Peierls instability. Due to reflection symmetry, there is no difference between a shift to the right or a shift to the left; the material chooses one or the other, thus breaking spontaneously the reflection symmetry, and giving rise to doubly degenerate vacua, called A and B.

If the displacement is described by a field ϕ which depends on the position x along the lattice, the so-called phonon field, then Peierls' instability, as well as detailed dynamical calculations indicate that the energy density $V(\phi)$, as a function of constant ϕ, has a double-well shape. The symmetric point $\phi = 0$ is unstable; the system in its ground state must choose one of the two equivalent ground states $\phi = \pm \mid \phi_0 \mid = \pm .04$Å. In the ground states, the phonon field has uniform values, independent of x.

By now it is widely appreciated that whenever the ground state is degenerate there frequently exist additional stable states of the system, for which the phonon field is non-constant. Rather, as a function of x, it interpolates, when x passes from negative to positive infinity, between the allowed ground states. These are the famous solitons, or kinks. For polyacetylene they correspond to domain walls which separate regions with vacuum A from those with vacuum B, and vice versa. One represents the chemical bonding pattern by a double bond connecting atoms that are closer together, and the single bond connecting those that are further apart.

Consider now a polyacetylene sample in the A vacuum, but with two solitons along the chain. Let us count the number of links in the sample without solitons and compare with number of links where two solitons are present. It suffices to examine the two chains only in the region where they differ, *i.e.* between the two solitons. Vacuum A exhibits 5 links, while the

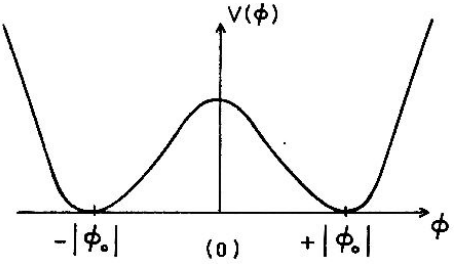


Figure 2. Energy density $V(\phi)$, as a function of a constant phonon field ϕ. The symmetric stationary point, $\phi = 0$, is unstable. Stable vacua are at $\phi = +|\phi_0|$, (A) and $\phi =$ $-|\phi_0|$, (B).

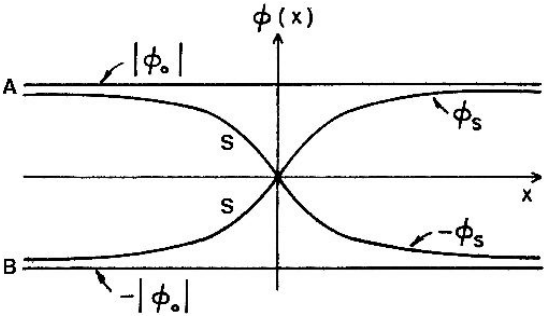


Figure 3. The two constant fields, $\pm \mid \phi_0 \mid$, correspond to the two vacua (A and B). The two kink fields, $\pm\phi_s$, interpolate between the vacua and represent domain walls.

addition of two solitons decreases the number of links to 4. The two soliton state exhibits a deficit of one link. If now we imagine separating the two solitons a great distance, so that they act independently of one another, then each soliton carries a deficit of half a link, and the quantum numbers of the link, for example the charge, are split between the two states. This is the essence of fermion fractionization.

It should be emphasized that we are not here describing the familiar situation of an electron moving around a two-center molecule, spending "half" the time with one nucleus and "half" with the other. Then one might say that the electron is split in half, on the average; however fluctuations in any quantity are large. But in our soliton example, the fractionization is without fluctuations; in the limit of infinite separation one achieves an eigenstate with fractional eigenvalues.

We must however remember that the link in fact corresponds to two states: an electron with spin up and another with spin down. This doubling

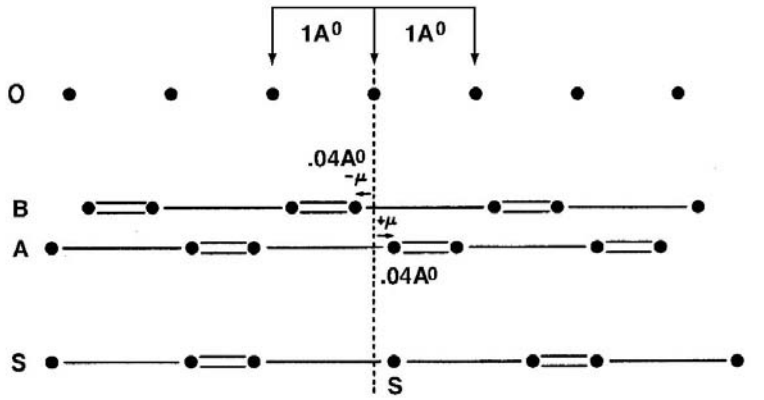


Figure 4. Polyacetylene states. The equally spaced configuration (O) possesses a left-right symmetry, which however is energetically unstable. Rather in the ground states the carbon atoms shift a distance μ to the left or right, breaking the symmetry and producing two degenerate vacua (A, B). A soliton (S) is a defect in the alteration pattern; it provides a domain wall between configurations (A) and (B).

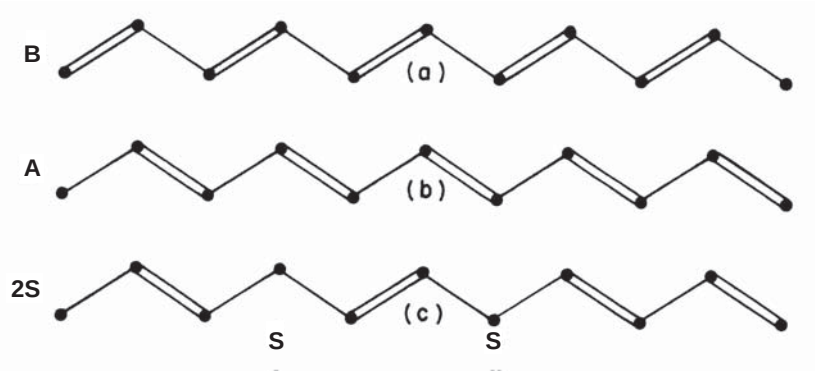


Figure 5. (a), (b) Pattern of chemical bonds in vacua A and B. (c) Two solitons inserted into vacuum A.

obscures the dramatic charge $\frac{1}{2}$ effect, since everything must be multiplied by 2 to account for the two states. So in polyacetylene, a soliton carries a charge deficit of one unit of electric charge. Nevertheless charge fractionization leaves a spur: the soliton state has net charge, but no net spin, since all of the electron spins are paired. If an additional electron is inserted into the sample, the charge deficit is extinguished, and one obtains a neutral state, but now there is a net spin. These spin-charge assignments (charged

– without spin, neutral – with spin) are unexpected, but in fact have been observed, and provide experimental verification for the soliton picture and fractionalization in polyacetylene.

Notice that in this simple counting argument no mention is made of topology. This feature emerges only when an analytic treatment is given. I now turn to this.

3. The Polyacetylene Story (Quantum Mechanics)

I shall now provide a calculation which shows how charge 1/2 arises in the quantum mechanics of fermions in interaction with solitons. The fermion dynamics are governed by an one-dimensional Dirac Hamiltonian, $H(\phi)$, which also depends on a background phonon field ϕ, with which the fermions intact. The Dirac Hamiltonian arises not because the electrons are relativistic. Rather it emerges in a certain well-formulated approximation to the microscopic theory, which yields a quantal equation that is a 2x2 matrix equation, like a Dirac equation. In the vacuum sector, ϕ takes on a constant value ϕ_0, appropriate to the vacuum. When a soliton is present, ϕ becomes the appropriate, static soliton profile ϕ_s. We need not be any more specific. We need not insist on any explicit soliton profile. All that we require is that the topology [*i.e.* the large distance behavior] of the soliton profile be non-trivial.

In the present lineal case the relevant topology is that infinity corresponds to two points, the end points of the line, and the phonon field in the soliton sector behaves differently at the points at infinity.

To analyze the system we need the eigenmodes, both in the vacuum and soliton sectors.

$$H(\phi_0)\psi_E^v = E\psi_E^v \tag{1}$$

$$H(\phi_s)\psi_E^s = E\psi_E^s \tag{2}$$

The Dirac equation is like a matrix-valued "square root" of the wave equation. Because a square root is involved, there will be in general negative energy solutions and positive energy solutions. The negative energy solutions correspond to the states in the valence band; the positive energy ones, to the conduction band. In the ground state, all the negative energy levels are filled, and the ground state charge is the integral over all space of the charge density $\rho(x)$, which in turn is constructed from all the negative

energy wave functions.

$$\rho(x) = \int_{-\infty}^{0} dE\, \rho_E\,(x),\ \ \rho_E(x) = \psi_E^*\,(x)\,\psi_E\,(x) \tag{3}$$

Of course integrating (3) over x will produce an infinity; to renormalize we measure all charges relative to the ground state in the vacuum sector. Thus the soliton charge is

$$Q = \int dx \int_{-\infty}^{0} dE\, \{\rho_E^s\,(x) - \rho_E^v\,(x)\}. \tag{4}$$

Eq. (4) may be completely evaluated without explicitly specifying the soliton profile, nor actually solving for the negative energy modes, provided H possesses a further property. We assume that there exists a conjugation symmetry which takes positive energy solutions of (1) and (2) into negative energy solutions. (This is true for polyacetylene.) That is, we assume that there exists a unitary 2x2 matrix M, such that

$$M\psi_E = \psi_{-E}. \tag{5}$$

An immediate consequence, crucial to the rest of the argument, is that the charge density at E is an even function of E.

$$\rho_E(x) = \rho_{-E}(x) \tag{6}$$

Whenever one solves a conjugation symmetric Dirac equation, with a topologically interesting background field, like a soliton, there always are, in addition to the positive and negative energy solutions related to each other by conjugation, self-conjugate, normalizable zero-energy solutions. That this is indeed true can be seen by explicit calculation. However, the occurrence of the zero mode is also predicted by very general mathematical theorems about differential equations. These so-called "index theorems" count the zero eigenvalues, and insure that the number is non-vanishing whenever the topology of the background is non-trivial. We shall assume that there is just one zero mode, described by the normalized wave function ψ_0.

To evaluate the charge Q in (4), we first recall that the wave functions are complete, both in the soliton sector and in the vacuum sector.

$$\int_{-\infty}^{\infty} dE\, \psi_E^*\,(x)\,\psi_E(y) = \delta(x-y) \tag{7}$$

As a consequence, it follows that

$$\int_{-\infty}^{\infty} dE\,[\rho_E^s\,(x)\,-\rho_E^v\,(x)] = 0. \tag{8a}$$

In the above completeness integral over all energies, we record separately the negative energy contributions, the positive energy contributions, and for the soliton, the zero-energy contribution. Since the positive energy charge density is equal to the negative one, by virtue of (6), we conclude that (8a) may be equivalently written as an integral over negative E.

$$\int_{-\infty}^{0} dE\,[2\rho_E^s\,(x) - 2\rho_E^v\,(x)] + \psi_0^*\,(x)\,\psi_0\,(x) = 0 \tag{8b}$$

Rearranging terms give

$$Q = \int dx \int_{-\infty}^{0} dE[\rho_E^s(x) - \rho_0^v(x)] = -\frac{1}{2}\int dx\psi_0(x)\psi_0(x) = -\frac{1}{2}. \tag{9}$$

This is the final result: the soliton's charge is $-\frac{1}{2}$; a fact that follows from completeness (7) and conjugation symmetry (6). It is seen in (9) that the zero-energy mode is essential to the conclusion. The existence of the zero mode in the conjugation symmetric case is assured by the non-trivial topology of the background field. The result is otherwise completely general.

4. The Polyacetylene Story (Quantum Field Theory)

The quantum mechanical derivation that I just presented does not address the question of whether the fractional half-integer charge is merely an uninteresting expectation value or whether it is an eigenvalue. To settle this, we need a quantum field theory approach, that is we need to second quantize the field. For this, we expand Ψ, which now is an anti-commuting quantum field operator, in eigenmodes of our Dirac equation in the soliton sector as

$$\begin{aligned}\Psi &= \sum^{E}(b_E\,\psi_E^s + d_E^\dagger\,\psi_{-E}^s) + a\psi_0 \\ \Psi^\dagger &= \sum^{E}(b_E^\dagger\,\psi_E^{s*} + d_E\,\psi_{-E}^{s*}) + a^\dagger\psi_0.\end{aligned} \tag{10}$$

The important point is that while the finite energy modes $\psi_{\pm E}^s$ enter with annihilation particle (conduction band) operators b_E and creation anti-particle (valence band) operators $d_E^\dagger$, the zero mode does not have a partner

and is present in the sum simply with the operator a. The zero energy state is therefore doubly degenerate. It can be empty $|->$, or filled $|+>$, and the $a, a^\dagger$ operators are realized as

$$a\,|+>=|->,\ a^\dagger\,|+>=0,\ a\,|->=0,\ a^\dagger\,|+>=|+>\,. \tag{11}$$

The charge operator $Q = \int dx \psi^\dagger \psi$ must be properly defined to avoid infinities. This is done, according to Schwinger's prescription in the vacuum sector, by replacing the formal expression by

$$Q = \frac{1}{2}\int dx\,(\psi^\dagger\psi - \psi\psi^\dagger). \tag{12}$$

We adopt the same regularization prescription for the soliton sector and insert our expansion (10) into (12). We find with the help of the orthonormality of wave functions

$$\begin{aligned} Q &= \frac{1}{2}\sum_E (b_E^\dagger\, b_E + d_E\, d_E^\dagger - b_E\, b_E^\dagger - d_E^\dagger\, d_E) + \frac{1}{2}(a^\dagger a - a a^\dagger) \\ &= \sum_E (b_E^\dagger\, b_E - d_E^\dagger\, d_E) + a^\dagger a - \frac{1}{2}. \end{aligned} \tag{13}$$

Therefore the eigenvalues for Q are

$$Q\,|->=-\frac{1}{2}|->,\quad Q\,|+>=\frac{1}{2}|+>\,! \tag{14}$$

5. Conclusion

This then concludes my polyacetylene story, which has experimental realization and confirmation. And the remarkable effect arises from the non-trivial topology of the phonon field in the soliton sector.

Many other topological effects have been found in the field theoretic descriptions of condensed matter and particle physics. Yet we must notice that mostly these arise in phenomenological descriptions, not in the fundamental theory. In condensed matter the fundamental equation is the many-body Schrödinger equation with Coulomb interactions. This does not show any interesting topological structure. Only when it is replaced by effective, phenomenological equations do topological considerations become relevant for the effective description. Fundamental (condensed matter) Nature is simple!

Similarly in particle physics, our phenomenological, effective theories, like the Skyrme model, enjoy a rich topological structure. Moreover, even the Yang-Mills theory of our fundamental "standard particle physics model"

supports non-trivial topological structure, which leads to the QCD vacuum angle. In view of my previous observation, can we take this as indirect evidence that thisYang-Mills based theory also is a phenomenological, effective description and at a more fundamental level – yet to be discovered – we shall find a simpler description that does not have any elaborate mathematical structure. Perhaps in this final theory Nature will be described by simple counting rules – like my first polyacetylene story. Surely this will not be the behemoth of string theory.

This work is supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative research agreement DE-FC02-94ER40818.

References

1. This research was performed in collaboration with C. Rebbi, and independently by W.P. Su, J.R. Schrieffer and A. Heeger. For a summary see R. Jackiw and J.R. Schrieffer "Solitons with Fermion number 1/2 in Condensed Matter and Relativistic Field Theories" *Nucl. Phys.* **B190**, 253 (1981).

L.P. KADANOFF

Chapter 10

Overview of Phase Transitions

10.1 Thermodynamic Phases

In this chapter, we develop the beginnings of a description of thermodynamic phases and the transitions that occurs as the system changes from one phase to another. Before the beginning of science, people classified the materials in the world about them according to their properties. They saw gases, and liquids, and solids and their distinct behaviors. More recently, scientists have distinguished normal materials from superfluids, and studied a whole variety of magnetic materials, of liquid crystals, of ferroelectrics, and many additional phases of matter. One can make the distinction between the phases in question by observing the qualitative difference among the phases or alternatively by observing the *phase transition* which takes the system from one phase to another. Our prototype system, the Ising model, shows several distinct thermodynamic phases. Some of these are depicted in Fig. 10.1. On the left, we see a paramagnetic phase. There is a considerable randomness to the arrangement of spins and no long-ranged order appears. In contrast, in the right hand panel we see a system with the long-ranged order characteristic of the ferromagnetic phase. Through the entire system the spins tend to point upward with only a few islands of 'wrong' behavior, caused by fluctuations. The center panel depicts a situation intermediate between the ferromagnetic and the paramagnetic. In this so-called 'critical phase' there are

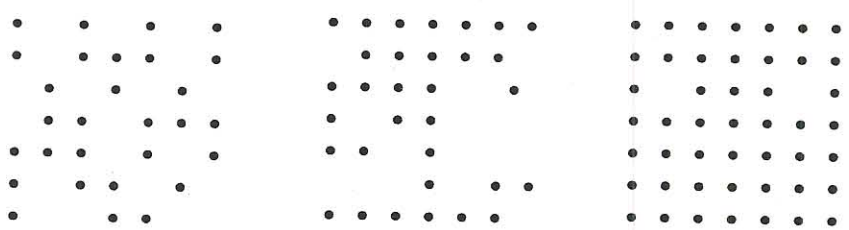


Fig. 10.1. Three different phases for an Ising model. The pictures show spins on a lattice. Spins pointing up are indicated by dots; those pointing down by blanks. The left hand picture shows a paramagnetic phase, the middle one a critical phase, and the right hand picture shows a ferromagnetic phase.

correlations among the different spins. Over large regions of the system, like spins tend to group together. However, one spin direction does not predominate over the other as in the ferromagnetic case. In the following chapters, we shall have much more to say about this critical phase.

10.2 Phase Transitions

The world would indeed be a dull place if there were no distinction among the different phases of matter. No separation of oil and water, no distinction between stream-bed and stream, none between sky and earth, no difference between airy snowflake and sitting water. But it is far from obvious why such distinctions and changes between phases should exist. If a phase transition is to occur, it must be true that two different thermodynamic phases can both exist under the same external thermodynamic conditions. Thus for example a dilute fluid, water vapor, and a dense fluid, liquid water, can be in contact with one another under identical conditions of temperature and pressure. Here the phases are different in character. Each phase of water has its own free energy which depends on number, volume, and temperature. When the free energies of the different phases are different, then the phase with the smaller free energy is the more stable. (This is part of a general rule which states that the relative probability of observing configurations of the system with different free energies is proportional to an exponential of the free energy difference divided by $k_{\rm B}T$. Thus when two configurations differ by a macroscopic energy difference, in a situation of thermodynamic equilibrium, only the phase with the lower free energy will be observed.)

Thus if two phases are in equilibrium, they must have the same free energy per particle. So coexisting water vapor and liquid water have equal free energy per particle. In fact, the line of coexistence is computed by calculating the free energy as a function of density and temperature in each phase, and then finding the density in which the free energy per particle is equal in the two phases. In this case, the equality of the free energy is viewed as just an accidental occurrence.

But, sometimes there is some symmetry relation which defines the behavior of the system. In that case, there will often be some kind of symmetry connecting the coexisting phases. In a magnetic material, for example, the magnetism is produced by spins and other magnetic moments which can line up spontaneously to produce a natural magnetism. A magnetic moment is a special kind of vector, one which changes sign under the operation of time reversal. (Change the direction of every velocity in a system. The electromagnetic currents will all change sign, and with them all magnetic moments and magnetic fields.) In the absence of a magnetic field applied from outside the material, the system will always have a natural degeneracy: If there is a state in which the average magnetization points in a given direction, then there will be an equivalent state in which it will point is exactly the opposite direction. In the simplest kind of magnetic material there is one 'easy axis' and the natural magnetism will occur with the spins lining up parallel or antiparallel to that axis. Call the spin in the direction of that axis σ or σ_z. Then phase transition involves a magnetization density, m, proportional to the average of this spin $\langle\sigma\rangle$. Notice that at the

first order phase transition, the system thus has to choose between two thermodynamically equivalent states, with different values of the magnetization density, m. The two phases have values of m's which differ by a sign. Furthermore, the value of the magnetization density is same throughout a very large region of the sample, one containing many, many spins.[1] The system spontaneously chooses a particular phase, based upon its history or some weak residual magnetic fields. And after this choice is made and the system has a single thermodynamic phase, then $\langle\sigma\rangle$ will have the same sign and the same value throughout the material. One of the most interesting question about this situation is: How can a huge number of atoms, perhaps 10^{23} of them, can chooses one of the possible phases, and apply the choice across the entire system?

10.3 Two Kinds of Transitions

In nature there are two different kinds of phase transitions. The first kind, observed when we boil water at home, involves a discontinuous change in some intensive thermodynamic quantity. This discontinuity is observed, for example, in the density expressed as a function of pressure and temperature. In a magnet, the direction of magnetization might suddenly change so that the magnetization density vector will change discontinuously. These changes are called *discontinuous* phase transition or *first order* phase transitions. Thermodynamic variables, like this density or magnetization, which exhibit the discontinuity, are called *order parameters*, since their value reflects the strength and kind of ordering in the system. Our two examples of first order transitions are the boiling of water or having a domain of iron jump from one direction of magnetization to another. In either case, if we increase the temperature the jump or discontinuity will become smaller. In the fluid, the vapor density will approach the liquid density. In the magnet, the magnitude of the magnetization will get closer to zero. In both cases, at some critical temperature, T_c, the discontinuity will vanish and the phase transition will change its character. At this special point, we say there is a *continuous* or *second order* phase transition.

All of this is illustrated in Figs. 10.2 and 10.3. These figures plot the magnetization density as it is determined by an approximate theory of the ferromagnetic behavior, called mean field theory. We describe this theory in this chapter and the next. For now, we wish to look at results. Figure 10.2 plots dimensionless magnetization, $\langle\sigma\rangle$ versus dimensionless magnetic field, h, for various values of the temperature. Notice how there is a discontinuity in $\langle\sigma\rangle$ at zero values of the field for $T < T_c$ — reflecting a first-order transition. This $h = 0$ discontinuity disappears at $T = T_c$ — producing a second order transition at that point. This discontinuity in m as we pass through vanishing values of h reflects one kind

[1] In what is called a single-domain ferromagnet, the magnetization is uniform throughout the same. But most ferromagnets actually have a tendency to split up into different domains, each with its own direction of the magnetization. This splitting occurs because of weak but long ranged magnetic forces. Magnetic energies are minimized by the splitting into different domains. The domains are usually quite large, containing many atoms. So the ferromagnet is well-represented as a system with long-ranged order over wide regions. In other kinds of ordered phases, the domains can be even larger — extending sometimes over microns or even centimeters.

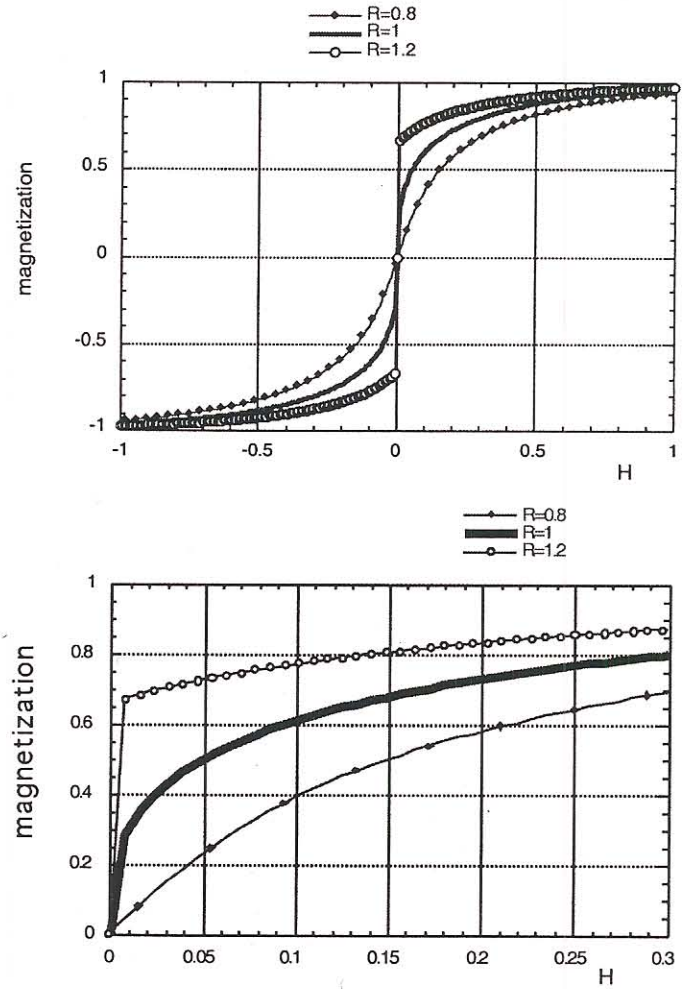


Fig. 10.2. Magnetization plotted as a function of magnetic field. The curves are plotted for three values of $R = T_c/T$, corresponding to above, at, and below the critical temperature. Notice how the curve for $T > T_c$ appears smooth while the one for $T < T_c$ shows a discontinuity at zero field. For $T = T_c$, the discontinuity disappears and is replace by an infinite slope in the magnetization versus field curve. The upper part shows a larger region of field, the lower is a blowup for small values of the magnetic field.

of singularity in this system. This singularity disappears if $T > T_c$, where magnetization is a smooth function of field. Exactly at the critical temperature, we see an infinite slope of m versus h, showing us that the mathematical singularity remains present at this second order transition.

This same data is plotted in a different way in Fig. 10.3. Here we look at m as a function of T. This figure contains several curves each for a different value of the applied magnetic field. The most interesting is the curve marked '$h = 0$' which shows a magnetization which is zero for $T > T_c$. In contrast, for $T < T_c$, m has two branches, one positive-one negative.

There is an exact symmetry which involves flipping the sign of both m and h. Thus if the function $m(h, T)$ gives the magnetization as a function of temperature and magnetic field,

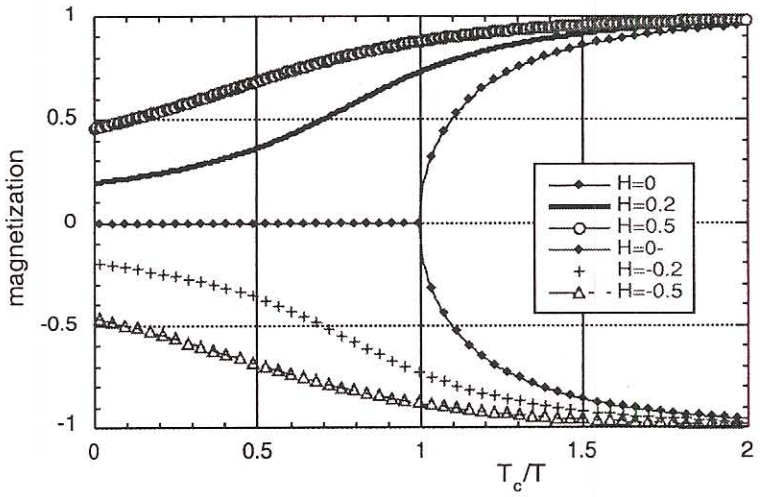


Fig. 10.3. Magnetization plotted as a function of the ratio, R of critical temperature to actual temperature for different values of the magnetic field. Notice the jump in the zero field magnetization for temperatures below the critical temperature.

another equally good solution is $-m(-h,T)$. Each curve with $h > 0$ has a corresponding curve — with oppositely signed m — that applies for the magnetic field with flipped sign. The symmetry is also realized at zero field. At high temperatures, the magnetization is zero for $h = 0$. Then the symmetry is trivially realized since changing the sign of the magnetization changes nothing. However, at low temperatures, for each value of the temperature, the magnetization can take on either of two possible values. The symmetry under flipping the sign of the magnetic field is here reflected in the equality of the magnitude of $\langle\sigma\rangle$ in each of these zero-field phases.

Sometimes we see phase transitions which do not easily fit into the framework given above. There is an interesting phase transition which involves the dissociation of uncharged atoms into charged components. When the charges are tightly bound together we call the material a dielectric. It is an insulator, which may polarize in an electric field but certainly does not conduct electricity. When the atoms dissociate, the charged components may move in opposite directions in response to an electric field. This kind of matter is called a plasma, and can conduct electricity. In general, one does not know whether this *metal to insulator* transition is first order or higher. There is a realization of this transition in three-dimensional condensed matter systems.[2] A very similar transition occurs in the early universe in which quarks bind together to form the hadrons and mesons we see today. In the Chapters 15 and 16, we shall discuss the two-dimensional version of this phase transition. Of course this transition also occurs in three-dimensional systems.

Transitions into and out of the glassy state, provides other examples in which the basic nature of the transition is unknown. A ordinary glass is a liquid in which some of the molecular motions are very substantially slowed down. The phase transitions become hard

[2]For a discussion of this transition see the several papers in P. W. Anderson *A Career in Theoretical Physics* (World Scientific, Singapore, 1994).

to see because one has to wait too long. We would like to extrapolate to the phase transition. But we are not quite sure how that might work.

10.4 Back to the Ising Model

We have already discussed in Chapter 4 the setup of an Ising problem which describes the simplest ferromagnets. Recall the previous discussion. We can set up a Hamiltonian which contains a set of spins $\sigma_{\mathbf{r}}$, where $\mathbf{r}$ is the lattice site defining the location of the spin. Each $\sigma_{\mathbf{r}}$ takes on two values: plus one and minus one. As in Chapters 2 and 3, we work with a simple hypercubic lattice

$$\mathbf{r} = a(n_1, n_2, \ldots, n_\mu, \ldots, n_d)\,, \tag{10.1}$$

where a is the lattice constant and each of the n's is an integer. To produce a ferromagnetic phase transition there must be an interaction between neighboring spins which gives a lower energy to spin configurations in which spins are aligned. An external magnetic field is taken into account by saying that there is a lower energy when a spin is aligned along the magnetic field than opposite to it. This kind of situation can be represented by a Hamiltonian with two terms

$$\mathcal{H} = -\sum_{\mathbf{r}} \sigma_{\mathbf{r}} B(\mathbf{r})\mu - J \sum_{\langle \mathbf{r},\mathbf{r}' \rangle} \sigma_{\mathbf{r}}\sigma_{\mathbf{r}'}\,. \tag{10.2}$$

Here $2J$ is the energy cost for moving a pair of neighboring spins from an aligned configuration to an anti-aligned one while $2B(\mathbf{r})\mu$ is the energy cost of moving the spin at $\mathbf{r}$ from being lined up with the field to the opposite. The first sum says that the latter energy applies to all spins. The notation $\langle \mathbf{r}, \mathbf{r}' \rangle$ means include all pairs of nearest neighboring spins. The number of nearest neighbors of a given lattice site is generally denoted by z. For the simple hypercubic lattice, $z = 2d$. We take both J and $B\mu$ to be positive. The partition function is obtained by summing an exponential of $-\beta\mathcal{H}$, which we write as

$$W[\sigma] = -\beta\mathcal{H} = \sum_{\mathbf{r}} h(\mathbf{r})\sigma_{\mathbf{r}} + K \sum_{\langle \mathbf{r},\mathbf{r}' \rangle} \sigma_{\mathbf{r}}\sigma_{\mathbf{r}'}\,. \tag{10.3}$$

Here $h(\mathbf{r})$ is a dimensionless rendition of the magnetic field and K is a dimensionless version of the coupling strength. Both of these go to zero at infinite temperature and to infinity at zero temperature. In most of our calculations we shall let the field $h(\mathbf{r})$ be independent of position. In that case, we shall write it as simply h.

Notice that the Hamiltonian is left invariant if we change the sign of both the magnetic field, h, and also of each of the spins, $\sigma_{\mathbf{r}}$.

We have already talked about the possibility (and difficulties) of calculating the consequences of the Hamiltonian (10.3) for statistical systems. The calculation can be carried out for one dimension and at zero field for two dimensions. (See Chapter 4.) Beyond that, little can be done exactly; it is just too hard. However there is one special case which is

quite easy: the one in which the coupling among the spins is zero. Then, as we have already seen, the spin is determined by a single particle Hamiltonian, ϵ, with

$$-\beta H = h\sigma_{\mathbf{r}}\,, \tag{10.4}$$

a simple calculation gives the average magnetization as

$$\langle\sigma_{\mathbf{r}}\rangle = \tanh h\,. \tag{10.5}$$

But Eq. (10.5) tells us nothing very direct about the phase transition.

10.5 Mean Field Theory of Magnets

But we can get some of the qualitative properties of the phase transition by examining an approximate theory called 'mean field theory'. The many variants of this theory work reasonably well and, in fact, often provide a qualitatively correct theory of phase transitions and other physical properties. The idea behind this theory is to make the approximation that each spin behaves as if it were an independent spin sitting in the mean field produced by all the other spins. To reflect this idea we focus upon the particular spin at $\mathbf{r}$ and consider how the Hamiltonian depends upon that spin alone. We see that

$$-\beta\mathcal{H} = h\sigma_{\mathbf{r}} + K\sigma_{\mathbf{r}} \sum_{\mathbf{r}' \text{ nn to } \mathbf{r}} \sigma_{\mathbf{r}'} + \text{constant}\,. \tag{10.6}$$

The last sum is over all $\mathbf{r}'$ which are nearest neighbors to $\mathbf{r}$. The 'constant' represents all terms which do not depend upon the spin $\sigma_{\mathbf{r}}$. Now here comes the approximation. Assume that one can replace the fluctuating sum in Eq. (10.6) by a sum of the averages values of all the neighboring spins. Thus, Eq. (10.6) becomes

$$-\beta H = \sigma_{\mathbf{r}} \left(h + K \sum_{\mathbf{r}' \text{ nn to } \mathbf{r}} \langle\sigma_{\mathbf{r}'}\rangle \right) + \text{constant} \tag{10.7a}$$

which can be equally well represented as

$$-\beta H = \sigma_{\mathbf{r}} h_{\text{eff}}(\mathbf{r}) + \text{const}\,, \tag{10.7b}$$

where the effective field has the value

$$h_{\text{eff}}(\mathbf{r}) = h(\mathbf{r}) + K \sum_{\mathbf{r}' \text{ nn to } \mathbf{r}} \langle\sigma_{\mathbf{r}'}\rangle\,. \tag{10.8}$$

As we see, the mean field, h_{eff}, is composed of two parts, first the externally applied field h, and second the internally generated field. Since the spin at $\mathbf{r}$ has its probabilities determined by this effective field, its average is given by equation in this approximation the

average is given by Eq. (10.5) with h replaced by h_{eff}. Thus, the mean field theory implies that

$$\langle \sigma_{\mathbf{r}} \rangle = \tanh h_{\text{eff}}(\mathbf{r}) . \tag{10.9}$$

In the special case in which the system is translationally invariant, we write $\langle \sigma \rangle$ for $\langle \sigma_{\mathbf{r}} \rangle$ and find that Eq. (10.9) is the statement:

$$\langle \sigma \rangle = \tanh h_{\text{eff}} \tag{10.10}$$

where h_{eff} is given by Eq. (10.8) as

$$h_{\text{eff}} = h + Kz\langle \sigma \rangle . \tag{10.11}$$

In this expression, z is, as we have mentioned, the number of nearest neighbors of a given site. In the hypercubic lattice, $z = 2d$, where d is the dimension of the system.

We might expect Eq. (10.10) to be accurate in the limit as z becomes very large. For, when z is large, we have an effective field which is a sum of many terms. The fluctuation in this field gets smaller and smaller as z gets larger and larger. In the limit, the spin in question sees the summed effects of many, many different spins and one might expect to replace the sum by its average. Thus, in particular, we might expect the mean field theory to be particularly accurate for systems in very high dimensions. Notice that K is inversely proportional to the temperature. Therefore, Kz can be thought of as T_{c}/T, which T_{c} is some kind of characteristic temperature for the problem. As we shall see that temperature is, in fact, the critical temperature at which we have a continuous transition. We thus can write the effective field for the problem as

$$h_{\text{eff}} = h + \langle \sigma \rangle \frac{T_{\text{c}}}{T} . \tag{10.12}$$

10.6 The Phases

Equations (10.10) and (10.11) constitute the mean field theory for the equilibrium properties of this Ising model. Now we are ready to get the thermodynamic properties of the system by solving Eq. (10.10). Take first the case in which h and $\langle \sigma \rangle$ are both quite small. Then we can expand Eq. (10.10) in a series in both. To first order we find

$$\langle \sigma \rangle = h_{\text{eff}} = h + \langle \sigma \rangle Kz . \tag{10.13}$$

Here z is a symbol for the number of nearest neighbors, which is $2d$ for a hypercublic lattice. Since Kz is inversely proportional to the temperature, we can write it as T_{c}/T and solve Eq. (10.13) to find

$$\langle \sigma \rangle = \frac{h}{1 - T_{\text{c}}/T} . \tag{10.14}$$

Equation (10.14) shows a singularity at $T = T_c$. Therefore we interpret $T_c = TKz$ as the critical temperature.

From Eq. (10.14), so long as the temperature is above T_c, $\langle\sigma\rangle$ and h point in the same direction while, according to Eq. (10.14), when the temperature is below the critical temperature, $\langle\sigma\rangle$ and h are antiparallel. However the latter is an impossible result. To see its impossibility use Eq. (10.14) to find that the magnetic susceptibility, which is the derivative of magnetization with respect to field, has the value

$$\chi = \left.\frac{\partial\langle\sigma\rangle}{\partial h}\right|_T = \frac{1}{1 - T_c/T} . \tag{10.15}$$

Equation (10.15) was derived from the mean field theory (10.10) using as the sole assumption that $\langle\sigma\rangle$ and h were both sufficiently small so that one can do an expansion in them.

However for $T < T_c$ Eq. (10.15) must be nonsense. Recall our discussion of operators and field in Chapter 7. The total magnetization operator $M = \sum_{\mathbf{r}} \sigma_{\mathbf{r}}$ is the operator conjugate to the field h. As a result, we know that

$$\langle M\rangle = \frac{\partial}{\partial h} \ln Z = N\langle\sigma_{\mathbf{r}}\rangle . \tag{10.16a}$$

Here the derivative is to be calculated at fixed coupling, K and constant number of sites, N. A higher order calculation is done in which the same way. We see that the derivative of the average spin is exactly the susceptibility

$$\chi = \frac{\partial}{\partial h}\langle\sigma_{\mathbf{r}}\rangle = \frac{\langle(M - \langle M\rangle)^2\rangle}{N} .$$

Since the last expression is certainly positive, so is the susceptibility. Equation (10.15), then, must fail as soon as T falls below the critical temperature, T_c — since it then gives the wrong sign of the susceptibility. We thus derive a wrong result in this region.

This incorrectness should not be a great surprise to us. If T is far enough below T_c, the magnetization and $\langle\sigma\rangle$ are never very small. Hence one can never expand in them. To see the difficulty, let us try to solve Eq. (10.10) for $\langle\sigma\rangle$ at $h = 0$. To do this, we use a graphical method. We generate graphs for two functions of $\langle\sigma\rangle = x$. The first is just $f(x) = x$. This will represent the left hand side of Eq. (10.10). The second is $g(x) = \tanh(xT_c/T)$. This will be the right hand side. When f crosses g we have a solution for the magnetization. Figures 10.4 and 10.5 show this graphical calculation. The first one, Fig. 10.4, is for $T > T_c$ and shows the one expected root at $x = \langle\sigma\rangle = 0$. The second panel, Fig. 10.5, is for $T < T_c$ and shows three roots! One of these is $\langle\sigma\rangle = 0$ — which has given us trouble. We might suspect that this root is not physically realized. There are two additional roots at nonzero values of the spontaneous magnetization. Because the problem is symmetric under the operation of changing the sign of both σ and h, the two roots are equal in magnitude and opposite in sign.

Somehow the system must choose one of these roots. Which one? We shall see.

A graphical understanding is a beginning. But analytical work often gives us a richer understanding than can be obtained from pure numerics. Therefore, we now turn to obtaining an analytical understanding of the mean field theory for low temperatures.

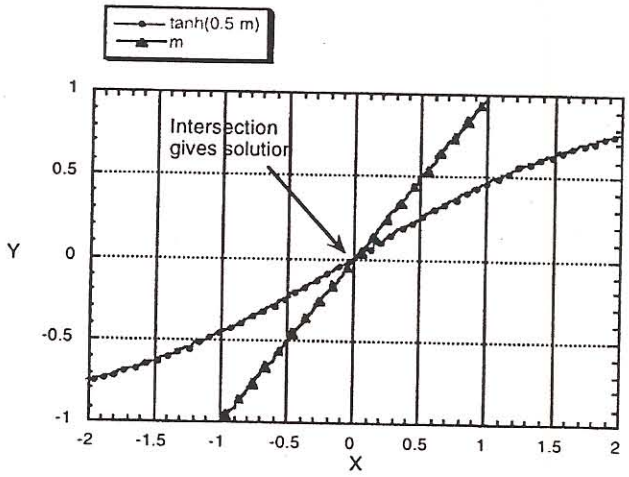


Fig. 10.4. Graphical method of equation solving. Two curves give respectively the left and right hand sides of Eq. (10.9), We calculate the magnetization for zero magnetic field and $T_c/T = 0.5$, using Eq. (10.9). The X-value is $\langle\sigma\rangle$. We draw two curves as a function of X. In one curve, the Y-value is the right hand side of Eq. (10.9) expressed as a function of $\langle\sigma\rangle$ for fixed values of K. For the other it is $\langle\sigma\rangle$ itself. Where these cross we have solutions to Eq. (10.9).

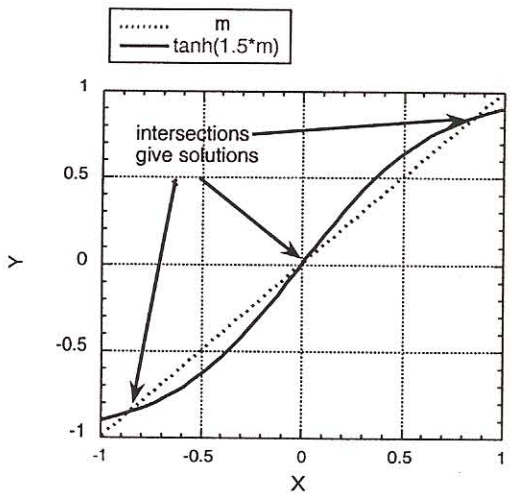


Fig. 10.5. Another graphical calculation of magnetization. Just like Fig. 10.4, except that T is one half of T_c. We follow the same method of equation solving. As in the last figure, the X-axis represents $\langle\sigma\rangle$ while the Y-axis plots $\langle\sigma\rangle$ and the mean field approximation to that quantity. Now there are three solutions: zero magnetization, and the two values of the spontaneous magnetization.

10.7 Low Temperature Result

The quantity $\langle\sigma\rangle T_c/T$ as the interpretation of being the portion of the effective magnetic field induced by the interaction among the spins. For low temperatures, this induced field tends to be very large in magnitude. Let us assume that both this field and the usual magnetic field h, are positive. Thus the induced field is lined up with the applied field. In

this case, the argument of the hyperbolic tangent in Eq. (10.10) is very large and one can use the expression:

$$\tanh x = 1 - 2e^{-2x}, \tag{10.17}$$

which is true for large x. Then Eq. (10.10) becomes:

$$\langle\sigma\rangle = 1 - 2\exp\left[-2\left(h + \frac{\langle\sigma\rangle T_c}{T}\right)\right].$$

To lowest order in an expansion appropriate for low temperatures, one can replace the $\langle\sigma\rangle$ on the right hand side of this expression by unity and find:

$$\langle\sigma\rangle = 1 - 2\exp\left[-2\left(h + \frac{T_c}{T}\right)\right], \qquad \text{(phase +)}. \tag{10.18a}$$

When one calculates the susceptibility from this expression one finds a perfectly positive (and small) value consistent with everything we know from thermodynamics and statistical mechanics. We label this phase as + since it has a positive value of the magnetization. Equation (10.18a) looks like a perfectly acceptable solution even for negative values of h. There is another solution which is constructed from the approximation $\langle\sigma\rangle \approx -1$, which we call phase −. Its magnetization has the form:

$$\langle\sigma\rangle \approx -1 + 2\exp\left[-2\left(-h + \frac{T_c}{T}\right)\right], \qquad \text{(phase −)}, \tag{10.18b}$$

and yet a third solution, which we obtained before as Eq. (10.14)

$$\langle\sigma\rangle = \frac{-hT}{T_c - T}, \qquad \text{(phase 0)}. \tag{10.18c}$$

All three solutions look OK. In the next section, we shall argue that only one solution at a time is acceptable. The conclusion as the temperature goes to zero (10.18a) is an acceptable solution only for $h \geq 0$, while (10.18b) is acceptable only for $h \leq 0$, and (10.18c) is never true for low temperatures. This last phase is labeled phase zero since it will have zero magnetization at zero field.

10.8 Free Energy Selection Argument

Equation (10.18) give us three presumably possible thermodynamic states for the system when h is small and the temperature is low. Which one(s) of these situation represents a possible behavior of the real system? To make the right selection we follow a principle that whatever the thermodynamic state achieved by the system is the one which produces the minimum of the free energy, F, or the maximum of the quantity

$$\exp(-\beta F) = \mathrm{Tr}\,\exp(-\beta H). \tag{10.19}$$

In writing this expression, we use our convention that the quantum sum over states and the classical sum over phase space will be represented by the symbol Tr. To find the free energy we notice that the derivative of the free energy with respect to β is given in terms of the average value of the energy:

$$-\frac{\partial \beta F}{\partial \beta} = \langle H \rangle \,. \tag{10.20}$$

In mean field theory, to calculate the average energy we start with H and replace each spin in the Hamiltonian by its average. Thus, Eq. (10.20) becomes

$$-\beta \frac{\partial(\beta F)}{\partial \beta} = h \sum_{\mathbf{r}} \langle \sigma \rangle + K \sum_{\langle \mathbf{r},\mathbf{r}' \rangle} \langle \sigma \rangle^2 = N \langle \sigma \rangle \left[h + \frac{K z \langle \sigma \rangle}{2} \right] . \tag{10.21}$$

Here, N is the number of lattice sites while Kz has the physical interpretation of being the ratio T_{c}/T. Now consider the case $h = 0$. We wish to know how the free energy depends upon our choice of state. For $T > T_{\mathrm{c}}$, there is no choice. For $T < T_{\mathrm{c}}$, in the unmagnetized state $\langle \sigma \rangle$ is and remains zero. Then if F_{c} is the value of the free energy at the critical point, Eq. (10.21) reads

$$F = F_{\mathrm{c}} \quad \text{for } h = 0 \text{ and all } T \qquad (\text{phase } 0) \,. \tag{10.22}$$

Then the free energy is F_{c} for this configuration. In the magnetized configuration, Eq. (10.21) reads

$$T \left(\frac{\partial(\beta F)}{\partial T} \right)_h = N \frac{T_{\mathrm{c}}}{T} \frac{\langle \sigma \rangle^2}{2} \,. \tag{10.23}$$

All factors in this expression are non-negative.

Notice that Eq. (10.23) describes all three phases at once. To get the actual free energy at $T < T_{\mathrm{c}}$, one integrates this equation downward in temperature starting at T_{c}, using the boundary condition that $\beta F = \beta_{\mathrm{c}} F_{\mathrm{c}}$ at the critical temperature. One visualizes doing the integral separately in each of the three phases, in each case stopping the integration at some temperature, T, below the critical temperature. Then one finds that in phase zero the right hand side is zero so that βF remains at the critical value $\beta_{\mathrm{c}} F_{\mathrm{c}}$. On the other hand, in the other two phases, the right hand side of Eq. (10.23) is positive and thus gives a negative contribution to the downward integral, so that

$$\beta F < \beta_{\mathrm{c}} F_{\mathrm{c}} \text{ for } h = 0 \text{ and } T < T_{\mathrm{c}} \qquad (\text{phase} + \text{ and phase} -) \,. \tag{10.24}$$

The general rule is that configurations of lower free energy are more likely. In big systems, in which the number of degrees of freedom is approaching infinity, a small difference in free energy per site can results in vastly different probabilities. Here, for $T < T_{\mathrm{c}}$, we have a lower free energy for the two magnetized phases. The system will pick one of these two, and not the unmagnetized phase. A similar argument shows that in a system with a very

large number of lattice sites, N, if h is positive only the positive magnetization phase will be realized. Conversely for negative h only phase — will be achieved. There two phases will be cleanly separated for sufficiently large values of $|h|$. In homework Problem 10.3 we ask you to show that the condition '$|h|$ is sufficiently large' gets weaker and weaker as the number of sites gets larger. Thus, in an infinite system all nonzero values of $|h|$ give us only one thermodynamically stable phase.

So far, we have not emphasized one essential fact about all these phase transitions. *Phase transitions only occur in infinitely large systems.* Phase transitions are signaled by qualitative differences between two phases of a thermodynamic system. They occur when there is some kind of mathematical discontinuity in the behavior of the system as a function of parameters like K and h. For example, in our Ising example as N goes to infinity the magnetization jumps discontinuously as a function of h. However, it is easy enough to prove mathematically that such a jump cannot occur for a finite system.[3] Thus, somehow, phase transitions must be linked to the fact that there are an effectively infinite number of particles in the system. So they are not just a result of mechanics. They are reflections of the mechanics of large systems.

10.9 Behaviors of Different Phases

It is tautological to say that different phases behave differently. A crystalline solid is different from a liquid, which is different from a superfluid, and that is different from a ferromagnet. There is a direct, but deep, reason for the difference. In many, but not all phase transitions, one or more of the phases has a nontrivial set of symmetry properties. In this case 'nontrivial' means that the configuration of the system has less than the full symmetry of the Hamiltonian. For example, a crystalline solid has a Hamiltonian which displays both a rotational and a translational invariance. However, the produced lattice shows neither symmetry. Each lattice picks out a set of directions for its crystalline axes. A rotation of the crystal produces another, equivalent, but different configuration. Similarly every lattice picks out preferred positions, which are the equilibrium position of the constituents. Displace the entire lattice a little, and you have other equilibrium positions. Thus, the full translational symmetry is broken. One can often see the broken rotational order. Crystalline solids produce *crystals*, that is pieces of material with surfaces arranged in facets along directions set by the crystal axes. The translational order cannot be seen with the naked eye. Nonetheless these ordering priorities have consequences which are immediately apparent. In contrast to a fluid, a solid is resistant to a shear. It holds its shape. The difference between the disorder of a fluid and the order of a solid is particularly evident in its dynamical behavior. A fluid will propagate only one kind of sound, a longitudinal wave.

[3]The magnetization is the logarithmic derivative of the partition function. Such a derivative can never show a discontinuity in any region in which Z is analytic in h. But, for a finite system, Z is a finite sum of exponentials of h. It is analytic. Therefore no discontinuities. Indeed, finite-but-large systems can show changes which appear quite abrupt. But, in each case, a close examination of the 'jump' region will show a continuous behavior. The abrupt jumps only occur for infinite N.

Typically, a solid will respond to a transient shear by producing a transverse sound wave. This kind of excitation is directly produced because the shear tries to break the produced order and then the system responds dynamically to restore the order. It is generically true that interesting phases show some kind of broken symmetry, by producing a kind of ordering. It is further true that when this ordering is locally disturbed, the system will produce a dynamical response characteristic of the phase in question.

The Bose condensed system provides another example. The superfluid ordering is achieved by having a macroscopic number of particles fall into a single quantum state. This state is defined by a complex wave function. The ordering determines the wave function of this state, defining its magnitude but not its phase angle.[4] In the Bose condensed system, the condensate wave function has a phase angle, which defines the thermodynamic phase, and takes on all possible values between 0 and 2π. Thus the thermodynamic phase has a symmetry under the rotation of the phase angle. This symmetry is called a 'U(1)' symmetry.

In equilibrium the phase angle does not vary in space. The system responds to variations in the phase angle in two different ways. If there is a fixed spatial variation produced by external forces, then the system will produce a 'supercurrent' in which there is a mass-flow along the gradient of the phase. This current is entirely different from ordinary mass flow in that this kind of flow can occur without friction, and without energy losses from heating. In contrast, if we apply no forces, but start the system in a situation with a spatially varying phase angle, then it will produce a new characteristic form of sound wave called 'second sound'. We should not be surprised if a U(1) symmetry can produce waves, after all photons are produced in much the same fashion.

One almost always finds that a broken symmetry produces an ordering and that disturbing that ordering produces some characteristic type of dynamical response. This response is called a 'Goldstone–Nambu Boson'. Jeffrey Goldstone, now at MIT, and Yoshiro Nambu from Chicago first made this general argument. The word 'boson' appears in the argument because one important application of the argument is to broken symmetry states in particle physics. The reader has been prepared, specifically in Chapter 4, for the fact that there is an intimate connection between statistical physics and particle physics. We shall see more of this connection in the next few chapters. We start here. The particle physicist is often interested in understanding the connection between the more 'fundamental' short-ranged and high energy interactions and the more accessible lower energy and longer-ranged behavior. The statistical physicist is also interested in relating microscopic interactions to macroscopic behavior. Most of the richness provided by microscopic interactions are really present only at the micro scale. However a few effects persist out to the macro scale. These are:

(1) The effects of inherently long-ranged forces like electromagnetic interactions or gravity.

[4]Notice that we are in the process of using two different meanings of the word 'phase'. There is a phase of a thermodynamic system, e.g. liquid or gaseous. To be more precise, we use the words 'thermodynamic phase' to specify this meaning. But, there is also the phase of a complex number. We refer to the latter as a 'phase angle'.

(2) The special effects of long-ranged correlation produced near the critical point. The theory of critical behavior, which develops this class of application is the subject of succeeding chapters.

(3) Long-ranged interaction which occur in broken symmetry states.[5]

The treatment of long-ranged forces are almost identical in particle physics and statistical physics. Thus, in recent years, the paths of statistical theory and particle theory have often run parallel. The discussion of the Higgs particle in high energy physics followed the lines of the treatment of the Anderson mode in superconductors.[6] The critical state has formed the basis for the calculation of the masses and interactions which occur in particle physics.

This chapter has been concerned with the mean field theory of first order phase transitions. This theory neglects fluctuations in the ordering which drives the phase transition. Of course, in an finite-range, finite-system the fluctuations are always there. Thus, our analysis has left something out. The neglected fluctuations are important. Sometimes the system is in the 'wrong' phase — that is it is a liquid when, by free energy considerations, it should be a solid. In another case, its magnetization points opposite to the magnetic field. In these situations the real system will undergo a dynamical process which will bring it to the 'right' phase. The process starts with a fluctuation. In a transitory process, the system produces a region within the 'wrong' phase that is essentially a small piece of the 'right' phase. If that region is large enough it will tend to grow and bring the entire system into the equilibrium phase. Naturally, in order to understand, the process, we must understand fluctuations and it is precisely these fluctuations which are left out in mean field theory. One must go beyond mean field theory to understand some of the most interesting problems in first order phase transitions: the long-lived existence of states of the wrong phase (termed *metastability*), the particularly long-lived metastable states called glasses, and the dynamical processes by which equilibrium is restored. These issues are not fully understood to this day. So we go on, to the somewhat easier problem of second order phase transitions.

Homework

Problem 10.1 (Magnetization versus T_c/T). Draw plots of the solutions to the mean field equation for the magnetization as a function of T_c/T for $h = 0$, $h = 0.1$, $h = 0.2$, and $h = 1.0$.

Problem 10.2 (Anti-ferromagnetism). An anti-ferromagnetic has spins at different sites lined up in opposite directions, so that the total magnetization is zero. Mean field theory gives a description of anti-ferromagnetism directly analogous to the ferromagnetism we discussed. It arises for negative values of the coupling J and hence negative values of K.

[5]Particle physics is much concerned with showing that long-ranged forces can always be understood in terms of microscopic mechanism like #2 and #3. The first point is that there are some inherently long-ranged forces in nature. Statistical physics uses these forces; particle physics must try to explain them, perhaps in terms of #2 and #3.

[6]P. W. Anderson, Phys. Rev. **130**, 62 (1962).

(1) Find the appropriate mean field theory for antiferromagnetism.

(2) Calculate the magnetic susceptibility for this system.

Problem 10.3 (Phase Probabilities). Below Eq. (10.24) we discussed a free energy evaluation for the relative probabilities for the different phases of the system. Imagine a system with N particles, at $h = 0$ with a temperature a little below the critical temperature. Using mean field theory, estimate the probability that the system will be in the unmagnetized phase for $N = 100$ and $N = 10^{20}$.

Now apply a small positive magnetic field. For $T < T_c$, what is the condition that makes the phase with negative magnetization have a probability of being observed greater than one part in 10^5.

Problem 10.4 (Magnetic Susceptibility for $T < T_c$). Calculate an expression for the magnetic susceptibility in mean field theory for $T < T_c$ and $h = 0$. Evaluate this expression more explicitly for low temperatures and for temperatures close to the critical temperature.

G.A. SAWATZKY

Spectral-weight transfer: Breakdown of low-energy-scale sum rules in correlated systems

M. B. J. Meinders
Materials Science Centre, Department of Solid State and Applied Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

H. Eskes
Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, D-7000 Stuttgart 80, Federal Republic of Germany

G. A. Sawatzky
Materials Science Centre, Department of Solid State and Applied Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

(Received 4 March 1993)

In this paper we study the spectral-weight transfer from the high- to the low-energy scale by means of exact diagonalization of finite clusters for the Mott-Hubbard and charge-transfer model. We find that the spectral-weight transfer is very sensitive to the hybridization strength as well as to the amount of doping. This implies that the effective number of low-energy degrees of freedom is a function of the hybridization and therefore of the volume and temperature. In this sense it is not possible to define a Hamiltonian which describes the low-energy-scale physics unless one accepts an effective nonparticle conservation.

I. INTRODUCTION

In spite of a lot of theoretical and experimental studies, there is still little understanding about the normal-state excitation spectrum, and therefore of the low-energy-scale physics, of (strongly) correlated systems. Because the full many-body Schrödinger equation is not solvable, one is forced to make approximations to describe the (low-energy) physics of correlated materials. The first thing usually done is to construct a model Hamiltonian, which incorporates the most important characteristics of strongly correlated systems. However, even the one-dimensional Mott-Hubbard (MH) model, which is one of the least complicated realistic model Hamiltonians, is not fully understood. Although one has established that the low-energy physics of the one-dimensional MH model, which has been solved by Lieb and Wu,[1] can be described in terms of spinons and holons,[2] the extraction of the relevant physical information is still a problem.[3] The two- and three-dimensional MH model as well as the three-band charge-transfer (CT) model, where an uncorrelated band exists between the lower Hubbard band (LHB) and the upper Hubbard band (UHB) (as in the high T_c's) have not been solved yet, and the description of the low-energy physics is a very interesting field of proposals and intelligent guesses.

For understanding the low-energy physics, further approximations have to be made. Often one treats the on-site Coulomb repulsion in a mean field way as in a homogeneous or inhomogeneous[4] Hartree-Fock (HF) calculation with the addition of random-phase approximation to describe optical and spin excitations.[5] To reduce the Hilbert space and thereby hopefully the complexity of the problem, one projects out the high-energy states (doubly occupied states) completely, as one does when reducing the three-band Emery model[6] or one-band Hubbard model to a t-J model.[7] A different strategy is to use a slave operator technique to project out double occupation. This results in a mean-field starting point for the calculation but has the advantage that the usual many-body techniques can be used to include correlations.[8]

Another approach is to diagonalize the model Hamiltonian exactly or use Monte Carlo methods, but then one is restricted to small finite clusters resulting in large energy spacings between the levels. These energy spacings for a one-dimensional chain of 20 sites or a two-dimensional 4×4 cluster is of the order of 50 meV or 600 K. This is much too large to describe the low-energy physics in detail.

The main purpose of all these mentioned approximations is generally aimed at finding some suitable effective Hamiltonian, which can be used to describe the low-energy-scale properties of solids. However, this search may be futile if it turns out that the low- and high-energy scales cannot be decoupled, or that such a decoupling is only valid provided one is willing to leave the concepts of fermion or Bose statistics[9] or perhaps even the comforts of well established sum rules.

In this paper we study, using exact diagonalization, one remarkable feature, out of many, of the normal-state excitation spectra, namely, the doping and hybridization dependence of the spectral weight transfer from the high- to the low-energy scale. A nice experimental example of this phenomenon is the O 1s x-ray absorption study[10,11] of the $La_{2-x}Sr_xCuO_4$ system, where, upon hole doping in the O 2p band, spectral intensity of the upper-Hubbard band (high-energy scale) is transferred to the low-energy scale near the Fermi edge. Similar behavior has experimentally been found for several correlated systems.[12,13] In a previous paper[14] we described the general physical

origin of this redistribution of spectral intensities and its doping dependence, while Hybertsen *et al.*[15] gave a description of the O K edge in particular. This occurrence of spectral weight transfer, which we believe is a fingerprint for correlation effects, is now commonly accepted and has been observed in several numerical calculations[16,17] of correlated systems.

In this paper we will study the spectral weight transfer in greater detail and describe the important consequences for the description of the low-energy physics. Furthermore, different approaches such as the so-called Hubbard I solution,[18] the t-J model and mean-field theories will be discussed with respect to the spectral weight transfer.

In Sec. II we review the differences between doping a semiconductor, a localized MH and a CT system. In Sec. III the influence of the hybridization for the MH (Sec. III A) and CT system (Sec. III B) is taken into account. The results will be discussed and compared with other theories in Sec. IV, while also the consequences will be considered. Conclusions are given in Sec. IV.

II. COUNTING PRINCIPLE

An interesting question is what happens with the low-energy scale states if we dope a strongly correlated insulator as for example in $Li_xNi_{1-x}O$ or $La_{2-x}Sr_xCuO_4$. We may study this by looking at the one electron Green's function. Assuming that there are N electrons in the ground state, we define the one electron removal Green's function as

$$G_\sigma^-(\mathbf{k},\omega)=\sum_m \frac{A_\sigma^-(\mathbf{k},\omega_m)}{\omega-E_m^{N-1}+E_{\rm GS}^N+i\eta} \ . \tag{1}$$

Here the summation runs over all $N-1$ particle final states with wave functions ψ_m^{N-1} and eigenenergies E_m^{N-1}. The pole strength is defined as

$$A_\sigma^-(\mathbf{k},\omega_m)=|\langle\psi_m^{N-1}|c_{\mathbf{k}\sigma}|\psi_{\rm GS}^N\rangle|^2 \ ,$$

where $c_{\mathbf{k}\sigma}$ annihilates an electron with momentum $\mathbf{k}$ and spin σ from the N-particle ground state $\psi_{\rm GS}^N$. The angular integrated photoelectron spectrum $W^{\rm PES}$ is now given by

$$W^{\rm PES}(\omega)=\sum_{\mathbf{k}\sigma} W_{\mathbf{k}\sigma}^{\rm PES}(\omega)=\sum_{\mathbf{k}\sigma}\frac{1}{\pi}\lim_{\eta\downarrow 0}{\rm Im}G_\sigma^-(\mathbf{k},\omega) \ . \tag{2}$$

In the same way the one-electron addition Green's function is given by

$$G_\sigma^+(\mathbf{k},\omega)=\sum_m \frac{A_\sigma^+(\mathbf{k},\omega_m)}{\omega-E_m^{N+1}+E_{\rm GS}^N-i\eta} \ . \tag{3}$$

with pole strength

$$A_\sigma^+(\mathbf{k},\omega_m)=|\langle\psi_m^{N+1}|c_{\mathbf{k}\sigma}^\dagger|\psi_{\rm GS}^N\rangle|^2 \ ,$$

where $c_{\mathbf{k}\sigma}^\dagger$ creates an electron with momentum $\mathbf{k}$ and spin σ into the ground state. The inverse photoelectron spectrum $W^{\rm IPES}$ is then given by

$$W^{\rm IPES}(\omega)=\sum_{\mathbf{k}\sigma} W_{\mathbf{k}\sigma}^{\rm IPES}(\omega)=\sum_{\mathbf{k}\sigma}\frac{1}{\pi}\lim_{\eta\downarrow 0}{\rm Im}G_\sigma^+(\mathbf{k},\omega) \ . \tag{4}$$

We note that the pole strengths $A_\sigma^\pm(\mathbf{k},\omega_m)$ are just the overlap between the eigenstates $\psi_m^{N\pm1}$ and the state obtained by suddenly adding or removing an electron from the ground state; thus

$$\begin{aligned} c_{\mathbf{k}\sigma}^\dagger\psi_{\rm GS}^N&=\sum_m A_\sigma^+(\mathbf{k},\omega_m)\psi_m^{N+1} \ , \\ c_{\mathbf{k}\sigma}\psi_{\rm GS}^N&=\sum_m A_\sigma^-(\mathbf{k},\omega_m)\psi_m^{N-1} \ . \end{aligned} \tag{5}$$

For an insulator, doped with x holes, we define the low-energy spectral weight (LESW) for a certain spin and momentum as

$$\begin{aligned} \Lambda_{\mathbf{k}\sigma}(x)&=\int_0^{\omega_g} W_{\mathbf{k}\sigma}^{\rm IPES}(\omega)d\omega \\ &=\int_0^{\omega_g} A_\sigma^+(\mathbf{k},\omega_m)\delta(\omega-E_m^{N+1}+E_{\rm GS}^N)d\omega \ , \\ \Lambda(x)&=\sum_{\mathbf{k}\sigma}\lambda_{\mathbf{k}\sigma}(x) \ . \end{aligned} \tag{6}$$

The LESW for an electron-doped insulator is defined as

$$\begin{aligned} \Lambda_{\mathbf{k}\sigma}(x)&=\int_0^{\omega_g} W_{\mathbf{k}\sigma}^{\rm PES}(\omega)d\omega \\ &=\int_0^{\omega_g} A_\sigma^-(\mathbf{k},\omega_m)\delta(\omega-E_m^{N-1}+E_{\rm GS}^N)d\omega \ , \\ \Lambda(x)&=\sum_{\mathbf{k}\sigma}\Lambda_{\mathbf{k}\sigma}(x) \ . \end{aligned} \tag{7}$$

The upper limit of integration ω_g is chosen somewhere in the correlation gap, which divides the high-energy states from the low-energy states. The chemical potential is set equal to 0. Thus $\Lambda(x)$ may be seen as the effective number of degrees of freedom in the low-energy regime, responsible for the low-energy physics.

Let us consider a semiconductor (which can be described in an independent particle framework) with an occupied valence band and an unoccupied conduction band, separated by an energy gap $E_{\rm gap}$. In this case the poles $G_\sigma(\mathbf{k},\omega)$ are positioned at just the one-particle energies $\epsilon_{\mathbf{k}\sigma}$, as derived from some sort of band theory. The pole strength $A_\sigma^\pm(\mathbf{k},\epsilon_{\mathbf{k}\sigma})$ is 1 or 0, depending on the occupation of the one electron levels. For the undoped semiconductor the total electron removal and addition spectrum will look as sketched in Fig. 1. If the total number of sites equals $\mathcal{N}$ then there are $2\mathcal{N}$ occupied states and $2\mathcal{N}$ unoccupied states, separated $E_{\rm gap}$. Suppose we chemically dope the semiconductor, resulting in the addition of one hole, then the chemical potential will shift into the former occupied band, provided we can neglect the impurity potential of the dopant. The total electron removal spectral weight will be $2\mathcal{N}-1$ (just the number of electrons in the ground state) and the total electron addition spectral weight will be $2\mathcal{N}+1$ (total number of holes in the ground state). We may divide the electron addition spectrum into two parts, a high-energy-scale part (the conduction band) and a low-energy-scale part, which is the unoccupied part of the valence band (see Fig. 1). We immediately see now that

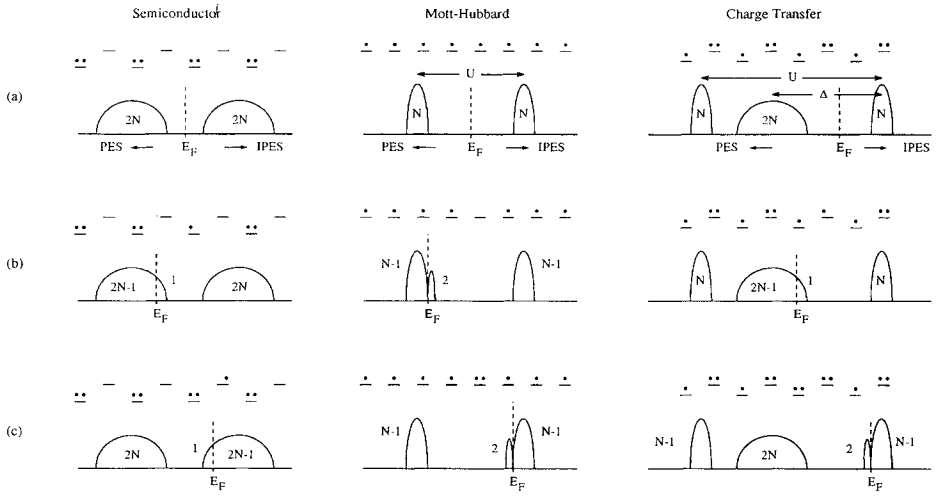


FIG. 1. A schematic drawing of the electron-removal and electron-addition spectra for a semiconductor (left), a Mott-Hubbard system in the localized limit (middle) and a charge transfer system in the localized limit (right). (a) Undoped (half filling), (b) one-electron doped, and (c) one-hole doped. The bars just above the figures represent the sites and the dots represent the electrons. The on-site repulsion U and the charge-transfer energy Δ are also indicated.

the LESW equals 1. The same arguments hold for an electron-doped semiconductor. Thus, when doping a semiconductor, the total spectrum is just given by a repositioning of the chemical potential and the LESW grows as x (with x the amount of doping). At the same time the spectral weight of the high-energy band is not changed. Thus, there is no redistribution of spectral intensities upon doping a simple semiconductor.

Consider now a correlated system described by the MH Hamiltonian. The one-band Hubbard Hamiltonian is defined as

$$H=t\sum_{\langle i,j\rangle,\sigma}(c_{i\sigma}^{\dagger}c_{j\sigma}+\text{H.c.})+U\sum_{i}c_{i\uparrow}^{\dagger}c_{i\uparrow}c_{i\downarrow}^{\dagger}c_{i\downarrow}\ , \tag{8}$$

where t is the hybridization between nearest-neighbor orbitals and U is the on-site repulsion. $\langle ij\rangle$ runs over all nearest-neighbor pairs, i runs over all sites. The $c_{i\sigma}^{\dagger}$ ($c_{i\sigma}$) creates (annihilates) an electron at site i with spin $\sigma=\uparrow,\downarrow$.

We first examine the one-particle Green's function for an $\mathcal{N}$-site MH cluster in the localized limit [$t=0$ in Eq. (8)] as a function of doping. In Fig. 1 (top middle) the total photoelectron and inverse photoelectron spectrum at half filling is shown. The total electron-removal spectral weight is equal to the number of occupied levels, while the total electron-addition spectral weight is equal to the number of unoccupied levels. Therefore, each has an intensity equal to $\mathcal{N}$. Upon doping the system with one hole, there are $\mathcal{N}-1$ singly occupied sites so the total electron removal spectral weight will be $\mathcal{N}-1$. For electron addition there are $\mathcal{N}-1$ ways for adding the electron to a site which was already occupied. Therefore the intensity of the UHB will also be $\mathcal{N}-1$ (not $\mathcal{N}$). We are left with the empty site for which there are two ways of adding an electron (spin up, spin down), both belonging to the LHB. Thus we find $\mathcal{N}-1$ electron removal states near the Fermi-level, two electron addition states near the Fermi level and $\mathcal{N}-1$ electron addition states in the UHB. The same arguments hold for the electron doped case. Thus (normalizing the total PES+IPES spectrum to 2), a doping concentration x yields a LESW $\Lambda(x)=2x$ and the high-energy spectral weight is $1-x$. There have been $\mathcal{N}x$ states transferred from high to low energy.

For the high-T_c superconductors, an oxygen band is located between the LHB and UHB. These systems therefore have to be described as a CT system.[19] A prototype charge-transfer Hamiltonian for CuO_2 planes in the cuprates reads

$$\begin{aligned}H=&\epsilon_d\sum_i d_{i\sigma}^{\dagger}d_{i\sigma}+\epsilon_p\sum_j p_{j\sigma}^{\dagger}p_{j\sigma}\\&+t_{pp}\sum_{\langle jj'\rangle,\sigma}(-1)^{\alpha_{jj'}}(p_{j\sigma}^{\dagger}p_{j'\sigma}+\text{H.c.})\\&+t_{pd}\sum_{\langle ij\rangle,\sigma}(-1)^{\alpha_{ij}}(p_{j\sigma}^{\dagger}d_{i\sigma}+\text{H.c.})\\&+U_{dd}\sum_i d_{i\uparrow}^{\dagger}d_{i\uparrow}d_{i\downarrow}^{\dagger}d_{i\downarrow}\ ,\end{aligned} \tag{9}$$

where i runs over all Cu sites and j over all O sites and $\langle ij\rangle$ represents a nearest-neighbor pair. $d_{i\sigma}^{\dagger}$ ($d_{i\sigma}$) creates (annihilates) a hole on a copper site, while $p_{i\sigma}^{\dagger}$ ($p_{i\sigma}$) creates (annihilates) a hole on an oxygen site. The charge-transfer energy Δ is defined as the difference between the on-site energy of O and Cu ($\Delta=\epsilon_p-\epsilon_d$). The hybridization between the nearest-neighbor oxygen orbitals has strength t_{pp} while the hybridization between a copper orbital and a nearest-neighbor oxygen orbital is t_{pd}. U_{dd} is the on-site Coulomb interactions between two holes on a copper site. The signs of t_{pd} and t_{pp} are described by α_{ij} and $\alpha_{jj'}$ which are 0 or 1, depending on the relative position of a nearest-neighbor copper-oxygen and oxygen-oxygen pair, respectively.

Assume first that we are dealing with a CT system in the localized limit, i.e., we do not allow any hybridization between the oxygen and copper sites ($t_{pd}=0$). Assume no holes in the oxygen band (as in the insulating compounds), then the one particle Green's function will look like that sketched in Fig. 1 (top right). When we dope this system with electrons, the situation is similar to the MH case and we find a spectral weight transfer from high to low energy. Thus in this case the LESW $\Lambda(x)=2x$. However, upon hole doping, the situation is similar to that of the semiconductor without any spectral weight

transfer. So, for the CT system in the localized limit, we find a fundamental asymmetry between hole and electron doping.

In the examples above we could find the LESW just by counting the available states, which can be reached when we do a photoelectron experiment or an inverse photoelectron experiment. This is because the oscillator strength $A_\sigma^{\pm}(\mathbf{k},\omega_m)$ is always one or zero, depending on whether the state is occupied or empty. This is always the case in theories where the Coulomb repulsion is treated in a mean-field way and where the photoelectron spectrum is equal to the density of states (DOS), i.e., when "spectral weight" can be replaced by "states." In local-density approximation (LDA) or restricted Hartree-Fock (HF) calculations the added holes will occupy $\mathbf{k}$ states, and the change in potential due to these holes will be evenly distributed over all sites and can therefore be neglected when $x \ll 1$. Therefore, we will always find a LESW $\Lambda(x)=x$. Inhomogeneous Hartree-Fock calculations[4,5] reveal that, for large U/t, upon doping a state can be pushed out from the upper Hubbard band to the Fermi level. In this case the LESW goes as $\Lambda=2x$, which looks like the ionic MH model. In this case the extra hole will form a local potential and therefore the counting principle for the ionic Hubbard model applies. Lowering U one can expect that this hole will spread out. This means that the extra state moves back to the upper Hubbard band or lies somewhere in the gap. This may be interpreted as $\Lambda(x)$ somewhere between x and $2x$. We shall show that this is not the expected behavior for a correlated system. Note that the definition of $\Lambda(x)$ becomes meaningless when the gap fills up.

Close to half filling and for $U \gg t$ the Hubbard model transforms into the t-J model. This model describes the antiferromagnetic interaction between two spins on neighboring sites and it allows for a restricted hopping between neighboring sites. The restriction consists of the fact that no doubly occupied sites are allowed. It means that we are only dealing with empty and singly occupied states and therefore again one can apply the above reasoning. Thus the LESW goes as $\Lambda(x)=2x$. While the factor $2x$ for the MH and CT model is really only valid in the ionic limit (t or $t_{pd}=0$), in the t-J model it is independent of the actual value of t and/or J. Thus the t-J model behaves as a single-band Hubbard model with $t=0$ for all values of the hybridization.

All these examples show a linear LESW with respect to the doping concentration. We also found that for the hole-doped CT model the LESW behaves like a simple semiconductor, i.e., $\Lambda(x)=x$. This is contrary to what is found experimentally for the high T_c's.[10,11] In the next section we will show that the hybridization yields a strong deviation from linear behavior of the LESW for the correlated MH and CT models and explains the experimental results.

III. INFLUENCE OF HYBRIDIZATION: DYNAMICAL SPECTRAL WEIGHT TRANSFER

We use exact diagonalization on small clusters to calculate the one electron Green's function $G_\sigma^-(\mathbf{k},\omega)$ for the MH and CT model. The Lanczos method as described in, for instance, Refs. 20 or 21 is used. We perform calculations for one- and two-dimensional MH clusters for different values of U/t as well as the CT model for different parameter values and we adopt periodic boundary conditions.

A. Mott-Hubbard model

All the calculations for the MH model [Eq. (8)] are performed with an on-site Coulomb repulsion $U=10$ eV. In Fig. 2 the total electron removal and addition spectra are plotted for a 10-site one-dimensional ring with $t=-1.0$ eV and for different number of electrons N. The hole (electron) doping concentration is given by $x=|(N-\mathcal{N})/\mathcal{N}|$ with the number of sites $\mathcal{N}=10$ ($N<\mathcal{N}$ corresponds to hole doping and $N>\mathcal{N}$ corresponds to electron doping). From the figure we see immediately that if we go from the one-hole- to one-electron-doped case, the chemical potential shifts by just the amount of the insulating gap. Therefore, there is no sign at all of a chemical potential which remains roughly halfway between the two bands due to doping-induced midgap

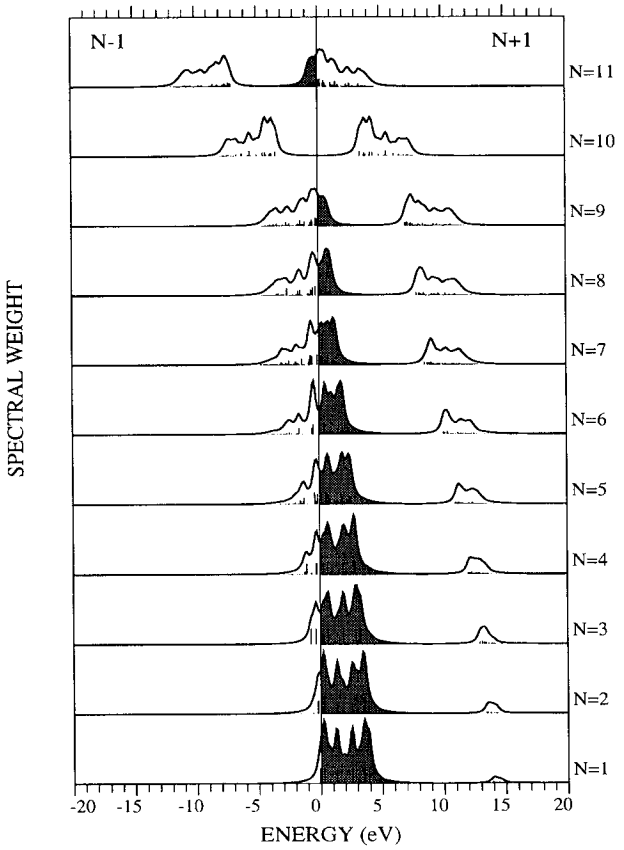


FIG. 2. One-particle Green's function for a one-dimensional Hubbard-ring of $N=10$ sites for $U=10$ eV and $t=1$ eV. The number of electrons in the ground state N are indicated. The low-energy electron-addition spectral weight is obtained by integration over the shaded area.

states,[22] in agreement with Ref. 23 but in contrast to the suggestion by Allen *et al.*[24] We observe the same kind of chemical potential shift for other values of t.

The LESW, as defined by Eq. (6), is obtained by integrating the shaded area. In Fig. 3 we plotted the low-energy electron addition spectral weight for the 10-site cluster from $t=-0.5$ to -2.0 eV in steps of 0.25 eV. For strong hybridization it is no longer possible to define ω_g properly for all doping concentrations because, due to hybridization, incoherent parts of $G(\mathbf{k},\omega)$ are spread through the gap. There is therefore no clean distinction between the low- and high-energy scale. However, for large doping concentrations and large hybridization it may still be possible to define ω_g because of an increase of the gap with doping. This phenomenon is clearly seen in Fig. 2. The observed increase of the original insulating gap with increasing doping concentration is contrary to what is usually found in mean-field-like theories, where the effective on-site repulsion is proportional to the expectation value of the electron density, $U_{\text{eff}}=U\langle n_i\rangle$. Therefore, the gap always has the tendency to collapse with doping because of a decrease of the effective interaction. The increase of the gap in the exact diagonalization with doping has a simple physical origin. Increasing the hole-doping concentration results in an increase of the number of $N+1$-particle states at the Fermi edge with which the states in the UHB can hybridize. This will push these two bands apart and cause an increase of the gap between the low- and high-energy scale states.

It is also interesting to note that mean-field theory gives correct first moment, i.e., the energy average. The first moment for the combined photo-emission and inverse-photo-emission spectrum is given by

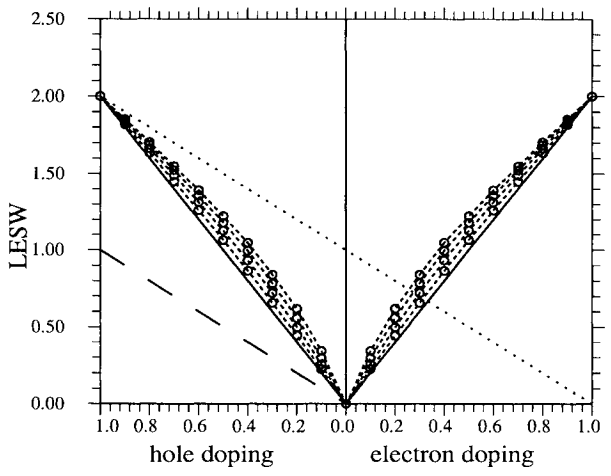


FIG. 3. The integrated LESW divided by the number of sites as a function of the doping concentration x for the one-dimensional ten-site Hubbard ring ($U=10$ eV). The solid lines correspond to the localized limit $t=0$, the dotted line to the free particle limit (hole doped) and the dashed line to hole doping a semiconductor. The data points are from the calculations: $t=-0.5$ eV (lowest) to $t=-2$ eV steps of 0.5 eV.

$$m_1=\frac{1}{\mathcal{N}}\sum_{i,\sigma}\langle\{[c_{i\sigma}^{\dagger},H],c_{i\sigma}\}_+\rangle \tag{10}$$

where $\langle\ \rangle$ is the expectation value in the exact ground state. Substituting the MH Hamiltonian [Eq. (8)] we find for the first moment $m_1=Un=U(N/\mathcal{N})$, where n is the number of electrons per site. For paramagnetic Hartree-Fock we have

$$H_{\text{HF}}=\sum_{\mathbf{k},\sigma}(\epsilon_{\mathbf{k}}+U\langle n_{i\sigma}\rangle)c_{\mathbf{k}\sigma}^{\dagger}c_{\mathbf{k}\sigma}\ . \tag{11}$$

Since the average of the band energies $\epsilon_{\mathbf{k}}$ equals 0, the HF and exact first moments are identical for a homogeneous charge distribution. The same holds for antiferromagnetic HF. However, in HF this correct first moment is generally obtained by changing the eigenenergies with doping, while the weights remain unchanged (equal to 1). This is contrary to exact diagonalization, where the eigenenergies change only little and the correct first moment is obtained by modifying the weights.

In Fig. 3 we also show $\Lambda(x)=x$ corresponding to hole doping a semiconductor as well as the free-particle limit. In the free particle limit there is no high-energy scale so the LESW goes as $\Lambda(x)=x+1$, i.e., $\Lambda(x)=1$ for zero doping (half-filled band) and $\Lambda(x)=2$ for an empty or full band.

From the figure we see that the low-energy spectral weight grows even faster than twice the amount of doping and that it is increasing with increasing $|t|$. At first glance this result is surprising because on increasing $|t|$ we would expect the LESW to go towards the one-particle theory result, which is $\partial\Lambda(x)/\partial x=1$, but instead we go even further away from it. Note that the slope for large doping, close to the empty or full band situation, goes towards the independent particle limit.

The-lowing doping regime for finite t can be understood as follows. We may write the final-state and ground-state wave function as a sum of states belonging to the UHB and of states belonging to the LHB. When calculating the oscillator strength

$$A_{\sigma}^{+}(\mathbf{k},\omega_m)=|\langle\psi_m^{N+1}|c_{\mathbf{k}\sigma}^{\dagger}|\psi_{\text{GS}}^{N}\rangle|^2$$

the phases of the wave functions will constructively interfere for transitions to the low-energy regime and destructively interfere for transitions to the high-energy regime, thereby increasing the LESW. This is a situation similar to one encountered in both core-level and valence-band spectral weights of CE, Kondo-like systems, in which the lowest-energy state obtains more weight than expected from the occupation numbers.[25] $\Lambda(x)>2x$ appears strange because it looks like there is room for much more than one electron per spin and per $\mathbf{k}$ state in the low-energy scale whereas we know that the total number of electrons we can add in the low-energy regime to reach again the insulator cannot be more than the hole-doping level. However, in a many-body system the phase space available to add one electron is larger than the number of electrons you can add in the low-energy scale.

If we substract the static part [the $\Lambda(x)=2x$ part, which comes from the counting of available states] we are left with the dynamical part of the LESW [DLESW: $\Lambda_D(x)=\Lambda(x)-2x$], due to the kinetic energy of the elec-

trons. In Fig. 4 the DLESW for the MH system is shown as well as $\Lambda_D(x)=1-x$, corresponding with the free-particle limit and $\Lambda_D(x)=-x$, corresponding with the semiconductor.

Also shown are the results of Harris and Lange[26] for two values of t (-0.5 and -2.0 eV). They worked out a first-order expression in t/U for the spectral weights in the lower and upper Hubbard bands. For the total integrated (i.e., PES and IPES) spectral weight of the LHB (which is the zeroth-order moment $m^0_{\rm LHB}$) they found

$$m^0_{\rm LHB}=\sum_\sigma \left[1-\langle c^\dagger_{i\bar\sigma}c_{i\bar\sigma}\rangle-\frac{2t}{N}\sum_{\langle ij\rangle}\langle c^\dagger_{i\bar\sigma}c_{j\bar\sigma}\rangle\right], \quad (12)$$

where $\langle ij\rangle$ runs over all nearest-neighbor pairs. Making use of the sum rules and assuming a homogeneous charge distribution, it is now a straightforward exercise to derive from Eq. (12) the LESW:

$$\Lambda(x)=2x-\frac{2t}{N}\sum_{\langle ij\rangle,\sigma}\langle c^\dagger_{i\bar\sigma}c_{j\bar\sigma}\rangle . \quad (13)$$

The first term describes the LESW for the MH model in the localized limit [$\Lambda(x)=2x$]. The second term is always positive and proportional to the expectation value of the kinetic energy. To calculate this expectation value, one has to use the zeroth-order ($U\to\infty$) wave functions, which are equivalent to the spinless-fermion wave function. They are given by filling up the one-particle momentum levels [$\epsilon_k=2t\cos(ka)$] with only one electron per state up to the Fermi wave vector. Doing so, we find a simple expression for the first order in t/U contribution to the DLESW:

$$\Lambda_D(x)=\frac{4|t|}{\pi U}\sin(\pi x) . \quad (14)$$

From the figure we see that this first-order term (proportional to the kinetic energy) gives a good approximation to the DLESW for small t as expected. However, the Harris-Lange result is symmetric around $x=0.5$. This is quite different from the exact diagonalizations for larger $|t|$ values where we find that this dynamical part is highly asymmetric and that the maximum of the DLESW lies somewhere in the doping range $x<0.4$. For small doping concentration the DLESW obtained from the exact calculations is substantially larger than that from the first-order approximation.

An often used approximation for the MH model is the Hubbard I decoupling scheme, proposed by Hubbard.[18] This approximation yields two poles for every momentum and spin with energies and pole strengths dependent on the occupation number in the ground state. The energies $E^{(1)}_{k\sigma}$ and $E^{(2)}_{k\sigma}$ ($E^{(1)}_{k\sigma}<E^{(2)}_{k\sigma}$) are given by

$$E^{(1),(2)}_{k\sigma}=\tfrac{1}{2}(U+\epsilon_k)\pm\tfrac{1}{2}\sqrt{(U-\epsilon_k)+4n_{-\sigma}U\epsilon_k} , \quad (15)$$

where ϵ_k are the one-particle energies (the center of the band is chosen to be 0). The pole strength for each pole can be found by differentiating the pole energies with respect to ϵ_k; thus,

$$A^{(1),(2)}_{k\sigma}=\frac{\partial E^{(1),(2)}_{k\sigma}}{\partial\epsilon_k} . \quad (16)$$

If we put the occupation number in the ground state for the up spins equal to the occupation number for the down spins it is straightforward to calculate the DLESW by filling the momentum energy levels up to the chemical potential, weighted by the pole strengths. The results are the solid lines ($t=0.5$, 1, and 2 eV) plotted in Fig. 4. The open squares are obtained by substituting the occupation numbers for the up and down spins as obtained from the exact diagonalized ground state of the 10-site cluster. From Fig. 4 it can be seen that the dynamical contribution to the LESW is negative, contrary to what is found in the exact diagonalization. As far as the LESW is concerned, the Hubbard I approximation has the tendency to go to the situation of doping a semiconductor, while the exact diagonalization has the tendency to go to the right free-particle limit for large hole or electron doping away from half filling.

In Figs. 5 and 6 we plot the derivative of the LESW with respect to the doping concentration x and the hybridization strength t, respectively. We again see that the LESW is a strong function of the hybridization t and doping concentration x. This means that the effective number of degrees of freedom in the low-energy regime (responsible for the low-energy physics) is strongly dependent on the hybridization and therefore on the volume and temperature of the system.

It is interesting to see what will happen with the electrons if the volume of a system is changed. Suppose we contract one half of a material, which results in an in-

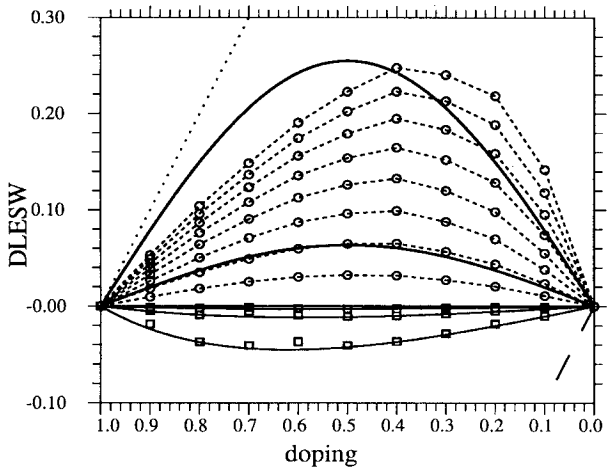


FIG. 4. The dynamical part of the LESW [$\Lambda_D(x)=\Lambda(x)-2x$] as a function of (hole) doping concentration x_h for the MH model. The dotted line corresponds to the free-particle limit (hole doped) and the dashed line to the semiconductor. The data points (○) are from the calculations: $U=10$, $t=-0.25$ eV (lowest) to $U=10$, $t=-2$ eV (highest) in steps of 0.25 eV. The first order in t/U results are shown in the upper part of the figure for $U=10$, $t=-0.5$ (lower solid curve) and $t=-2.0$ eV (upper solid curve). Also shown are results of the Hubbard I approximation in the lower part of the figure. The data points (□) correspond to the Hubbard I decoupling using ground state occupation numbers.

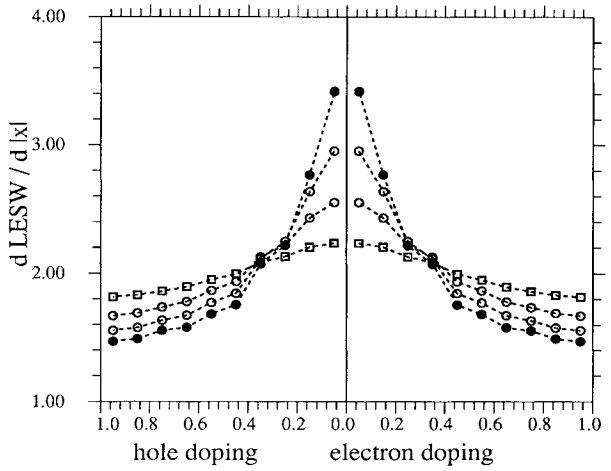


FIG. 5. The derivative of the LESW (see Fig. 3) with respect to the doping concentration for the MH model ($U=10$ eV): $t=-0.5$ (□) to $t=-2$ (●) in the steps of 0.5 eV.

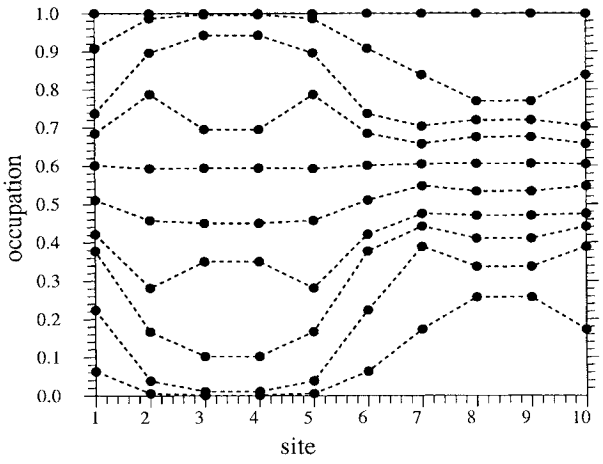


FIG. 7. The number of electrons per site for a one-dimensional MH chain ($U=10$ eV) for different values of the (hole-) doping concentration x_h. The lowest corresponds to $x=0.9$ and the highest to the undoped case ($n=1$, for all sites). The hybridization strength in the left part of the chain is 0.5 eV, while in the right part of the chain it is 1.0 eV.

crease of the hybridization in this region. If this material can be described by an independent particle picture, the electrons will go to that part of the sample with the largest hybridization as long as the band is less than half filled. Thus, if we are dealing with a hole-doped system, the number of electrons in the part with the largest volume will be lower than that in the part with smaller volume. However, if we are dealing with a correlated system the situation is really different. This is illustrated in Fig. 7. We consider a 10-site MH cluster. The Coulomb repulsion on every site is again 10 eV and the hybridization in the "left" part (sites 1–5) is 0.5 eV, while the hybridization in the right part of the cluster (sites 6–10) is 1.0 eV. We plotted the number of electrons per site for different values of the doping concentration. The lowest curve corresponds to a doping concentration $x=0.9$ and the upper curve to half filling.

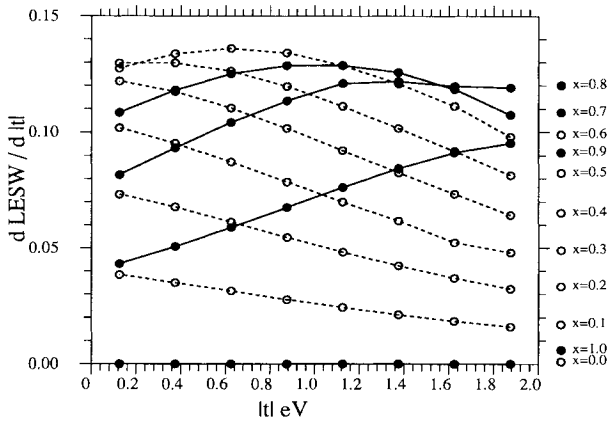


FIG. 6. The derivative of the LESW (see Fig. 3) with respect to the hybridization for the MH model ($U=10$ eV). The doping concentrations are indicated.

From the figure it can be seen that for small electron concentrations (large doping concentrations) the situation resembles more or less the independent particle limit where the electrons tend to go to the part with the largest hybridization. However, this picture changes when we approach a critical doping concentration of roughly 0.4. At this point the electrons are distributed homogeneously. When we decrease the doping concentration still further the role of electrons and holes are interchanged. For low hole doping concentration, the holes tend to go to the region with largest hybridization and not the electrons.

This may be compared with the electron-hole distribution in the $U\rightarrow\infty$ ground state. This is the spinless fermion ground state and is constructed by filling up the one-particle moment levels with only one electron. The one-dimensional band structure for spinless fermions for this system consists of two bands, one with high dispersion (corresponding with the right side of the sample) and one with lower dispersion (corresponding with the left side). Upon filling the system, we start filling up the band with highest dispersion until the one-particle energy level reaches the bottom of the other hand. Until this point is reached, all electrons go to the right side of the sample, where they gain most kinetic energy. From here, we fill up both bands. At exactly quarter filled, the number of electrons at both sides are equal, after which the role of electrons and holes are interchanged. This turning point can be seen to coincide with the critical doping concentration for which a maximum in the DLESW is obtained. This critical point moves towards lower doping concentrations when t/U is increased.

We also studied the effect of the cluster size as well as the dimensionality of the system on the LESW. We performed calculations on one-dimensional six-, eight-, and ten-site as well as for two-dimensional nine- and ten-site MH clusters. In order to be able to compare the one-

dimensional and two-dimensional systems, we adjust the hybridization in such a way that the same effective bandwidth is obtained. We define the one particle effective bandwidth as the root mean square value of the one particle energies, which equals the second moment $(zt^2$, with z the number of nearest neighbors, thus $t_{1D}=\sqrt{2}t_{2D}$). The DLESW for the one- ($t=-1$ eV) and two-dimensional ($t=-\sqrt{1/2}$ eV) clusters are shown in Fig. 8. For the one-dimensional systems, no substantial differences in the DLESW can be seen, from which we may conclude that the cluster size is of no qualitative importance as far as the DLESW is concerned. Also the two-dimensional clusters compare very well with the one-dimensional clusters provided the one-particle effective bandwidth is kept constant. From this comparison we may conclude that the phenomenon of spectral weight transfer is qualitatively and quantitatively independent of cluster size and dimensionality.

B. Charge-transfer model

It can be shown that the periodic two-dimensional Cu_4O_8 cluster, studied before as representing the CuO_2 planes of the cuprates,[20,27–29] is equivalent to the periodic one-dimensional Cu_4O_4 cluster if the onsite O-O repulsion is neglected (as we do). Because in the Cu_4O_8 cluster there are four pairs of equivalent oxygens, one can take linear combinations of these pairs. One linear combination does not couple to the Cu sites and forms nonbonding levels. Therefore, the two-dimensional Cu_4O_8 cluster reduces to a one-dimensional Cu_4O_4 chain with $\alpha_{jj'}=a_{ij}=0$ in Eq. (9), and $t_{pd,1D}=\sqrt{2}t_{pd,2D}$ and $t_{pp,1D}=-2t_{pp,2D}$. We perform calculations on the one-dimensional four-unit-cell (eight site) CT model (Cu_4O_4) with a copper on-site repulsion $U_{dd}=8$ eV, charge-transfer energy $\Delta=\epsilon_p-\epsilon_d=4$ eV, and oxygen-oxygen hybridization $t_{pp}=-0.25$ eV. The above relations for the hybridizations can be used to translate our results to the two-dimensional Cu_4O_8 cluster.

The total electron removal and addition spectra for the CT system with $t_{pd}=1$ eV and for different hole fillings are plotted in Fig. 9. The oxygen spectral weight is represented by the dotted line and the copper spectral weight by the solid line. The number of holes N in the ground state is indicated. The insulating CT system corresponds to the spectrum with four holes in the ground state. Thus the hole doping concentration is given by $x_h=(N-\mathcal{N})/\mathcal{N}(N\geq\mathcal{N})$ with $\mathcal{N}=4$ the number of unit cells. The electron doping concentration is given by $x_e=(\mathcal{N}-N)/\mathcal{N}(N\leq\mathcal{N})$. As in the MH case we find that the chemical potential shifts by just the amount of the insulating gap when we go from the one hole to the one-electron-doped system.

The LESW, obtained by integrating the d and p spectral weights over the shaded area, is shown in Fig. 10 for different values of t_{pd}. Again, we only focus on those values of t_{pd}, which ensures a sensible definition of ω_g. From the figure it can seen that the LESW as a function of the hybridization for electron-doped CT system

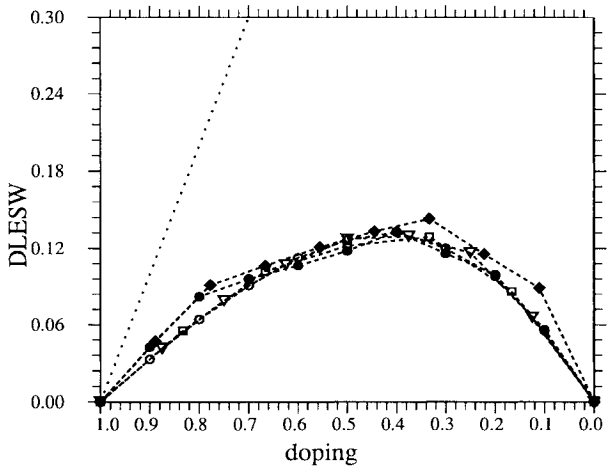


FIG. 8. The dynamical part of the LESW $[\Lambda_D(x)=\Lambda(x)-2x]$ as a function of (hole-) doping concentration x_h for the MH model for the one-dimensional ten-site (○), eight-sites (▽), and six-site (□) clusters with $U=10$ and $t=-1$ eV and two-dimensional ten-site (●) and nine-site (■) clusters with $U=10$ and $t=-0.71$ eV.

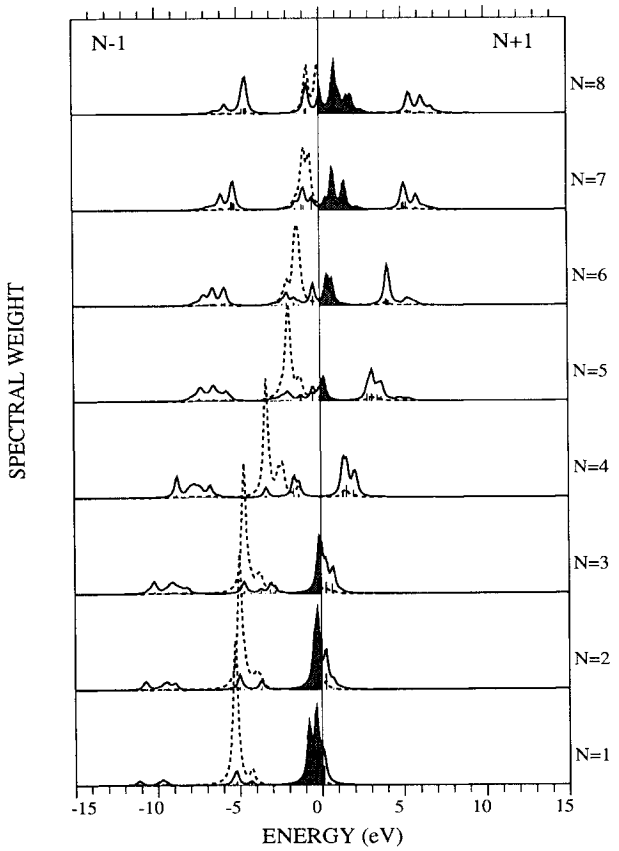


FIG. 9. One-particle Green's function for a one-dimensional four-unit-cell CT model. The dotted lines correspond to the p spectral weight and the solid lines to the d spectral weight. $U_{dd}=8$ eV, $t_{pd}=1$, $t_{pp}=-0.25$, and $\Delta=4$ eV. The number of holes in the ground state N are indicated. The LESW is obtained by integration over the shaded area.

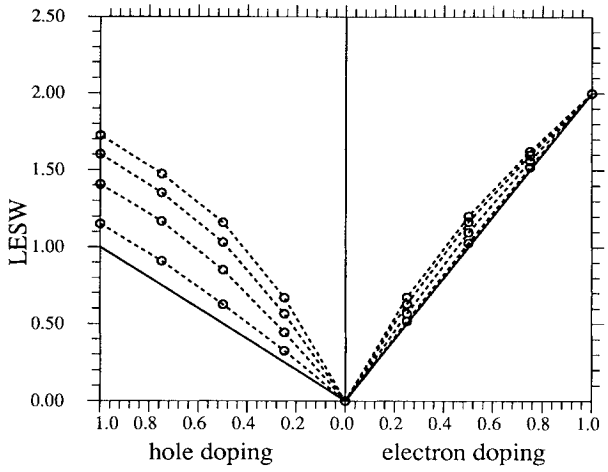


FIG. 10. The integrated total-LESW divided by the number of sites as a function of the doping concentration x for the one-dimensional four-unit-cell CT system ($U_{dd}=8$ eV), $t_{pp}=0.25$, $\Delta=4$ eV). The solid lines correspond to the localized limit $t_{pd}=0$. The data points are from the calculations: $t_{pd}=0.5$ eV (lowest) to $t_{pd}=2$ eV in steps of 0.5 eV.

behaves more or less the same as found for the MH system. However, for the hole-doped case, the situation is quite different. For small hybridization between the free-particle-like oxygen orbitals and correlated copper orbitals it is found that the LESW behaves semiconductorlike and every added hole adds a weight 1 to the total spectrum. But when the hybridization is increased, the LESW for the hole-doped CT system rapidly increases and the LESW becomes almost symmetric with respect to hole or electron doping for $t_{pd}=-2$ eV. This can also be seen from the derivative of the LESW with respect to the doping concentration as plotted in Fig. 11, which also shows a more or less symmetric figure with respect to hole and electron doping for large hybridization. It is interesting to note that the high-T_c superconductors lie in the regime with large hybridization, so the holes in the hole-doped high T_c's will behave as strongly correlated particles.

Due to the hybridization between the correlated Cu and free-electron-like O orbitals a sort of mirror LHB appears at the low-energy side of the oxygen band and the d spectral weight is strongly enhanced. This can also be seen in Fig. 9. Holes in this sort of Zhang-Rice singlet band[30] will behave as strongly correlated particles with restriction on double occupancy. This is also shown by Feiner,[31] who projected the CT model onto CuO_4 cell-eigenstates. This permits a description of the low-energy physics by an effective MH model with an effective U in the order of Δ. The set of cell eigenstates is a reduced set of eigenstates of a CuO_4 cluster, namely the zero-hole vacuum state, one-hole state (linear combination of p and d) and the two-hole Zhang-Rice singlet. The LESW can now be described as a sum of intracellular [result of the relation between the physical particles (holes) and the effective particles (cell-eigenstates)] and intercellular contributions (result of an effective MH model working on the effective particles). The intracell contribution is linear with doping. For electron doping it goes as $2x_e$, while for hole doping the intracell LESW equals cx_h with c between 1 ($t_{pd}=0$) and 2 ($t_{pd}\rightarrow\infty$). It turned out that the prefactor c, which reflects the internal hybridized nature of the cell states and can be calculated from the model parameters, is equal to the LESW at $x_h=1$, as found in the exact diagonalizations. The further increase of the LESW results from the intercell contributions and is shown in Fig. 12. This part may be seen as the kinetic

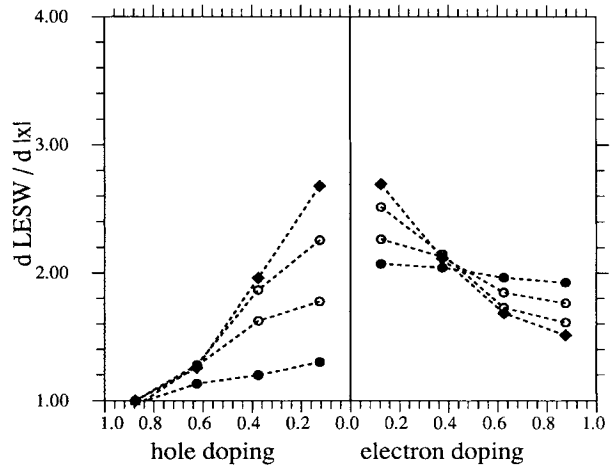


FIG. 11. The derivative of the LESW (see Fig. 10) with respect to the doping concentration for the CT model ($U_{dd}=8$, $t_{pp}=-0.25$, $\Delta=4$): $t_{pd}=0.5$ (□) to $t_{pd}=2$ eV (●) in steps of 0.5 eV.

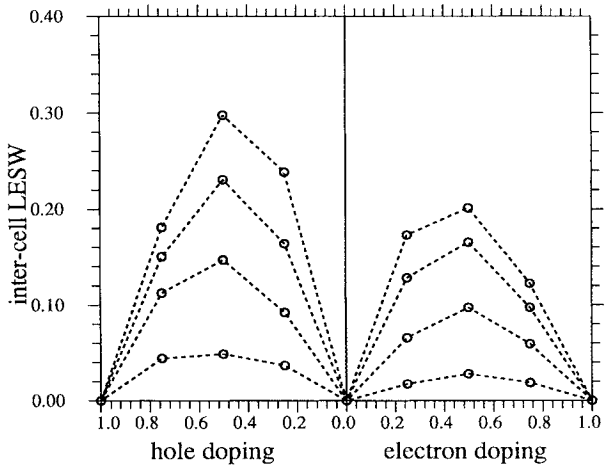


FIG. 12. The intercell contributions (see Ref. 31) to the LESW (see Fig. 10) with respect to the doping concentration for the CT model ($U_{dd}=8$, $t_{pp}=0.25$, $\Delta=4$). The intercell contribution equals LESW$-2x_e$ for electron doping and LESW$-cx_h$ for hole doping with c describing the internal hybridized nature of the CuO_4 cell eigenstates. The data points are from the calculations: $t_{pd}=0.5$ eV (lowest) to $t_{pd}=2$ eV in steps of 0.5 eV.

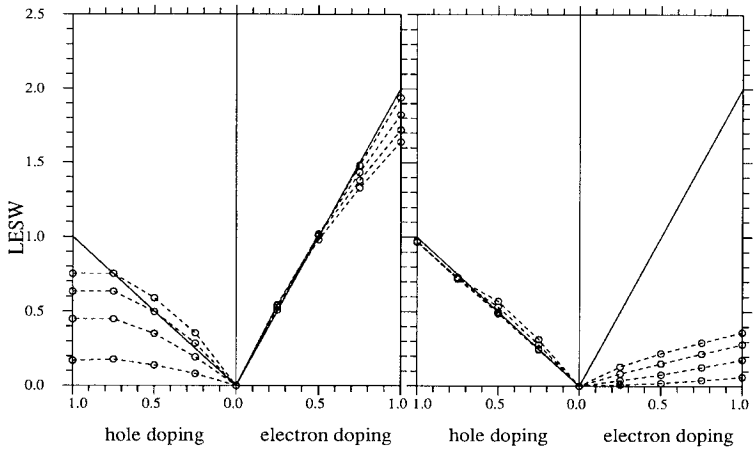


FIG. 13. The integrated d-LESW (left) and p-LESW (right) divided by the number of sites as a function of the doping concentration x for the one-dimensional four-unit-cell CT system ($U_{dd}=8$ eV, $t_{pp}=0.25$, $\Delta=4$ eV). The solid lines correspond to the localized limit $t_{pd}=0$. The data points are from the calculations: $t_{pd}=0.5$ eV (lowest) to $t_{pd}=2$ eV in steps of 0.5 eV.

effect on the LESW due to the *effective* MH model comparable to the DLESW as found in the preceding section. However, because the t_{pd}-dependent prefactor c already contains some kinetic effect it is not that straightforward to define a DLESW as in the MH case.

In Fig. 13 the d part (left) and p (right) part of the LESW is shown. It is interesting to see that for a hole-doped CT system the dynamical part of the LESW, due to the hybridization between neighboring oxygen and copper orbitals, is almost entirely due to a spectral weight transfer from the d-like UHB to the ligand band near the Fermi level, while no extra p spectral weight is observed in the UHB. This can be seen from the fact that the p-LESW coincides more or less with the $\Lambda(x_h)=x_h$ curve, corresponding with hole doping the CT system without intersite hybridization. For the electron doped CT system and for low doping concentrations, the situation is reversed. Now the dynamical contribution to the LESW is mostly due to extra p spectral weight while the d-LESW follows the $\Lambda(x_e)=2x_e$ curve, corresponding with electron doping the CT system in the localized limit.

The effect of the cluster size on the LESW is also studied. In Fig. 14 we show the LESW for three-, four-, and five-unit-cell CT clusters with $U_{dd}=8$, $t_{pp}=-0.25$, $\Delta=4$, and $t_{pd}=1$ eV. From the figure it can be seen that, as in the MH case, the size of the cluster is of little or no importance to the LESW.

IV. DISCUSSION AND CONCLUSIONS

For the Mott-Hubbard (MH) system, in which the low-energy spectral weight (LESW) is symmetric with respect to hole or electron doping, we found that the LESW grows faster than two times the amount of doping when the hybridization is taken into account. The LESW may be divided into two parts, the static LESW (which is just the result of counting the available states) and a dynamical part (DLESW), which is the kinetic contribution. The DLESW is strongly dependent on the intersite hybridization and on the amount of doping but independent on cluster size and dimensionality. We could identify a sort of critical doping concentration for which the DLESW reaches its maximum. This critical concentration also seems to be related to the turn over from electron to hole-like behavior. This turn over was studied by looking at the lattice parameter dependence of the electron density. For a nearly empty band in a MH model, electrons move towards regions with smaller lattice parameters and larger bandwidths, whereas at an electron concentration above the critical concentration of roughly 0.6–0.7 (doping concentration 0.4–0.3) the electrons move towards the lower bandwidth regions. For large doping concentrations the situation is similar to that of an independent particle picture where electrons like to be in that part where they gain the most kinetic energy.

The CT system shows asymmetric behavior between electron and hole doping when the intersite copper-oxygen hybridization is small. Holes residing on oxygen now occupy almost free-particle levels and scatter weakly off the copper spins. In this region it is therefore not possible to reduce the problem to a (strongly correlated) single-band Hubbard or t-J model. However, when the

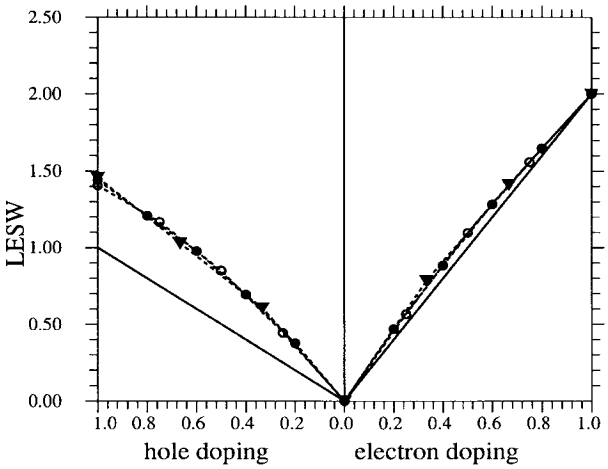


FIG. 14. The LESW as a function of the doping concentration for the CT model for the one-dimensional five-unit-cell (●), five-unit-cell (○) and three-unit-cell (▽) clusters with $U_{dd}=8$, $t_{pp}=0.25$, $\Delta=4$, and $t_{pd}=1$ eV.

hybridization is large, as is the case in the high T_c's the LESW becomes similar to that of the MH system and the electrons as well as the holes show correlated behavior. In this case Cu and O degrees of freedom can no longer be separated in zeroth order. Furthermore, it is shown that the LESW is independent of the cluster size, as for the MH systems, and that the hybridization dependent enhancement of the LESW for hole doping is almost entirely due to the Cu d spectral weight, while for small electron doping concentrations it is almost entirely due to the O p spectral weight.

In mean-field theories, where one describes the eigenstates by a single Slater determinant, the LESW will always be linear with respect to the doping concentration, because the oscillator strength will just be 1 or 0, since spectral weight is now identical to counting states. Unrestricted Hartree-Fock can reproduce the $U \rightarrow \infty$ behavior of the MH model if the holes localize. Linear behavior of the LESW with respect to doping is also found for the t-J model, which behaves the same as the Hubbard model in the localized limit for all values of t and J. The Hubbard I decoupling approximation gives oscillator strengths, which need not be 1 or 0. However, the DLESW calculated in this approximation yields a negative contribution to the total LESW instead of a positive one. The Harris and Lange first order in t/U theory shows that for small t/U the DLESW is proportional to the kinetic-energy expectation value. However, for larger t/U values and for low doping concentrations the first-order expression highly underestimates the DLESW.

The strong dependency of the LESW on the hybridization and doping concentration has interesting consequences for the model Hamiltonians used to describe the low-energy physics. This is because of the fact that the number of effective particles in the low-energy regime is a function of the hybridization and therefore of the volume, temperature and electron-phonon interaction. Therefore, it may well be impossible to define a low-energy-scale Hamiltonian, which describes temperature and pressure dependent properties unless one introduces a kind of effective nonparticle conservation and fractional statistics.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with D. I. Khomskii, M. J. Rice, A. M. Oleś, and L. F. Feiner. This investigation was supported by the Netherlands Foundation for Chemical Research (SON) and the Netherlands Foundation for Fundamental Research on Matter (FOM) with financial support from the Netherlands Organization for the Advancement of Pure Research (NWO).

[1]E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968).
[2]P. W. Anderson and Y. Ren, *Proceedings of the Los Alamos Symposium on High Temperature Superconductivity,* edited by K. S. Bedell *et al.* (Addison-Wesley, Reading, MA, 1989).
[3]H. J. Schultz, Phys. Rev. Lett. **64**, 2831 (1990).
[4]J. B. Grant and A. K. McMahan, Phys. Rev. B **46**, 8440 (1992).
[5]J. Lorenzana, Ph.D. thesis, International School for Advanced Studies, Trieste, 1992.
[6]V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987).
[7]K. A. Chao, J. Spatek, and A. M. Oleś, J. Phys. C **10**, 2271 (1977).
[8]C. Castellani *et al.*, Phys. Rev. Lett. **69**, 2009 (1992).
[9]F. D. M. Haldane, Phys. Rev. Lett. **67**, 937 (1991).
[10]C. T. Chen *et al.*, Phys. Rev. Lett. **66**, 104 (1991).
[11]H. Romberg *et al.*, Phys. Rev. B **42**, 8768 (1990).
[12]P. Kuiper *et al.*, Phys. Rev. Lett. **62**, 221 (1989).
[13]S. L. Cooper *et al.*, Phys. Rev. B **41**, 11 605 (1990).
[14]H. Eskes, M. B. J. Meinders, and G. A. Sawatzky, Phys. Rev. Lett. **67**, 1035 (1991).
[15]M. S. Hybertsen, E. B. Stechel, W. M. C. Foulkes, and M. Schlüter, Phys. Rev. B **45**, 10 032 (1992).
[16]Y. Ohta *et al.*, Phys. Rev. B **46**, 14 022 (1992).
[17]P. Unger and P. Fulde (unpublished).
[18]J. Hubbard, Proc. R. Soc. London, Ser. A **276**, 238 (1968).
[19]J. Zaanen, G. A. Sawatzky, and J. W. Allen, Phys. Rev. Lett. **55**, 418 (1985).
[20]E. Dagotto *et al.*, Phys. Rev. B **41**, 9049 (1990).
[21]I. Sega and P. Prelovšek, Phys. Rev. B **42**, 892 (1990).
[22]H. Matsumoto, M. Sasaki, and M. Tachiki, Solid State Commun. **71**, 829 (1989).
[23]M. A. van Veenendaal, R. Schlatmann, G. A. Sawatzky, and W. A. Groen, Phys. Rev. B **47**, 446 (1993).
[24]J. W. Allen *et al.*, Phys. Rev. Lett. **64**, 595 (1990).
[25]See, e.g., O. Gunnarsson and K. Schönhammer, Phys. Rev. B **28**, 4315 (1983).
[26]A. B. Harris and R. V. Lange, Phys. Rev. **157**, 259 (1967).
[27]A. Moreo and E. Dagotto, Phys. Rev. B **42**, 4786 (1990).
[28]H. J. Schmidt and Y. Kuramoto, Phys. Rev. B **42**, 2562 (1990).
[29]E. R. Gagliano, C. A. Balseiro, and M. Avignon, Europhys. Lett. **12**, 259 (1990).
[30]F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759 (1988).
[31]L. F. Feiner (unpublished).

R. SHANKAR

The renormalization group and Fermi liquids

R.Shankar
Sloane Physics Lab
Yale University
New Haven CT 06520

Chapter 1

1.1 The RG: what, why and how

Imagine that you have some problem in the form of a partition function

$$Z(a,b) = \int dx \int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \tag{1.1}$$

where a, b are the parameters.

First consider $b = 0$, *the gaussian model*. Suppose that you are just interested in x, say in its fluctuations. Then you have the option of integrating out y and working with the new partition function

$$Z(a) = N \int dx e^{-ax^2} \tag{1.2}$$

where N comes from doing the y-integration. We will ignore such an x-independent prefactor here and elsewhere since it will cancel in any averaging process.

Consider now the nongaussian case with $b \neq 0$. Here we have

$$\begin{aligned} Z(a', b'...) &= \int dx \left[\int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \right] \\ &\equiv \int dx e^{-a' \ x^2} e^{-b'x^4 - c'x^6 + ...} \end{aligned} \tag{1.3}$$

where a', b' etc., define the parameters of the effective field theory for x. These parameters will reproduce exactly the same averages for x as the original ones. *This evolution of parameters with the elimination of uninteresting degrees of freedom, is what we mean these days by renormalization, and as such has nothing to do with infinities; you just saw it happen in a problem with just two variables.*

The parameters b, c etc., are called *couplings* and the monomials they multiply are called *interactions*. The x^2 term is called the *kinetic* or *free-field* term.

Notice that to get the effective theory we need to do a nongaussian integral. This can only be done perturbatively. At the simplest *tree Level*, we simply drop y and find $b' = b$. At higher orders, we bring down the nonquadratic exponential and integrate in y term by term and generate effective interactions for x. This procedure can be represented by Feynman graphs in which variables in the loop are limited to the ones being eliminated.

Why do we do this? Because certain tendencies of x are not so apparent when y is around, but surface to the top, as we zero in on x. For example, we are going to consider a problem in

which x stands for low-energy variables and y for high energy variables. Upon integrating out high energy variables a numerically small coupling can grow in size (or initially impressive one diminish into oblivion), as we zoom in on the low energy sector.

This notion can be made more precise as follows. Consider the gaussian model in which we have just $a \neq 0$. We have seen that this value does not change as y is eliminated since x and y do not talk to each other. This is called a *fixed point of the RG*. Now turn on new couplings or "interactions" (corresponding to higher powers of x, y etc.) with coefficients b, c and so on. Let a', b' etc., be the new couplings after y is eliminated. The mere fact that $b' > b$ does not mean b is more important for the physics of x. This is because a' could also be bigger than a. So we rescale x so that the kinetic part, x^2, has the same coefficient as before. If the quartic term still has a bigger coefficient, (still called b'), we say it is a *relevant* interaction. If $b' < b$ we say it is irrelevant. This is because in reality y stands for many variables, and as they are eliminated one by one, the coefficient of the quartic term will run to zero. If a coupling neither grows not shrinks it is called *marginal*.

There is another excellent reason for using the RG, and that is to understand the phenomenon of universality in critical phenomena. I must regretfully pass up the opportunity to explain this and refer you to Reference [1].

We will now see how this method is applied to interacting fermions in $d = 2$. Later we will apply these methods to quantum dots.

1.2 The problem of interacting fermions

Consider a system of nonrelativistic spinless fermions in two space dimensions. The one particle hamiltonian is

$$H = \frac{K^2}{2m} - \mu \tag{1.4}$$

where the chemical potential μ is introduced to make sure we have a finite density of particles in the ground state: all levels up the Fermi surface, a circle defined by

$$K_F^2/2m = \mu \tag{1.5}$$

are now occupied and occupying these levels lowers the ground-state energy.

Notice that this system has gapless excitations above the ground state. You can take an electron just below the Fermi surface and move it just above, and this costs as little energy as you please. Such a system will carry a dc current in response to a dc voltage. An important question one asks is if this will be true when interactions are turned on. For example the system could develop a gap and become an insulator. What really happens for the $d = 2$ electron gas?

We are going to answer this using the RG. Let us first learn how to do RG for noninteracting fermions. To understand the low energy physics, we take a band of of width Λ on either side of the Fermi surface. This is the first great difference between this problem and the usual ones in relativistic field theory and statistical mechanics. Whereas in the latter examples low energy means small momentum, here it means small deviations from the Fermi surface. Whereas in these older problems we zero in on the origin in momentum space, here we zero in on a surface. The low energy region is shown in Figure 1.1.

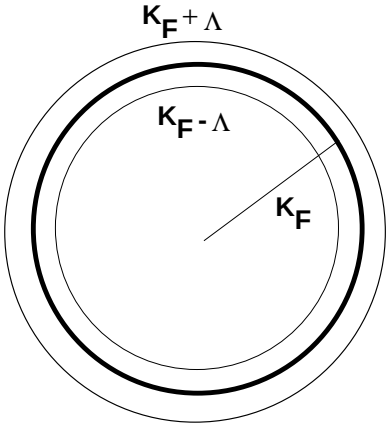


Figure 1.1: The low energy region for nonrelativistic fermions lies within the annulus concentric with the Fermi circle.

To apply our methods we need to cast the problem in the form of a path integral. Following any number of sources, say [2] we obtain the following expression for the partition function of free fermions:

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{1.6}$$

where

$$S_0 = \int d^2K \int_{-\infty}^{\infty} d\omega \overline{\psi}(\omega, \mathbf{K}) \left(i\omega - \frac{(K^2 - K_F^2)}{2m} \right) \psi(\omega, \mathbf{K}) \tag{1.7}$$

where ψ and $\overline{\psi}$ are called Grassmann variables. They are really weird objects one gets to love after some familiarity. Most readers can assume they are ordinary integration variables. The dedicated reader can learn more from Ref. [2].

We now adapt this general expression to the annulus to obtain

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{1.8}$$

where

$$S_0 = \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} d\omega \int_{-\Lambda}^{\Lambda} dk \overline{\psi}(i\omega - v\ k)\psi. \tag{1.9}$$

To get here we have had to approximate as follows:

$$\frac{K^2 - K_F^2}{2m} \simeq \frac{K_F}{m} \cdot k = v_F\ k \tag{1.10}$$

where $k - K - K_F$ and v_F is the fermi velocity, hereafter set equal to unity. Thus Λ can be viewed as a momentum or energy cut-off measured from the Fermi circle. We have also replaced KdK by $K_F dk$ and absorbed K_F in ψ and $\overline{\psi}$. It will seen that neglecting k in relation to K_F is irrelevant in the technical sense.

Let us now perform mode elimination and reduce the cut-off by a factor s. Since this is a gaussian integral, mode elimination just leads to a multiplicative constant we are not

interested in. So the result is just the same action as above, but with $|k| \leq \Lambda/s$. Let us now do make the following additional transformations:

$$\begin{aligned} (\omega', k') &= s(\omega, k) \\ (\psi'(\omega', k'), \overline{\psi}'(\omega', k')) &= s^{-3/2}(\psi(\frac{\omega'}{s}, \frac{k'}{s}), \overline{\psi}(\frac{\omega'}{s}, \frac{k'}{s})). \end{aligned} \tag{1.11}$$

When we do this, the action and the phase space all return to their old values. So what? Recall that our plan is to evaluate the role of quartic interactions in low energy physics as we do mode elimination. Now what really matters is not the absolute size of the quartic term, but its size relative to the quadratic term. Keeping the quadratic term identical before and after the RG action makes the comparison easy: if the quartic coupling grows, it is relevant; if it decreases, it is irrelevant, and if it stays the same it is marginal. The system is clearly gapless if the quartic coupling is irrelevant. Even a marginal coupling implies no gap since any gap will grow under the various rescalings of the RG.

Let us now turn on a generic four-Fermi interaction in path-integral form:

$$S_4 = \int \overline{\psi}(4)\overline{\psi}(3)\psi(2)\psi(1)u(4,3,2,1) \tag{1.12}$$

where $\int$ is a shorthand:

$$\int \equiv \prod_{i=1}^{3} \int d\theta_i \int_{-\Lambda}^{\Lambda} dk_i \int_{-\infty}^{\infty} d\omega_i \tag{1.13}$$

At the tree level, we simply keep the modes within the new cut-off, rescale fields, frequencies and momenta , and read off the new coupling. We find

$$u'(k', \omega', \theta) = u(k'/s, \omega'/s, \theta) \tag{1.14}$$

This is the evolution of the coupling function. To deal with coupling constants with which we are more familiar, we expand the functions in a Taylor series (schematic)

$$u = u_o + ku_1 + k^2 u_2 ... \tag{1.15}$$

where k stands for all the k's and ω's. An expansion of this kind is possible since couplings in the Lagrangian are nonsingular in a problem with short range interactions. If we now make such an expansion and compare coefficients in Eqn. (1.14), we find that u_0 is marginal and the rest are irrelevant, as is any coupling of more than four fields. Now this is exactly what happens in ϕ_4^4, scalar field theory in four dimensions with a quartic interaction. The difference here is that we still have dependence on the angles on the Fermi surface:

$$u_0 = u(\theta_1, \theta_2, \theta_3, \theta_4)$$

Therefore in this theory we are going to get coupling functions and not a few coupling constants.

Let us analyze this function. Momentum conservation should allow us to eliminate one angle. Actually it allows us more because of the fact that these momenta do not come form

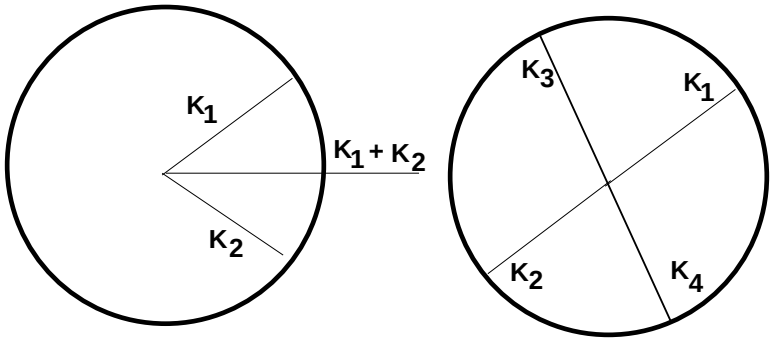


Figure 1.2: Kinematical reasons why momenta are either conserved pairwise or restricted to the BCS channel.

the entire plane, but a very thin annulus near K_F. Look at the left half of Figure 1.2. Assuming that the cutoff has been reduced to the thickness of the circle in the figure, it is clear that if two points 1 and 2 are chosen from it to represent the incoming lines in a four point coupling, the outgoing ones are forced to be equal to them (not in their sum, but individually) up to a permutation, which is irrelevant for spinless fermions. Thus we have in the end just one function of two angles, and by rotational invariance, their difference:

$$u(\theta_1, \theta_2, \theta_1, \theta_2) = F(\theta_1 - \theta_2) \equiv F(\theta). \tag{1.16}$$

About forty years ago Landau came to the very same conclusion[3] that a Fermi system at low energies would be described by one function defined on the Fermi surface. He did this without the benefit of the RG and for that reason, some of the leaps were hard to understand. Later detailed diagrammatic calculations justified this picture [4]. The RG provides yet another way to understand it. It also tells us other things, as we will now see.

The first thing is that the final angles are not slaved to the initial ones if the former are exactly opposite, as in the right half of Figure 1.2. In this case, the final ones can be anything, as long as they are opposite to each other. This leads to one more set of marginal couplings in the BCS channel, called

$$u(\theta_1, -\theta_1, \theta_3, -\theta_3) = V(\theta_3 - \theta_1) \equiv V(\theta). \tag{1.17}$$

The next point is that since F and V are marginal at tree level, we have to go to one loop to see if they are still so. So we draw the usual diagrams shown in Figure 3. We eliminate an infinitesimal momentum slice of thickness $d\Lambda$ at $k = \pm\Lambda$.

These diagrams are like the ones in any quartic field theory, but each one behaves differently from the others and its its traditional counterparts. Consider the first one (called ZS) for F. The external momenta have zero frequencies and lie of the Fermi surface since ω and k are irrelevant. The momentum transfer is exactly zero. So the integrand has the following schematic form:

$$\delta F \simeq \int d\theta \int dk d\omega \left(\frac{1}{(i\omega - \varepsilon(K))} \frac{1}{(i\omega - \varepsilon(K))} \right) \tag{1.18}$$

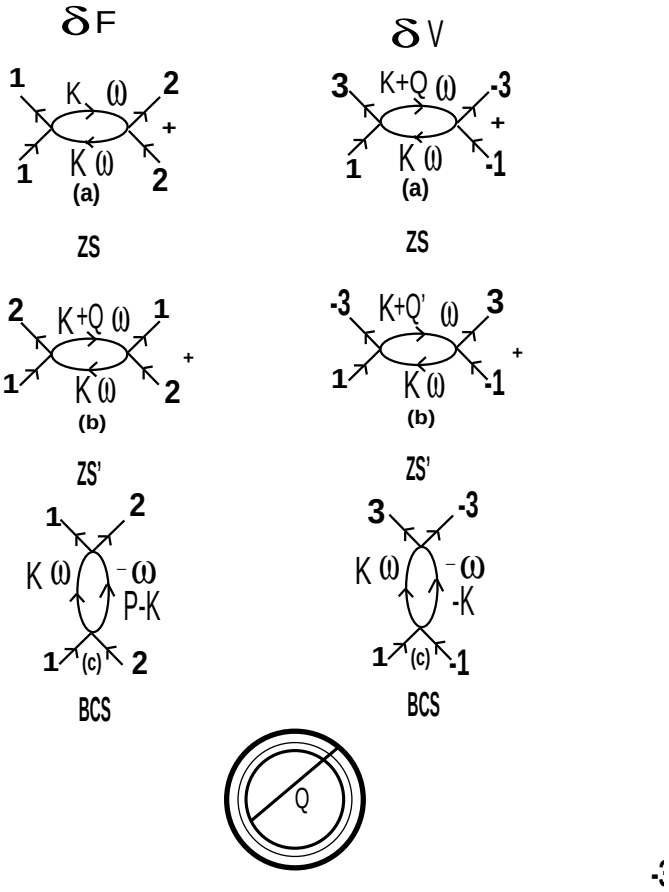


Figure 1.3: One loop diagrams for the flow of F and V. The last at the bottom shows that a large momentum Q can be absorbed only at two particular initial angles (only one of which is shown) if the final state is to lie in the shell being eliminated.

The loop momentum K lies in one of the two shells being eliminated. Since there is no energy difference between the two propagators, the poles in ω lie in the same half-plane and we get zero, upon closing the contour in the other half-plane. In other words, this diagram can contribute if it is a particle-hole diagram, but given zero momentum transfer we cannot convert a hole at $-\Lambda$ to a particle at $+\Lambda$. In the ZS' diagram, we have a large momentum transfer, called Q in the inset at the bottom. This is of order K_F and much bigger than the radial cut-off, a phenomenon unheard of in say ϕ^4 theory, where all momenta and transfers are bounded by Λ. This in turn means that the loop momentum is not only restricted in the direction to a sliver $d\Lambda$, but also in the angular direction in order to be able to absorb this huge momentum Q and land up in the other shell being eliminated (see bottom of Fig. (1.3). So we have $du \simeq dt^2$, where $dt = d\Lambda/\Lambda$. The same goes for the BCS diagram. Thus F does not flow at one loop.

Let us now turn to the renormalization of V. The first two diagrams are useless for the same reasons as before, but the last one is special. Since the total incoming momentum is zero, the loop momenta are equal and opposite and no matter what direction K has, $-K$ is guaranteed to lie in the same shell being eliminated. However the loop frequencies are now equal and opposite so that the poles in the two propagators now lie in opposite half-planes. We now get a flow (dropping constants)

$$\frac{dv(\theta_1-\theta_3)}{dt} = -\int d\theta v(\theta_1-\theta)\ v(\theta-\theta_3) \tag{1.19}$$

Here is an example of a flow equation for a coupling function. However by expanding in terms of angular momentum eigenfunctions we get an infinite number of flow equations

$$\frac{dv_m}{dt} = -v_m^2. \tag{1.20}$$

one for each coefficient. These equations tell us that if the potential in angular momentum channel m is repulsive, it will get renormalized down to zero (a result derived many years ago by Anderson and Morel) while if it is attractive, it will run off, causing the BCS instability. This is the reason the V's are not a part of Landau theory, which assumes we have no phase transitions. This is also a nice illustration of what was stated earlier: one could begin with a large positive coupling, say v_3 and a tiny negative coupling v_5. After much renormalization, v_3 would shrink to a tiny value and v_5 would dominate.

1.3 Large-N approach to Fermi liquids

Not only did Landau say we could describe Fermi liquids with an F function, he also managed to compute the response functions at small ω and q in terms of the F function even when it was large, say 10, in dimensionless units. Again the RG gives us one way to understand this. To this end we need to recall the the key ideas of "large-N" theories.

These theories involve interactions between N species of objects. The largeness of N renders fluctuations (thermal or quantum) small, and enables one to make approximations which are not perturbative in the coupling constant, but are controlled by the additional small parameter $1/N$.

As a specific example let us consider the Gross-Neveu model[5] which is one of the simplest fermionic large-N theories. This theory has N identical massless relativistic fermions interacting through a short-range interaction. The Lagrangian density is

$$\mathcal{L} = \sum_{i=1}^{N} \bar{\psi}_i\ \partial\!\!\!/ \psi_i - \frac{\lambda}{N}\Big(\sum_{i=1}^{N} \bar{\psi}_i\psi_i\Big)^2 \tag{1.21}$$

Note that the kinetic term conserves the internal index, as does the interaction term: any index that goes in comes out. You do not have to know much about the GN model to to follow this discussion, which is all about the internal indices.

Figure 1.4 shows the first few diagrams in the expression for the scattering amplitude of particle of isospin index i and j in the Gross-Neveu theory. The "bare" vertex comes with a factor λ/N. The one-loop diagrams all share a factor λ^2/N^2 from the two vertices. The first one-loop diagram has a free internal summation over the index k that runs over N values, with the contribution being identical for each value of k. Thus, this one-loop diagram acquires a compensating factor of N which makes its contribution of order λ^2/N, the *same order in* $1/N$ *as the bare vertex*. However, the other one-loop diagrams have no such free internal summation and their contribution is indeed of order $1/N^2$. Therefore, to leading order in $1/N$, one should keep only diagrams which have a free internal summation for every

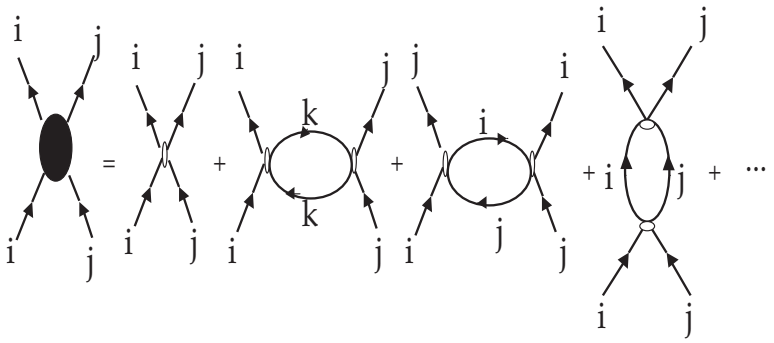


Figure 1.4: Some diagrams from large-N theory

vertex, that is, iterates of the leading one-loop diagram, which are called bubble graphs. *For later use remember that in the diagrams that survive (do not survive), the indices i and j of the incoming particles do not (do) enter the loops.* Let us assume that the momentum integral up to the cutoff Λ for one bubble gives a factor $-\Pi(\Lambda, q_{ext})$, where q_{ext} is the external momentum or frequency transfer at which the scattering amplitude is evaluated. To leading order in large-N the full expression for the scattering amplitude is

$$\Gamma(q_{ext}) = \frac{1}{N}\frac{\lambda}{1+\lambda\Pi(\Lambda, q_{ext})} \tag{1.22}$$

Once one has the full expression for the scattering amplitude (to leading order in $1/N$) one can ask for the RG flow of the "bare" vertex as the cutoff is reduced by demanding that the physical scattering amplitude Γ remain insensitive to the cutoff. One then finds (with $t = \ln(\Lambda_0/\Lambda)$)

$$\frac{d\Gamma(q_{ext})}{dt} = 0 \Rightarrow \frac{d\lambda}{dt} = -\lambda^2\frac{d\Pi(\Lambda, q_{ext})}{dt} \tag{1.23}$$

which is exactly the flow one would extract at one loop. Thus the one-loop RG flow is the exact answer to leading order in a large-N theory. All higher-order corrections must therefore be subleading in $1/N$.

1.3.1 Large-N applied to Fermi liquids

Consider now the $\bar{\psi}\psi - \bar{\psi}\psi$ correlation function (with vanishing values of external frequency and momentum transfer). Landau showed that it takes the form

$$\chi = \frac{\chi_0}{1+F_0}, \tag{1.24}$$

where F_0 is the angular average of $F(\theta)$ and χ_0 is the answer when $F = 0$. Note that the answer is not perturbative in F.

Landau got this result by working with the ground-state energy as a functional of Fermi surface deformations. The RG provides us with not just the ground-state energy, but an effective hamiltonian (operator) for all of low-energy physics. This operator problem can be solved using large N-techniques.

The value of N here is of order K_F/Λ, and here is how it enters the formalism. Imagine dividing the annulus in Fig. (1.2) into N patches of size (Λ) in the radial and angular directions. The momentum of each fermion $\mathbf{k}_i$ is a sum of a "large" part ($\mathcal{O}(k_F)$) centered on a patch labelled by a patch index $i = 1, ...N$ and a "small" momentum ($\mathcal{O}(\Lambda)$ within the patch[2].

Consider a four-fermion Green's function, as in Figure (1.4). The incoming momenta are labeled by the patch index (such as i) and the small momentum is not shown but implicit. We have seen that as $\Lambda \to 0$, the two outgoing momenta are equal to the two incoming momenta up to a permutation. At small but finite Λ this means the patch labels are same before and after. Thus the patch index plays the role of a conserved isospin index as in the Gross-Neveu model.

The electron-electron interaction terms, written in this notation, (with $\mathbf{k}$ integrals replaced by a sum over patch index and integration over small momenta) also come with a pre-factor of $1/N$ ($\simeq \Lambda/K_F$).

It can then be verified that in all Feynman diagrams of this cut-off theory the patch index plays the role of the conserved isospin index exactly as in a theory with N fermionic species. For example in Figure (1.4) in the first diagram, the external indices i and j do not enter the diagram (small momentum transfer only) and so the loop momentum is nearly same in both lines and integrated fully over the annulus, i.e., the patch index k runs over all N values. In the second diagram, the external label i enters the loop and there is a large momentum transfer ($\mathcal{O}(K_F)$). In order for both momenta in the loop to be within the annulus, and to differ by this large q, the angle of the loop momentum is limited to a range $\mathcal{O}(\Lambda/K_F)$. (This just means that if one momentum is from patch i the other has to be from patch j.) Similarly, in the last loop diagram, the angle of the loop momenta is restricted to one patch. In other words, the requirement that all loop momenta in this cut-off theory lie in the annulus singles out only diagrams that survive in the large N limit.

The sum of bubble diagrams, singled out by the usual large-N considerations, reproduces Landau's Fermi liquid theory. For example in the case of χ, one obtains a geometric series that sums to give $\chi = \frac{\chi_0}{1+F_0}$.

Since in the large N limit, the one-loop β-function for the fermion-fermion coupling is exact, it follows that the marginal nature of the Landau parameters F and the flow of V , Eqn. (1.20), are both exact as $\Lambda \to 0$.

A long paper of mine Ref. ([2]) explains all this, as well as how it is to be generalized to anisotropic Fermi surfaces and Fermi surfaces with additional special features and consequently additional instabilities. Polchinski [6] independently analyzed the isotropic Fermi liquid (though not in the same detail, since it was a just paradigm or toy model for an effective field theory for him).

This work was supported by NSF-DMR 0103639.

Bibliography

[1] M.E.Fisher *Critical Phenomena*, F. W. J. Hahne, Editor, Lecture Notes Number 186, Springer-Verlag, Berlin, (1983). These notes are from the Engelbrecht Summer School of 1983!

[2] R. Shankar, *Physica* **A177**, 530 (1991); R.Shankar, Rev. Mod. Phys. **66**, 129 (1994).

[3] L. D. Landau, Sov. Phys. JETP **3**, 920 (1956); Sov. Phys. JETP **5**, 101 (1957).

[4] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Dover Publications, New York, 1963.

[5] D. J. Gross and A. Neveu, Phys. Rev.**D10**, 3235 (1974).

[6] P.Polchinski In: *TASI Elementary Particle Physics*, ed by J. Polchinski and J.Harvey (World Scientific, 1992.)

P.C.E. STAMP

EUROPHYSICS LETTERS 15 March 1991

Europhys. Lett., 14 (6), pp. 569-574 (1991)

Quasi-Particle Quantum Numbers in Two and Three Dimensions.

P. C. E. STAMP

Physics Department, University of British Columbia
6224 Agricultural Rd., Vancouver, B.C., Canada V6T 2A6

(received 18 July 1990; accepted in final form 11 January 1991)

PACS. 67.40D – Quantum statistical theory; ground state, elementary excitations.
PACS. 67.50D – Normal phase.
PACS. 73.20D – Electron states in low-dimensional structures (inc. quantum wells, superlattices, layer structures and intercalation compounds).
PACS. 74.70T – Heavy-fermion superconductors.

Abstract. – It is shown how quasi-particle quantum numbers may be defined and calculated for interacting Fermi systems in 2 and 3 dimensions. Exact results are given for charged and neutral Fermi systems, both normal and superfluid, in 3 dimensions, and for Cherns-Simons implementations of anion theories in 2 dimensions. The latter is applied to the fractional Hall effect. In all cases, the local quasi-particle quantum numbers vary continuously with interactions and/or temperature.

There has been great interest recently in the possibility of exotic quasi-particle states in 2-dimensional Fermi systems. Such states are already known to exist in 1-dimensional systems [1, 2], and in the fractional quantum Hall effect (FQHE) [3], and many recent theories of high-T_c superconductivity depend crucially on the existence of quasi-particles with precisely defined fractional quantum numbers [4].

However the definition and calculation of these quantum numbers (or «charges») turn out to be full of surprises, even for 3-dimensional systems. Here a way of calculating both «local» and «global» charges will be given, along with exact results for a variety of systems. Apart from suggesting a number of interesting experiments, these calculations also considerably clarify the issues at stake in the discussion of exotic quasi-particles.

Definition of quasi-particle charges. – Consider some 2- or 3-dimensional system composed of interacting fermions, and eigenstates labelled by quantum numbers $\{\xi_j\}$. The expectation value of some local operator $\hat{X}(r, t)$ acting on the system in a state $|\alpha\rangle$, with one single quasi-particle, is $\langle \hat{X}_\alpha(r, t)\rangle = \langle \alpha|\hat{X}(r, t)|\alpha\rangle$. The Fourier transform $\langle \hat{X}_\alpha(Q)\rangle$ of this $\Lambda_\alpha^X(Q)$, the fully renormalized 3-point vertex describing interactions between the normalized «quasi-particles» and the field $X(Q)$ (here $Q = (q, \omega)$). In general we shall deal with quasi-particle wave-packets $|X\rangle$, which can nevertheless be labelled using the conserved quantities $\{\xi_j\}$ of the system.

We now define the functions $\bar{X}_\alpha(t)$ for different «charges» as

$$\bar{X}_\alpha(t) = \int \mathrm{d}r^D \, \theta(R - |r|) \, \langle X_\alpha(r, t) \rangle \,, \tag{1}$$

where the system size $L \gg R$, and we require $R \gg t\,\Delta p/m$, the free-particle wave-packet spread after time t (with momentum spread Δp); we also require $\Delta r(t=0) \ll R$. The «local quasi-particle charges» are given by $X_\alpha^{\mathrm{loc}} = \bar{X}_\alpha(t \to \infty)$ (but still keeping $R \gg t\,\Delta p/m$, in this long-time limit), while the «global quasi-particle charges» $X_\alpha^{\mathrm{glob}} \equiv \bar{X}_\alpha(t \to 0)$. Thus we see that the global charge X_α^{glob} refers to the expectation value of X_α averaged over the entire system (or over a small part of it at short times). However the local charge X_α^{loc} refers to that part of this charge that «stays together», in a somewhat distorted and slowly spreading «packet», as time goes on. Note that the shape and size of this packet (which is really a density matrix) is different for each different quantum number (see below). The difference between X_α^{loc} and X_α^{glob} arises solely from interactions.

3-dimensional systems. – It is very useful to start by considering some familiar examples. A *neutral* 3-dimensional Fermi liquid has 1-quasi-particle states $|p\sigma\rangle$, for which

$$X_{p\sigma}(Q) = \lambda^X_{\sigma\sigma'} \sum_{p'\sigma'} \left[\delta_{pp'} + \left(\frac{\boldsymbol{q} \cdot \boldsymbol{v}_{p'\sigma'}}{\boldsymbol{q} \cdot \boldsymbol{v}_{p'\sigma'} - \omega} \right) T^{\sigma'\sigma}_{p'p}(Q) \right] \tag{2}$$

(we consider wave-packets below). Here $\lambda^X_{\sigma\sigma'}$ is the bare 3-point vertex for quasi-particle interactions with the field X and $T^{\sigma\sigma'}_{pp'}(Q)$ is the *renormalized on-shell* quasi-particle T-matrix [5]. We assume that our initial quasi-particle energy $\varepsilon_{p\sigma}$ is considerably less than the typical fluctuation energies of the system (note $\varepsilon_{p\sigma}$ is a complex function of $\zeta_{p\sigma} = (p - p_{\mathrm{F}}^\sigma)\, v_{\mathrm{F}}^\sigma$; and $\varepsilon_{p\sigma} = \zeta_{p\sigma}$ for very low $\varepsilon_{p\sigma}$ [5]).

We may then solve (2) using microscopic Fermi-liquid theory, in terms of the Landau parameters F_l^S, F_l^A. The techniques are standard [5, 6], but the results are actually rather surprising. Considering for example the fermion number density $n_p(Q)$, and taking only $l = 0$, 1 parameters (as for ^{3}He liquid), one finds that

$$\langle n_p(Q) \rangle = \frac{1 + F_1^S[1/3 + \chi_0(\eta)(\eta^2 - \eta(\cos\theta_p - \chi_0(\eta)\cos^2 2\theta_p))]}{[1 + F_0^S\chi_0(\eta)][1 + F_1^S(1/3 + \eta^2\chi_0(\eta))] - \eta^2 F_0^S F_1^S \chi_0^2(\eta)} + O(\zeta_p^2 \ln \zeta) \,, \tag{3a}$$

$$\chi_0(\eta) = 1 - \frac{\eta}{2} \ln \left| \frac{1+\eta}{1-\eta} \right| + \imath \frac{\pi}{2} \eta \theta(1 - |\eta|) \,, \tag{3b}$$

where $\eta = \omega/qv_{\mathrm{F}}$, and $\theta_p = \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{q}}$. This very complex result contains all the details of the «decay down» of $|p\sigma\rangle$, via particle/hole and collective mode emission([1]). However although the Fourier transform is also very unwieldy (it is in fact the generalization of the 1st-order calculation of ref. [7] to all orders in perturbation theory), the long- and short-time results are very simple. Thus one finds $n_p^{\mathrm{glob}} = 1$, whilst $n_p^{\mathrm{loc}} = 1/(1 + F_0^S)$; and analogous calculations for spin and current give $\boldsymbol{S}_{p\sigma}^{\mathrm{glob}} = (1/2)\,\gamma\hbar\boldsymbol{\sigma}$, $\boldsymbol{J}_p^{\mathrm{glob}} = \boldsymbol{p}/m$, but $\boldsymbol{S}_{p\sigma}^{\mathrm{loc}} = (1/2)\,\gamma\hbar\boldsymbol{\sigma}/(1 + F_0^A)$, and $\boldsymbol{J}_p^{\mathrm{loc}} = \boldsymbol{p}/m(1 + 1/3F_1) = \boldsymbol{p}/m^*$. The difference between the global and local results describes «charge» that has escaped to (or been sucked in from) infinity. These fractions differ for each charge/quantum number, so that we have, *e.g.*, «partial spin/charge separation» at long times. Lest the reader doubt the applicability of our definitions here, it should be noted that

([1]) The details of the calculations of $\hat{X}_\alpha(Q)$ and its Fourier transform $\hat{X}_\alpha(r, t)$ are technically interesting but very lengthy, and will be given in a longer paper.

the functions $\langle n_p(Q)\rangle$, $\langle \boldsymbol{S}_{p\sigma}(Q)\rangle$, etc., are nothing but the Landau distribution functions for density, spin, etc., since $\Lambda_{p\sigma}(Q)$ solves the Landau-Boltzmann equation [6]. Thus our definitions of local and global quasi-particle charges correspond simply to the local and global parts of the relevant Landau distribution functions—which are themselves simply expectation values $\mathrm{Tr}\{\rho\hat{X}_\alpha\}$ over the reduced density matrix ρ [5].

These results are easily generalized to globally neutral electronic systems, but with one subtlety. At very short times, the standard calculation of the 3-point vertex gives $\Lambda_{p\sigma}(Q) = \varepsilon^{-1}(Q)\tilde{\Lambda}_{p\sigma}(Q)$, where $\tilde{\Lambda}_{p\sigma}(Q)$ is the «proper 3-point vertex» not containing direct Coulomb lines [6]. Now $\Lambda_{p\sigma}(Q)$ describes a very localized electronic wave-packet, whose electric charge is *not* locally compensated. But the correct description of the quasi-particles at long times is given by $\tilde{\Lambda}_{p\sigma}(Q)$, which satisfies the Landau-Silin equation; as is well known, this function describes, at long times, fermionic charge spread *uniformly*, thereby preserving *local* charge neutrality (cf. ref. [6]). Of course if we added electrons to the system, uncompensated by neutralizing charge, they would go to the walls [8]; but it is quite wrong to associate such excitations with quasi-particles, as usually defined.

Partial spin/charge separation also occurs—a fraction $F_0^A(1+F_0^A)^{-1}$ of the spin «escapes to infinity». F_0^A can be extracted from spin-wave measurements.

It is often assumed that the sharpness of quasi-particle charges may be restored if there are no gapless excitations. While this is often true in 1 dimension, it is incorrect in 3 dimensions. Consider, *e.g.*, a general singlet neutral superfluid. For short times one finds the usual results $n_p^{\mathrm{glob}} = (|u_p|^2 - |v_p|^2)$, $\boldsymbol{J}_p^{\mathrm{glob}} = \boldsymbol{p}/m$, and $\boldsymbol{S}_{p\sigma}^{\mathrm{glob}} = (1/2)\,\gamma\hbar\boldsymbol{\sigma}$, and straightforward generalization of the method given above yields the long-time limits

$$\begin{cases} n_p^{\mathrm{loc}} = (|u_p|^2 - |v_p|^2)\,Y(T)(1+F_0^S\,Y(T))^{-1}, \\ \boldsymbol{S}_{p\sigma}^{\mathrm{loc}} = \dfrac{1}{2}\gamma\hbar\boldsymbol{\sigma}Y(T)(1+F_0^A\,Y(T))^{-1}, \\ J_{p_i}^{\mathrm{loc}} = \left[\dfrac{Y(T)}{1+1/3F_1^S\,Y(T)}\right]_{ij}\dfrac{P_j}{m}, \end{cases} \tag{4}$$

where $\hat{Y}_{ij}(T)$ and $Y(T)$ are the matrix and scalar Yosida functions [9] for the appropriate gap function (*s*-wave, *d*-wave, etc.). Again partial (and only partial) spin/charge separation occurs. Moreover this partial separation is *not* changed, if we add Coulomb interactions to the system—exactly as for the metal described above, quasi-particles are neutral in the long-time limit, and $\boldsymbol{S}_{p\sigma}^{\mathrm{loc}}$ is still given by (4). Thus it is incorrect to regard the quasi-particles in 3-d superconductors as spinons [8].

In view of these results, one is led to ask how to properly define quasi-particle *statistics*. In 3-d systems this is normally done quite unambiguously via their *global* commutation relations [6]. This is equivalent to the global fermionic charge defined above, which is equal to unity for fermionic systems. The *local* fermionic charge n_σ^{loc} is not the same. In fact it corresponds to the Berry phase ϕ_p that one would obtain by slowly moving one quasi-particle around a second one, on a circle of radius R centered on the second (and with $R \gg t\,\Delta p/m$). The demonstration that $\phi_p = 2\pi n_p^{\mathrm{loc}}$ is then essentially the same as that in the anion literature (see, *e.g.*, Arovas *et al.* [10]), since the excess phase accumulated corresponds to the excess enclosed fermionic charge. However in 3 dimensions this Berry phase definition is somewhat artificial (since we can always deform the circle into a quite different curve, with a different ϕ_p), so it is best to stick to the definition of n_p^{loc} given previously.

2 *dimensions.* – Elementary consideration of the 2-particle scattering matrix for point particles shows that in 3 dimensions it is forced by undistinguishability and unitarity to take the form $K(\Omega) = f(\boldsymbol{\Omega}) \pm f(\boldsymbol{\Omega} + \pi)$. However in 2-d the more general

$$K(\theta) = \sum_{n=-\infty}^{\infty} \exp[2i\alpha n] f(\theta + 2\pi n)$$

is allowed [10], yielding «anions», with statistics and fermionic charge α. This result follows for point particles when $\alpha \neq 0$ because the diverging centrifugal force (as $|\boldsymbol{r}_1 - \boldsymbol{r}_2| \to 0$) prevents world lines from crossing. But how do we deal with *quasi-particles*, which are always smeared out in space?

A common answer to this is to argue that, if the quasi-particles are widely separated, then the above argument (or its more rigorous braid group formulation [10]) is still applicable, since world lines will then rarely cross. The argument would then justify *a posteriori* the use of Berry's phase to define quasi-particle statistics in, *e.g.*, the fractional Hall effect (FQHE); it gives anions with fractional fermionic charge $n^{\mathrm{loc}} = \nu = \pm 1/(2l+1)$, where $l = 1, 2, \ldots$, and ν is the Landau level filling fraction. At temperature $T = 0$ this result follows from Laughlin's wave function [3], and is easily shown using the methods above, since the charge does not spread at $T = 0$. Hence we find [3, 10] that $\phi_p = 2\pi n_p^{\mathrm{loc}} = 2\pi\nu$.

However at finite T things are more subtle. It has not yet been possible to generalize the Laughlin theory to finite T, but we can resort to the effective action theories that have been recently devised [11]. The simplest versions of these have a Lagrangian density

$$L(r,t) = \Psi^+\left[(i\partial_t - e(A_0 + a_0)) - \frac{K}{2}(\nabla - i(e\boldsymbol{A} + \boldsymbol{a}))^2\right]\Psi + \beta|\Psi|^2 - \lambda|\Psi|^4 - \frac{e^2\nu}{4\pi}\varepsilon^{\lambda\nu\sigma} a_\lambda \partial_\nu a_\sigma, \qquad (5)$$

where the fields $\Psi(r, t)$ can be interpreted, following Read [11], as the amplitude for finding a particle at $(\boldsymbol{r}, t)$. At $T = 0$ the vortices in the «statistical gauge field» $a_\lambda(\boldsymbol{r}, t)$, of form $a_\lambda(\boldsymbol{r}, t) \sim (\hat{\boldsymbol{z}} \times \hat{\boldsymbol{r}})/er$, collect a local charge $n^{\mathrm{loc}} = \nu$ around themselves (note that $A_\nu(\boldsymbol{r}, t)$ is the e.m. field).

Now it might be assumed that, because there is an energy gap $\Delta = \beta/\lambda$ in this theory, the charge is bound to the vortex cores in «sub-gap» states, as in superconducting vortices. But this is quite wrong. The eigenfunctions for (Ψ, Ψ^+) in the presence of a single vortex are easily found, and have the form (for $r^2 \gg l_{\mathrm{H}}^2 = \hbar/eB$):

$$\begin{pmatrix} \Psi_{km}(\boldsymbol{r}) \\ \Psi^+_{km}(\boldsymbol{r}) \end{pmatrix} = \left(\frac{E_p - \Delta}{2E_p}\right)^{1/2} \begin{pmatrix} J_{|m-\nu|}(kr)\exp[im\phi] \\ J_{|m+1-\nu|}(kr)\exp[i(m+1)\phi] \end{pmatrix}, \qquad (6)$$

where $\boldsymbol{p} = \hbar\boldsymbol{k}$, the quasi-particle energy $E_p \geq \Delta$, and $m = 0, \pm 1, \pm 2, \ldots$ (we assume $\nu < 1$). Then there are no bound states, for any T (if $\nu < 1$), and n^{loc} arises *entirely* from continuum states. The situation is the same as that prevailing in $(2+1)$-dimensional QED [12], and indeed we could not have a fractional n^{loc} if the states were bound!

It is then revealing to calculate n^{loc} around a vortex at finite T. A simple Boltzmann average then promotes charge higher up these states, and assuming $kT \ll \Delta$ (the Lagrangian (5) is unlikely to be meaningful otherwise), we find

$$n^{\mathrm{loc}} \simeq \nu(1 - \exp[-\Delta/2kT]) \qquad (7)$$

so that some charge has escaped (note that this result could also be obtained [1] by applying

trace identities to (5)). In a real FQHE system there will be corrections to this arising from other quasi-particles or quasi-holes—these have long-range interactions. Nevertheless (7) clearly shows that the $T = 0$ Berry phase definition of n^{loc} will eventually fail at finite T (although if we had a finite-T microscope generalization of Laughlin's theory, presumably we could recover (7) as a Berry phase at finite T).

Experimental tests. – Let us briefly examine what is possible here. Recent experiments [13] have indicated how one may measure n^{loc} in the FQHE, and similar experiments should be capable of checking (7), thereby testing the effective action theories. A good way of testing the 3-d results in normal and superfluid ^{3}He would be via ballistic quasi-particles experiments involving thin wires [14], since these experiments see n^{loc} (not n_p^{glob}) for a quasi-particle «wave-packet». Similar experiments involving spin could be done by spin wave transmission (in metals or normal ^{3}He). In superconductors a convenient method would be to make a ballistic point contact spectroscopic measurement (using a polarized tip, if one is interested in $\boldsymbol{S}_p^{\text{loc}}$). Detailed discussion of such experiments will be given elsewhere.

Thus, to conclude, we see that the «separation» of quasi-particle charges (*i.e.* the sometimes quite large differences between the local values of, *e.g.*, spin and fermionic charge) is a quite general phenomenon in both 2 and 3 dimensions—as is the distinction between the local and global values of each charge. This phenomenon arises because of interactions in 3 dimensions, at any temperature; and in 2 dimensions, even if there are topological terms in the effective action which may enforce quantized local charges at $T = 0$, these constraints break down at finite T. These local charges are often accessible experimentally, in both 2 and 3 dimensions.

* * *

I should like to thank Profs. G. SEMENOFF, W. KOHN and S. KIVELSON for discussions of these results. This work was supported by an NSERC-URF grant.

REFERENCES

[1] NIEMI A. J. and SEMENOFF G., *Phys. Rep.*, **135** (1986) 99.
[2] AFFLECK I., *J. Phys. Condensed Matter*, **1** (1989) 304.
[3] LAUGHLIN R. B., *The Quantum Hall Effect*, edited by R. E. PRANGE and S. M. GIRVIN (Springer, Berlin) 1987, p. 223.
[4] See, *e.g.*, CHEN Y. H. *et al.*, *Int. J. Mod. Phys. B*, **3** (1989) 1001 and references therein.
[5] BAYM G. and PETHICK C. J., *The Physics of Liquid and Solid Helium*, Vol. II (Wiley, New York, N.Y.) 1977, Chapt. 1.
[6] NOZIÈRES P., *Theory of Interacting Fermi Systems* (Benjamin, New York, N.Y.) 1964, Chapt. 6.
[7] HEINONEN O. and KOHN W., *Phys. Rev. B*, **36** (1987) 3565.
[8] KIVELSON S. A. and ROKHSAR D. S., ITP preprint (June 1989).
[9] The Yosida functions are defined in weak coupling in LEGGETT A. J., *Rev. Mod. Phys.*, **47** (1975) 331 and with strong-coupling corrections by SERENE J. W. and RAINER D., *Phys. Rep.*, **101** (1983) 221.
[10] LEINAAS J. M. and MYRHEIM J., *Nuovo Cimento B*, **37** (1977) 1. See also WU Y. S., *Phys. Rev. Lett.*, **52** (1984) 2103; **53** (1986) 111. Note that we are following standard terminology in talking about «fractional fermionic charge» (cf. ref. [1]) although it might be more consistent to talk about «anionic charge». The relation to Berry's phase in the FQHE is given by AROVAS D., SCHREIFFER J. R. and WILCZEK F., *Phys. Rev. Lett.*, **53** (1984) 722.

[11] GIRVIN S. M. and MACDONALD A. H., *Phys. Rev. Lett.*, **58** (1987) 1252; ZHANG S. C., HANSSON T. H. and KIVELSON S., *Phys. Rev. Lett.*, **62** (1989) 82; READ N., *Phys. Rev. Lett.*, **62** (1989) 86; these ideas have been generalized to the more complicated FQHE hierarchies by READ N., *Phys. Rev. Lett.*, **65** (1990) 1502.

[12] A similar phenomenon has been discussed in (2 + 1)-dimensional QED by DE VEGA H. J., *Phys. Rev. D*, **18** (1978) 2932 (at zero temperature only). Note that if $\nu > 1$, then there will be $[\nu]$ bound states at $T = 0$, plus $\nu - [\nu]$ local charge in continuum states.

[13] CLARK R. G. *et al.*, *Phys. Rev. Lett.*, **60** (1988) 1747; **62** (1989) 1536; SIMMONS J. A. *et al.*, *Phys. Rev. Lett.*, **63** (1989) 1731.

[14] See, *e.g.*, PICKETT G. R., *Rep. Prog. Phys.*, **51** (1988) 1295; LOUNASMAA O. V. and PICKETT G. R., *Sci. American*, **262**, No. 6 (June 1990), pp. 64-71.

2.1. Quantum environments

2.1.1. Extended environmental modes; oscillator baths

Research on the dynamics of polarons and related problems led Feynman in the early 1960s to a general discussion of the interaction of a quantum system with its background environment. Feynman and Vernon (1963) considered the case where each environmental mode coupled only weakly to the central system. Arguing that for this weak-coupling case, the effect of *any* environment could be mapped to that of a set of oscillators, they treated a model Hamiltonian in which a central system $\mathscr{S}$, with generalised coordinates P, Q and Hamiltonian $H_0(P, Q)$, interacted with an environment $\mathscr{E}$ of oscillators with generalised coordinates $\{p_q, x_q\}$ and Hamiltonian $H_{\rm env}^{\rm osc}(\{p_q, x_q\})$, via a simple bilinear coupling:

$$H_{\rm eff}^{\rm osc} = H_0 + H_{\rm int} + H_{\rm env}^{\rm osc}, \quad H_{\rm int}(Q, \{x_q\}) = \sum_{q=1}^{N} c_q x_q Q. \tag{1}$$

We assume that the entire Hamiltonian $H_{\rm eff}(\Omega_0)$ is defined with an ultraviolet cutoff energy Ω_0. The important points to bear in mind here are:

(i) the oscillators have bosonic statistics, and typically represent delocalised modes, extending over the whole region of the environment. Typical examples are phonons, magnons, electron–hole pairs, or photons, which are wave-like oscillations of some background field. These are the low-energy modes of the environment—at higher energies the model usually breaks down;
(ii) the couplings $\{c_q\}$ are weak—in fact $c_q \sim O(N^{-1/2})$, where N is the number of low-energy environmental modes (N is thus proportional to the size of the environmental domain). This typically follows because we must normalise the oscillator wave functions (so they are $\sim O(N^{-1/2})$). Typically N is very big, so that mathematical treatments often just adopt the 'thermodynamic limit' $N \to \infty$. Since the effect of each oscillator to second order is $\sim |c_q|^2 \sim O(1/N)$, their total effect is then independent of N, as it should be in this limit. Thus each oscillator is only very weakly affected by the system, but the system may be quite strongly affected by the oscillators.

Curiously, the work of Feynman and Vernon had no impact whatsoever on the discussion of quantum measurements or decoherence for two decades—possibly because it was phrased in the then unfamiliar language of path integrals, and because the community working on the foundations of quantum mechanics was less interested at that time in detailed models.

At the beginning of the 1980s Caldeira and Leggett (1983) introduced a somewhat generalised Feynman–Vernon model, in which the coupling $\sum_q c_q x_q Q$ was replaced by

$$H_{\rm int}^{\rm osc} = \sum_{q=1}^{N} [F_q(Q) x_q + G_q(P) p_q]. \tag{2}$$

The Hamiltonians (1) and (2) are effective ones, which means amongst other things that the couplings c_q, F_q, and G_q, the oscillator frequencies ω_q, and even the system Hamiltonian H_0 depend not only on the UV cutoff Ω_0 but also on the bath temperature T. This may seem strange to some (particularly readers more at home with the models used in particle physics). Recall however that *all* Hamiltonians in physics are effective ones,

written in a quantum system in terms of operators defined over some restricted Hilbert space, depending implicitly or explicitly on energy cutoffs, temperature, and possibly other boundary conditions.[8] It is only when dealing with a very rarified medium that one can ignore these complexities.

Caldeira and Leggett gave arguments for the very general applicability of such effective Hamiltonians to systems at low energy (along with specific application to superconducting SQUIDs). Consider some arbitrary environment, with eigenstates $\phi_\alpha(\mathbf{X})$ and eigenenergies ε_α defined over the environment's full multi-dimensional coordinate space $\mathbf{X}$. Assume the system interacts with this environment via some interaction $V(Q,\mathbf{X})$. Then the arguments go as follows:

(a) Certainly we can recover an oscillator bath model if the coupling between different eigenstates induced by the interaction $V(Q,\mathbf{X})$ is weak, i.e., under the Feynman–Vernon condition that

$$|V_{\alpha\beta}| \ll |(\varepsilon_\alpha - \varepsilon_\beta)| \tag{3}$$

for all relevant environmental states, where $V_{\alpha\beta} = \int \mathrm{d}\mathbf{X}\phi^*_\alpha(\mathbf{X})V(Q,\mathbf{X})\phi_\beta(\mathbf{X})$. The oscillator modes then correspond to the transitions between these states, and $\omega_q \equiv (\varepsilon_\alpha - \varepsilon_\beta)$.

(b) However, even if the weak-coupling condition is not obeyed, we can use a Born–Oppenheimer argument to derive a similar criterion. We first define adiabatic environmental eigenstates $\tilde{\phi}_\alpha(\mathbf{X},Q)$ and eigenenergies $\tilde{\varepsilon}_\alpha(Q)$, which depend on the instantaneous system coordinate Q. Now suppose that these states have a *fast* dynamics compared to the slower dynamics of the system coordinate Q (formally, that if Q moves on a frequency scale E_0, then $E_0 \ll \tilde{\varepsilon}_\alpha$). One then defines a fake 'gauge potential' $A_{\alpha\beta}$, describing the effect of the slowly changing Q on the bath modes, given by $\mathrm{i}A_{\alpha\beta} = \int \mathrm{d}\mathbf{X}\tilde{\phi}^*_\alpha(\mathbf{X})\partial/\partial Q\tilde{\phi}_\beta(\mathbf{X})$; there is no reference to the original interaction between Q and the bath modes, because this has already been incorporated into the renormalised $\tilde{\varepsilon}_\alpha$. Standard manoeuvres then show that we can make a mapping to an oscillator bath provided

$$|A_{\alpha\beta}| \ll |(\tilde{\varepsilon}_\alpha - \tilde{\varepsilon}_\beta)| \tag{4}$$

for all the relevant modes. If (4) is satisfied, then the oscillators now describe transitions between the new adiabatic bath modes, with frequencies $\omega_q \equiv (\tilde{\varepsilon}_\alpha - \tilde{\varepsilon}_\beta)$; and one can also derive the couplings F_q, G_q in terms of the gauge coupling in (4).

(c) Leggett et al. then argued that the low-T, low-energy quantum dynamics of such a system could be related to its higher T dissipative classical dynamics (cf. Fig. 1). From the classical dissipative dynamics one *infers* a low-energy effective Hamiltonian (having the form (1), with the generalised interaction in (2)); in particular, one finds the form of the couplings in (2). This is crucially important—instead of trying to derive the form of H_{eff} from some theory (a move which is always open to criticism given the

[8] The idea of the 'effective Hamiltonian' (or the effective Lagrangian) is rather subtle, and bound up in the recent history of physics with the idea of the renormalisation group (although discussions go back at least to the 19th century). See, e.g., Anderson (1984). For a recent discussion of the effective Hamiltonian, directed to a philosophical readership, see Stamp (to be published).

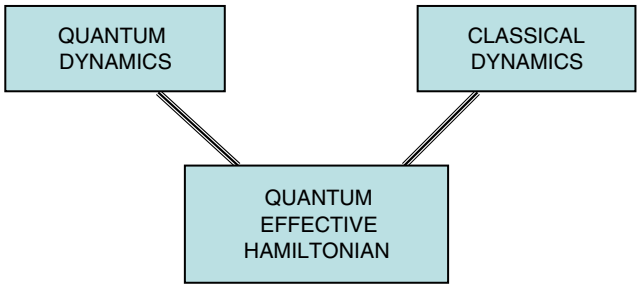


Fig. 1. The epistemological connection between the observable classical (usually high-temperature) dynamics and the low-temperature quantum dynamics (often not so easy to observe), for a system with many degrees of freedom. Both can be derived from the correct quantum effective Hamiltonian. Often (as in the approach of Caldeira and Leggett) one infers the quantum Hamiltonian from experiments on the classical behaviour.

huge complexity of large systems), one instead infers it directly from experiment.[9] One then *derives* the quantum dynamics of the system from this effective Hamiltonian.

At first glance the assumptions behind the oscillator bath model seem restrictive—small oscillations and weak coupling to each mode, use of a Born–Oppenheimer approximation, etc. However appearances are deceptive—oscillator bath models are quite robust in the real world. A large class of effective Hamiltonians (sometimes called a *universality class*), which will describe many physical systems, can be mapped to models of the oscillator bath type (Dubé & Stamp, 2001). Examples include: (i) itinerant fermion baths (e.g., a bath of interacting conduction electrons), in three, two or one dimensions; (ii) systems having weak higher-order 'anharmonic' couplings to extended bath modes—these can be absorbed into modified couplings to a new set of oscillators (the couplings and oscillator frequencies now being very strongly T-dependent); and (iii) systems where bath modes are strongly coupled to the system, provided the condition (4) is not violated (i.e., provided the effective coupling between two environmental states goes to zero fast enough as one reduces the energy difference between them). It is worth remarking here on a point which is crucially important for decoherence. The reduction in the strength of coupling to oscillator bath modes at low energies is a general feature of extended environmental states, whose density of states always goes down with energy, because of decreasing available phase space volume. This means that at as one lowers energies and temperatures towards zero, we can naively expect the decoherence from oscillator baths to also decrease to zero.

We have seen that oscillator bath models of quantum environments are thus much more general than is often assumed in the literature. However they certainly cannot always work, and they clearly fail in many solid-state systems at low temperatures. In order to understand why, we make a little diversion into the real world of low-energy physics.

[9] In Caldeira–Leggett theory, the interaction between system and environment is summarised in a 'spectral function' $J(\omega, T)$, a function of frequency and temperature. If the Caldeira–Leggett effective Hamiltonian applies to some physical system, *and* if one knows $J(\omega, T)$, then the behaviour can be derived theoretically in both classical and quantum regimes. More typically, one *infers* $J(\omega, T)$ from the classical and/or quantum behaviour in experiments.

2.1.2. Interlude: real condensed matter

Condensed matter is all around us—we are directly aware of little else. All measuring systems are made from condensed matter. It is clearly messy, and complex structures and order are evident everywhere (not least in living things). As a result, except for the He liquids (which go superfluid at low T and which can be made in essentially completely pure form) and rarified gases and plasmas, the low-energy effective Hamiltonians of *real* condensed matter systems are extraordinarily subtle (and very far from the descriptions usually given in student textbooks). There is a common misunderstanding that these subtleties have to do with 'dirt' effects (the 'squalid state', in Pauli's famous phrase). In fact they are mostly intrinsic, for the following reasons (footnote 8):

(i) *Topology.* Many-particle wave-functions have topological properties which restrict and sometimes control the dynamics. This often leads to new branches of low-energy 'topological excitations', with their counterpart in the effective Hamiltonian (Thouless, 1998).
(ii) *Lattices + interactions.* In solids, electrons are constrained to move between different atomic orbitals. Strong repulsive interactions between electrons can prevent more than one particle per orbital, imposing a highly non-trivial structure on the Hilbert space of the effective Hamiltonian and even causing the low-energy states to localise.
(iii) *Boundaries or edges.* All systems have boundaries. In conjunction with long-range forces and/or the topological properties of wave-functions, the boundaries and the states localised near them can control the low-energy properties of the whole system.
(iv) *Frustration.* Interactions between two different pairs of particles or spins are often 'incompatible' (i.e., lead to contradictory effects on any one of the particles). The result is typically a large number of almost degenerate low-energy states which hardly communicate.[10] The system can never reach its putative ground state (which then becomes a mere mathematical chimera). Because of frustration, most pure solids, without impurities, are intrinsically disordered. States pile up at low energies—many of these low-energy states are localised (footnote 10).

Clearly none of these effects come from 'junk' or 'dirt'; moreover, because they arise from very general mechanisms, they lead to effects that are ubiquitous in low-temperature experiments. These include peculiar structure in the low-energy density of states, complex and often non-linear long-time relaxation phenomena, including 'glassy' behaviour (the freezing out of dynamics caused by frustration), increasingly subtle kinds of quantum ordering as one lowers the temperature, etc. Over the last four decades a phenomenological description has emerged for these low-energy phenomena, in terms of a set of low-energy discrete modes (i.e., each having a discrete finite set of states, often only two, in the energy range of interest), appropriate to localised states (Anderson, 1994; Binder & Young, 1986; Esquinazi, 1998; Mézard et al., 1987). These states interact both amongst themselves, and with the extended 'oscillator modes'. Thus one ends up with a low energy description in terms of a set of interacting 'two-level systems'; usually the interactions are fairly weak,

[10]The only elementary review of some of the low-energy complexities in real solids seems to be the five short articles by Anderson on 'spin glasses' (Anderson, 1994). More sophisticated reviews are by Binder & Young (1986) and Mézard, Parisi, & Virasoro (1987); this latter book also makes the connection with work in computation and biology.

although they can have important effects. There is certainly no universal agreement about this picture (Yu & Leggett, 1988), but in many cases there is extensive evidence that it gives a good description of the low-energy physics (Anderson, 1994; Binder & Young, 1986; Esquinazi, 1998; Mézard et al., 1987). I emphasise again that these effects are pretty much universal in solids, although their effects are sometimes not obvious until very low temperatures. Their effects on ordinary transport and other dissipative properties can be very small (making them almost invisible at higher temperatures), but we shall see that their contribution to decoherence can be very large.

One is often met by surprise at this situation. How, it is asked, can a simple solid show such 'pathological behaviour', when after all it is made up electrons, protons, etc., which can be described by a simple continuum theory having none of these complexities? The fallacy in this argument is the assumption that the effective Hamiltonian of a composite system will somehow be analysable into that of its constituents.[11] This is not true—the effective theory of the constituents is still an effective theory, applicable only in a certain energy range and assuming a restricted Hilbert space. For this reason neither the vacuum nor the low-energy eigenstates of the high-energy Hamiltonians used in particle physics look anything like a condensed matter system (even though this is physically what a high-energy system becomes if it is cooled!). In many real solids, an infinite hierarchy of effective Hamiltonians, ever more complex, is expected as one lowers the energy scale (footnote 8), and we only have a dim understanding of what their structure might be. In other words, we do not really understand the basic structure of the lowest energy states or Hilbert spaces of most many-body systems. An understanding of this low-energy structure is one of the holy grails of condensed matter physics—in many ways it seems more elusive now than it did 30–40 (or even 100) years ago. One hundred years ago, with the vindication of the atomic hypothesis, but before quantum mechanics, a simple reductionist view of condensed matter looked very reasonable. Thirty to forty years ago, a unification of methods between quantum field theory and condensed matter physics looked imminent—the Ginzburg–Landau–Wilson theory of phase transitions, and the BCS theory of superconductivity, were shaping much of modern particle theory. This unification has happened, but only in the study of 'simple' systems. For a more realistic perspective see Anderson (1994), Binder and Young (1986), Esquinazi (1998) and Mézard et al. (1987).

If some day we ever have a ''complete theory of everything'', with a 'universal Hamiltonian' whose eigenstates (including the ground state) represent the real states of the universe, over all energy scales, then we would presumably find that the low-energy states of this Hamiltonian contain the full complexity of real condensed matter. Right now we have little idea if such a theory would even be meaningful (it is perhaps more likely that the whole Hamiltonian structure will be replaced by something more fundamental). We certainly have not the slightest idea whatsoever what it would look like. Current efforts towards progress range from theory at supra-Planck scale energies, to the exploration of coherence phenomena at temperatures below 10^{-9} K.

[11] It is commonly argued that the 'complexity' of low-energy physics comes only from the large number of constituents (this is certainly the point of view of 'reductionists'). This argument is refuted in a well-known paper by Anderson (1972), which inspired a very large subsequent literature.

2.1.3. Localised modes: spin baths

We return now to the question at hand, which is to understand the sources of decoherence at low energies in real condensed matter systems. The importance of the previous discussion is that we now see we must deal with the large number of low-energy localised states existing in solids, or more generally, low-energy modes having a finite Hilbert space, with discrete excitations. The general nature of these was described above; they include the eigenstates of nuclear spins, of topological defects, and of various more subtle modes associated with frustration, boundaries, and intrinsic disorder. In any real system there will also be 'junk' effects, coming from paramagnetic impurities, 'charge trap' excitations, etc. In many systems we may not know exactly what these discrete modes are, but as noted above, their presence is often very obvious in experiments (Anderson, 1994; Binder & Young, 1986; Esquinazi, 1998; Mézard et al., 1987).

Now one can always map a system having a set of M discrete states to a spin system, with spin σ, such that $2\sigma + 1 = M$. Thus we can in all cases describe an environment of these states as a '*spin bath*' (Prokof'ev & Stamp, 2000). Spin baths have the following general characteristics:

(i) The generic model for a quantum system interacting with a spin bath (corresponding to the generic oscillator bath model defined by Eqs. (1) and (2)) has the effective Hamiltonian:

$$H^{\rm sp}_{\rm eff}(\Omega_0) = H_0 + H^{\rm sp}_{\rm int} + H^{\rm sp}_{\rm env}, \tag{5}$$

where $H_0(P, Q)$ describes the system as before; but now the interaction term is a vector coupling to a set of 'spins' $\{\boldsymbol{\sigma}_k\}$ (which for simplicity we take here to be two-level systems, i.e., spin-$\frac{1}{2}$ systems):

$$H^{\rm sp}_{\rm int} = \sum_k^{N_s} \boldsymbol{F}_k(P, Q) \cdot \boldsymbol{\sigma}_k, \tag{6}$$

and the spin bath Hamiltonian itself has the form:

$$H^{\rm sp}_{\rm env} = \sum_k^{N_s} \mathbf{h}_k \cdot \boldsymbol{\sigma}_k + \sum_{k,k'}^{N_s} V^{\alpha\beta}_{kk'} \sigma^{\alpha}_k \sigma^{\beta}_{k'}, \tag{7}$$

with a set of external fields $\{\mathbf{h}_k\}$, and interspin interactions $V_{kk'}$. The generalisation of this model to bath modes having $M > 2$ discrete states is straightforward.

(ii) Each bath 'spin' interacts only weakly with its compatriots—formally we require that $\{|\boldsymbol{F}_k|\} \gg |V_{kk'}|$. If the $\{\boldsymbol{\sigma}_k\}$ describe localised modes, this is quite typical. The different bath excitation wave-functions do not overlap and can only communicate via weak long-range interactions $V_{kk'}$, whereas there is nothing limiting the size of the $\{|\boldsymbol{F}_k|\}$ (which are no longer $\sim O(1/N^{1/2})$). The bath dynamics is then under the direct control of the central system (note that inequality (4) is now violated), with its own 'intrinsic dynamics' playing second fiddle. Recall that this is exactly opposite to the oscillator bath system, where the intrinsic dynamics of the oscillator bath is only weakly perturbed by the central system, because the oscillator frequencies $\{\omega_q\}$ are much larger than either the $\{c_q\}$ or the F_q, G_q in (2). This situation is illustrated in Fig. 2.

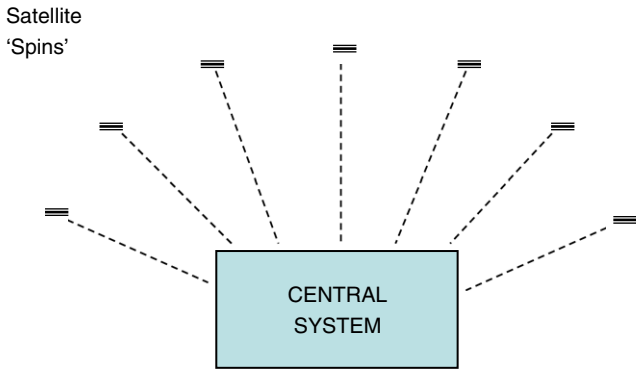


Fig. 2. The 'spin bath' environment—a set of satellite spins couples to the central quantum system of interest. The spins typically represent localised modes (not necessarily spins!) in the environment, each with a finite Hilbert space (often two-dimensional). The coupling between spins is weak compared to the coupling of each to the central system.

Clearly under some circumstances we can map the spin bath onto an oscillator bath. For example, if the interactions $V_{kk'}$ are strong (i.e., if $\{|\boldsymbol{F}_k|\} < |V_{kk'}|$ and if $|\mathbf{h}_k| < |V_{kk'}|$), then the bath spins can couple together to form extended 'spin waves', and $H_{\rm eff}^{\rm sp}(\Omega_0)$ then maps back to a Caldeira–Leggett model. If the central system dynamical energy scale $E_0 \gg |\boldsymbol{F}_k|$, then one goes to an anti-adiabatic (or 'anti-Born–Oppenheimer') limit, in which the system–bath couplings can be treated perturbatively. One can give more complete criteria for the mapping of spin baths to oscillator baths (Prokof'ev & Stamp, 2000), which we see must also involve the static fields $\{\mathbf{h}_k\}$.

In real physical systems the coupling energies $|\boldsymbol{F}_k|$ and static field strengths $\{|\mathbf{h}_k|\}$ are often spread over a very wide range, particularly in systems with frustration, disorder or impurities (note that 'impurities' include nuclear spins, which are almost everywhere; they live in some finite fraction of the nuclei of almost all the elements in solids). We cannot then use either or a Born–Oppenheimer or an anti-Born–Oppenheimer approximation, there are many environmental modes which must be treated directly as localised modes. Because these modes then have characteristic frequencies similar to those of the central system we are interested in, they cause a lot of decoherence.

2.2. Bath-induced decoherence and relaxation

Although the detailed calculation of the dynamics of decoherence is a complicated business, many of the main points can be understood by simple (although qualitative) arguments.

As noted earlier, in the early development of this subject, the idea of decoherence was very much bound up with quantum measurements. Decoherence was, in effect, viewed as a process in which the environment $\mathscr{E}$ 'measured' the state of the system $\mathscr{S}$ being decohered, via a transition

$$\sum_j c_j \psi_j \Phi_0 \;\rightarrow\; \sum_j c_j \chi_j \Phi_j, \tag{8}$$

J. UFFINK

on this issue and that it was a matter of taste.

4.2 The Boltzmann equation and H-theorem (1872)

In 1872 Boltzmann published one of his most important papers. It contained two celebrated results nowadays known as the Boltzmann equation and the H-theorem. The latter result was the basis of Boltzmann's renewed claim to have obtained a general theorem corresponding to the Second Law. This paper has been studied and commented upon by numerous authors, and an entire translation of the text has been provided by (Brush 1966). Thus, for the present purposes, a succinct summary of the main points might have been sufficient. However, there is still dispute among modern commentators about its actual content.

The issue at stake in this dispute is the question whether the results obtained in this paper are presented as necessary consequences of the mechanical equations of motion, or whether Boltzmann explicitly acknowledged that they would allow for exceptions. Klein has written:

> I can find no indication in his 1872 memoir that Boltzmann conceived of possible exceptions to the H-theorem, as he later called it (Klein 1973, p. 73).

Klein argues that Boltzmann only came to acknowledge the existence of such exceptions thanks to Loschmidt's critique in 1877. An opposite opinion is expressed by von Plato (1994). Calling Klein's view a "popular image", he argues that, already in 1872, Boltzmann was well aware that his H-theorem had exceptions, and thus "already had a full hand against his future critics". Indeed, von Plato states that

> Contrary to a widely held opinion, Boltzmann is not in 1872 claiming that the Second Law and the Maxwellian distribution are *necessary* consequences of kinetic theory (von Plato 1994, p. 81).

So it might be of some interest to try and settle this dispute.

Boltzmann (1872) starts with an appraisal of the role of probability theory in the context of gas theory. The number of particles in a gas is so enormous, and their movements are so swift that we can observe nothing but average values. The determination of averages is the province of probability calculus. Therefore, "the problems of the mechanical theory of heat are really problems in probability calculus" (Abh. I, p. 317). But, Boltzmann says, it would be a mistake to believe that the theory of heat would therefore contain uncertainties.

He emphasizes that one should not confuse incompletely proven assertions with rigorously derived theorems of probability theory. The latter are necessary consequences of their premisses, just like in any other theory. They will be confirmed by experience as soon as one has observed a sufficiently large number of cases. This last condition, however, should be no significant problem in

the theory of heat because of the enormous number of molecules in macroscopic bodies. Yet, in this context, one has to make doubly sure that we proceed with the utmost rigour.

Thus, the message expressed in the opening pages of this paper seems clear enough: the results Boltzmann is about to derive are advertised as doubly checked and utterly rigorous. Still, they are theoretical. Their relationship with experience might be less secure, since any probability statement is only reproduced in observations by sufficiently large numbers of independent data. Thus, Boltzmann would have allowed for exceptions in the relationship between theory and observation, but not in the relation between premisses and conclusion.

He continues by saying what he means by probability, and repeats its equivocation as a fraction of time and the relative number of particles that we have seen earlier in 1868:

> If one wants [...] to build up an exact theory [...] it is before all necessary to determine the probabilities of the various states that one and the same molecule assumes in the course of a very long time, and that occur simultaneously for different molecules. That is, one must calculate how the number of those molecules whose states lie between certain limits relates to the total number of molecules (Abh. I p. 317).

However, this equivocation is not vicious. For most of the paper the intended meaning of probability is always the relative number of molecules with a particular molecular state. Only at the final stages of his paper (Abh. I, p. 400) does the time-average interpretation of probability (suddenly) recur.

Boltzmann says that both Maxwell and he had attempted the determination of these probabilities for a gas system but without reaching a complete solution. Yet, on a closer inspection, "it seems not so unlikely that these probabilities can be derived on the basis of the equations of motion alone..." (Abh. I, p. 317). Indeed, he announces, he has solved this problem for gases whose molecules consist of an arbitrary number of atoms. His aim is to prove that whatever the initial distribution of state in such a system of gas molecules, it must inevitably approach the distribution characterized by the Maxwellian form (ibid. p. 320).

The next section specializes to the simplest case of monatomic gases and also provides a more complete specification of the problem he aims to solve. The gas molecules are contained in a fixed vessel with perfectly elastic walls. They interact with each other only when they approach each other at very small distances. These interactions can be mimicked as collisions between elastic bodies. Indeed, these bodies are modeled as hard spheres (Abh I, p. 320). Boltzmann represents the state of the gas by a time-dependent distribution function $f_t(\vec{v})$, called the "distribution of state", which gives us, at each time t, the relative number of molecules with velocity between $\vec{v}$ and $\vec{v} + d^3\vec{v}$.[29]

He also states two more special assumptions:

[29] Actually Boltzmann formulated the discussion in terms of a distribution function over kinetic energy rather than velocity. I have transposed this into the latter, nowadays more common formulation.

1. Already in the initial state of the gas, each direction of velocity is equally probable. That is:

$$f_0(\vec{v}) = f_0(v). \tag{47}$$

   It is assumed as obvious that this will also hold for any later time.

2. The gas is spatially uniform within the container. That is, the relative number of molecules with their velocities in any given interval, and their positions in a particular spatial region R does not depend on the location of R in the available volume.

The next and crucial assumption used by Boltzmann to calculate the change in the number of particles with a velocity $\vec{v}_1$ per unit time, is the *Stoßzahlansatz*, (29) and (30).

For modern readers, there are also a few unstated assumptions that go into the construction of this equation. First, the number of molecules must be large enough so that the (discrete) distribution of their velocities can be well approximated by a continuous and differentiable function f. Secondly, f changes under the effect of binary collisions only. This means that the density of the gas should be low (so that three-particle collisions can be ignored) but not too low (which would make collisions too infrequent to change f at all). These two requirements are already hard enough to put in a mathematically precise form. The modern explicitation is that of taking the so-called Boltzmann-Grad limit (cf. paragraph 6.4). The final (unstated) assumption is that all the above assumptions remain valid in the course of time.

He addresses his aim by constructing a differentio-integral evolution equation for f_t, by taking the difference of (29) and (30) and integrating over all variables except $\vec{v}_1$ and t. The result (in a modern notation) is the *Boltzmann equation*:

$$\frac{\partial f_t(\vec{v}_1)}{\partial t} = N \int_0^d b db \int_0^{2\pi} d\phi \int_{\mathbb{R}^3} d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \left(f_t(\vec{v}_1{}') f_t(\vec{v}_2{}') - f_t(\vec{v}_1) f_t(\vec{v}_2) \right) \tag{48}$$

which describes the change of f in the course of time, when this function at some initial time is given. (Recall from paragraph 3.3 that the primed velocities are to be thought of as functions of the unprimed velocities and the geometrical parameters of the collision: $\vec{v}_i' = \vec{v}_i{}'(\vec{v}_1, \vec{v}_2, b, \phi)$, and d denotes the diameter of the hard spheres.)

4.2.1 The H-theorem

Assuming that the Boltzmann equation (48) is valid for all times, one can prove, after a few well-known manipulations, that the following quantity

$$H[f_t] := \int f_t(\vec{v}) \ln f_t(\vec{v}) d^3\vec{v} \tag{49}$$

decreases monotonically in time, i.e.

$$\frac{dH[f_t]}{dt} \leq 0; \tag{50}$$

as well as its stationarity for the Maxwell distribution, i.e.:

$$\frac{dH[f_t]}{dt} = 0 \ \ (\forall t) \ \ \text{iff} \ \ f_t(v) = Ae^{-Bv^2}. \tag{51}$$

Boltzmann concludes Section I of the paper as follows:

> It has thus been rigorously proved that whatever may have been the initial distribution of kinetic energy, in the course of time it must necessarily approach the form found by Maxwell. [...] This [proof] actually gains much in significance because of its applicability to the theory of multi-atomic gas molecules. There too, one can prove for a certain quantity $[H]$ that, because of the molecular motion, this quantity can only decrease or in the limiting case remain constant. Thus, one may prove that because of the atomic movement in systems consisting of arbitrarily many material points, there always exists a quantity which, due to these atomic movements, cannot increase, and this quantity agrees, up to a constant factor, exactly with the value that I found in [(Boltzmann 1871c)] for the well-known integral $\int dQ/T$.
>
> This provides an analytical proof of the Second Law in a way completely different from those attempted so far. Up till now, one has attempted to proof that $\int dQ/T = 0$ for a reversible (*umkehrbaren*) cyclic[30] process, which however does not prove that for an irreversible cyclic process, which is the only one that occurs in nature, it is always negative; the reversible process being merely an idealization, which can be approached more or less but never perfectly. Here, however, we immediately reach the result that $\int dQ/T$ is in general negative and zero only in a limit case... (Abh. I, p. 345)

Thus, as in his 1866 paper, Boltzmann claims to have a rigorous, analytical and general proof of the Second Law. From our study of the paper until now, (i.e. section I) it appears that Klein's interpretation is more plausible than von Plato's. I postpone a further discussion of this dispute to paragraph 4.2.3, after a brief look at the other sections of the paper.

4.2.2 Further sections of Boltzmann (1872)

Section II is entitled "Replacement of integrals by sums" and devoted to a repetition of the earlier arguments, now assuming that the kinetic energies of the molecules can only take values in a discrete

[30]The term "cyclic" is missing in Brush's translation, although the original text does speak of "Kreisprozeß". The special notation $\oint$ for cyclic integrals was not introduced until much later.

set $\{0, \epsilon, 2\epsilon, \ldots, p\epsilon\}$. Boltzmann shows that in the limit $\epsilon \longrightarrow 0$, $p\epsilon \longrightarrow \infty$ the same results are recovered.

Many readers have been surprised by this exercise, which seems rather superfluous both from a didactic and a logical point of view. (However, some have felt that it foreshadowed the advent of quantum theory.) Boltzmann offers as motivation for the detour that the discrete approach is clearer than the previous one. He argues that integrals only have a symbolic meaning, as a sum of infinitely many infinitesimal elements, and that a discrete calculation yields more understanding. He does not argue, however, that it is closer to physical reality. Be that as it may, the section does eventually take the limit, and recovers the same results as before.

The third section treats the case where the gas is non-uniform, i.e., when condition 2 above is dropped. For this case, Boltzmann introduces a generalized distribution function $f_t(\vec{r}, \vec{v})$, such that $f_t d^3\vec{r} d^3\vec{v}$ represents the relative number of particles with a position in a volume element $d^3\vec{r}$ around $\vec{r}$ and a velocity in an element $d^3\vec{v}$ around $\vec{v}$.

He obtains a corresponding generalized Boltzmann equation:

$$\frac{\partial f_t(\vec{r}, \vec{v})}{\partial t} + \vec{v} \cdot \nabla_x f_t + \frac{\vec{F}}{m} \cdot \nabla_v f_t = \\ N \int b db d\phi d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \left(f_t(\vec{r}, \vec{v}_1')) f_t(\vec{r}, \vec{v}_2') - f_t(\vec{r}, \vec{v}_1)) f_t(\vec{r}, \vec{v}_2) \right) \tag{52}$$

where $\vec{F}$ denotes an external force field on the gas. The quantity H now takes the form $H[f_t] := \int f_t(\vec{r}, \vec{v}) d^3\vec{r} d^3\vec{v}$; and a generalization of the H-theorem $dH/dt \leq 0$ is obtained.

The last three sections are devoted to polyatomic molecules, and aim to obtain generalized results for this case too. The key ingredient for doing so is, of course, an appropriately generalized *Stoßzahlansatz*. The formulation of this assumption is essentially the same as the one given in his paper on poly-atomic molecules (1871a), which was later shown wrong and corrected by Lorentz. I will not go into this issue (cf. Lorentz 1887, Boltzmann 1887b, Tolman 1938).

An interesting passage occurs at the very end of the paper, where he expands on the relationship between H and entropy. He considers a monatomic gas in equilibrium. The stationary distribution of state is given as:

$$f^*(\vec{r}, \vec{v}) = V^{-1} \left(\frac{3m}{4\pi T} \right)^{3/2} \exp(\frac{-3mv^2}{4T}) \tag{53}$$

where V is the volume of the container. (Note that in comparison with (27), Boltzmann adopts units for temperature that make $k = 2/3$.) He shows that

$$H[f^*] := \int f^* \log f^* dx dv = -N \log V \left(\frac{4\pi T}{3m} \right)^{3/2} - \frac{3}{2} N; \tag{54}$$

which agrees (assuming $S = -kNH[f^*]$) with the thermodynamical expression for the ideal gas

(16) up to an additive constant. A similar result holds for the polyatomic gas.

4.2.3 Remarks and problems

1. The role of probability. As we have seen, the H-theorem formed the basis of a renewed claim by Boltzmann to have obtained a theorem corresponding to the full Second Law (i.e. including both parts) at least for gases. A main difference from his 1866 claim, is that he now strongly emphasizes the role of probability calculus in his derivation. It is clear that the conception of probability expounded here is thoroughly frequentist and that he takes 'the laws of probability' as empirical statements. Furthermore, probabilities can be fully expressed in mechanical terms: the probability distribution f is nothing but the relative number of particles whose molecular states lie within certain limits. Thus, there is no conflict between his claims that on the one hand, "the problems of the mechanical theory of heat are really problems in probability calculus" and that the probabilities themselves are derived on the basis of the equations of motion alone, on the other hand. Indeed, it seems to me that Boltzmann's emphasis on the crucial role of probability in this paper is only intended to convey that probability theory provides a particularly useful and appropriate language for discussing mechanical problems in gas theory. There is no indication in this paper yet that probability theory could play a role by furnishing assumptions of a non-mechanical nature, i.e., independent of the equations of motion (cf. Boltzmann & Nabl 1904, p. 520).

2. The role of the Stoßzahlansatz. Note that Boltzmann stresses the generality, rigour and "analyticity" of his proof. He puts no emphasis on the special assumptions that go into the argument. Indeed, the *Stoßzahlansatz*, later identified as the key assumption that is responsible for the time-asymmetry of the H-theorem, is announced as follows

> The determination [of the number of collisions] can only be obtained in a truly tedious manner, by consideration of the relative velocities of both particles. But since this consideration has, apart from its tediousness, not the slightest difficulty, nor any special interest, and because the result is so simple that one might almost say it is self-evident I will only state this result." (Abh. I, p. 323)

It thus seems natural that Boltzmann's contemporaries must have understood him as claiming that the H-theorem followed necessarily from the dynamics of the mechanical gas model.[31] I can find no evidence in the paper that he intended this claim to be read with a pinch of salt, as (von Plato 1991, p.. 81) has argued.

[31] Indeed this is *exactly* how Boltzmann's claims were understood. For example, the recommendation written in 1888 for his membership of the Prussian Academy of Sciences mentions as his main feat that Boltzmann had proven that, whatever its initial state, a gas must necessarily approach the Maxwellian distribution (Kirsten & Körber 1975, p.109).

Is there then no evidence at all for von Plato's reading of the paper? Von Plato refers to a passage from Section II, where Boltzmann repeats the previous analysis by assuming that energy can take on only discrete values, and replacing all integrals by sums. He recovers, of course, the same conclusion, but now adds a side remark, which touches upon the case of non-uniform gases:

> Whatever may have been the initial distribution of states, there is one and only one distribution which will be approached in the course of time. [...] This statement has been proved for the case where the distribution of states was already initially uniform. It must also be valid when this is not the case, i.e. when the molecules are initially distributed in such a way that in the course of time they mix among themselves more and more, so that after a very long time the distribution of states becomes uniform. This will always be the case, with the exception of very special cases, e.g. when all molecules were initially situated along a straight line, and were reflected by the walls onto this line (Abh. I, p. 358).

It is this last remark that, apparently, led to the view that after all Boltzmann did already conceive of exceptions to his claims. However, I should say that this passage does not convince me. True enough, Boltzmann in the above quote indicates that there are exceptions. But he mentions them only in connection with an *extension* of his results to the case when the gas is not initially uniform, i.e. when condition (2) above is dropped. There can be no doubt that under the assumption of the conditions (1) and (2), Boltzmann claimed the rigorous validity of the H-theorem. (Curiously, his more systematic treatment of the non-uniform gas (Section III of (1872)) does not mention any exception to the claim that "H can only decrease" (Abh. I p. 362).

As a matter of fact, when Loschmidt formulated the objection, it happened to be by means of an example of a non-uniform gas (although nothing essential depended on this). Thus, if Boltzmann had in 1872 a "full hand against his future critics", as von Plato claims, one would expect his reply to Loschmidt's objection to point out that Loschmidt was correct but that he had already anticipated the objection. Instead, he accused Loschmidt of a fallacy (see paragraph 4.3 below).

But apart from the historical issue of whether Boltzmann did or did not envisage exceptions to his H-theorem, it seems more important to ask what kind of justification Boltzmann might have adduced for the *Stoßzahlansatz*. An attempt to answer this question must be somewhat speculative, since, as we have seen, Boltzmann presented the assumption as "almost self-evident" and "having no special interest", and hence presumably as not in need of further explanation. Still the following remarks may be made with some confidence.

First, we have seen that Maxwell's earlier usage of the assumption was never far away from an argument from insufficient reason. Thus, in his approach, one could think of the *Stoßzahlansatz* as expressing that we have no reason to expect any influence or correlation between any pair of particles

that are about to collide. The assumption would then appear as a probabilistic assumption, reflecting a 'reasonable judgment', independent from mechanics.

In contrast, Boltzmann's critique of Maxwell's approach (cf. footnote 16) suggests that he did not buy this arguments from insufficient reason. But since the *Stoßzahlansatz* clearly cannot be conceived of as an assumption about dynamics —like the ergodic hypothesis—, this leaves only the option that it must be due to a special assumption about the mechanical state of the gas. Indeed, in the years 1895-6, when Boltzmann acknowledged the need for the *ansatz* in the proof of his H-theorem more explicitly —referring to it as "Assumption A" (Boltzmann 1895) or "the hypothesis of molecular disorder" (Boltzmann 1896)—, he formulated it as an assumption *about* the state of the gas.

Yet, even in those years, he would also formulate the hypothesis as expressing that "haphazard governs freely" (Boltzmann 1895, Abh. III, p. 546) or "that the laws of probability are applicable for finding the number of collisions" (Boltzmann 1895b). Similarly, he describes states for which the hypothesis fails as contrived "so as to intentionally violate the laws of probability"(Boltzmann 1896, §3). However, I think these quotations should not be read as claims that the *Stoßzahlansatz* was a consequence of probability theory itself. Rather, given Boltzmann's empirical understanding of "the laws of probability", they suggest that Boltzmann thought that, as a matter of empirical fact, the assumption would 'almost always' hold, even if the gas was initially very far from equilibrium.

3. The H-theorem and the Second Law. Note that Boltzmann misconstrues, or perhaps understates, the significance of his results. Both the Boltzmann equation and the H-theorem refer to a body of gas in a fixed container that evolves in isolation from its environment. There is no question of heat being exchanged by the gas during a process, let alone in an irreversible cyclic process. His comparison in the quotation on page 46 with Clausius' integral $\int dQ/T$ (i.e. $\oint đQ/T$ in equation (18) above) is therefore really completely out of place.

The true import of Boltzmann's results is rather that they provide (i) a generalization of the entropy concept to non-equilibrium states,[32] and (ii)a claim that this non-equilibrium entropy $-kH$ increases monotonically as the isolated gas evolves for non-equilibrium towards an equilibrium state. The relationship with the Second Law is, therefore, somewhat indirect: On the one hand, Boltzmann proves much more than was required, since the second law does not speak of non-equilibrium entropy, nor of monotonic increase; on the other hand it proves also less, since Boltzmann does not consider the increase of entropy in general adiabatic processes.

[32]Boltzmann emphasized that his expression for entropy should be seen as an *extension* of thermodynamic entropy to non-equilibrium states in (1877b, Abh. II, p. 218; 1896, §5) . Of course there is no guarantee that this generalization is the *unique* candidate for a non-equilibrium entropy.

6.4 Lanford's approach to the Boltzmann equation

We now turn to consider some modern approaches to non-equilibrium statistical mechanics. Of these, the approach developed by Lanford and others (cf. Lanford 1975, Lanford 1976, Lanford 1981, Spohn 1991, Cercignani, Illner &Pulvirenti 1994) deserves special attention because it stays conceptually closer to Boltzmann's 1872 work on the Boltzmann equation and the H-theorem than any other modern approach to statistical physics. Also, the problem Lanford raised and tried to answer is one of no less importance than the famous reversibility and recurrence objections. Furthermore, the results obtained are the best efforts so far to show that a statistical reading of the Boltzmann equation or the H-theorem might hold for the hard spheres gas.

The question Lanford raised is that of the consistency of the Boltzmann equation and the underlying Hamiltonian dynamics. Indeed, if we consider the microstate of a mechanical system such as a dilute gas, it seems we can provide two competing accounts of its time evolution.

(1) On the one hand, given the mechanical microstate x_0 of a gas, we can form the distribution of state $f(\vec{r},\vec{v})$, such that $f(\vec{r},\vec{v})d^3\vec{v}d^3\vec{r}$ gives the relative number of molecules with a position between $\vec{r}$ and $\vec{r}+d^3\vec{r}$ and velocity between $\vec{v}$ and $\vec{v}+d^3\vec{v}$. Presumably, this distribution should be uniquely determined by the microstate x_0. Let us make this dependence explicit by adopting the notation $f^{[x_0]}$. This function, then, should ideally serve as an initial condition for the Boltzmann equation (48), and solving this equation —assuming, that is, that it, that it has a unique solution— would give us the shape of the distribution function at a later time, $f_t^{[x_0]}(\vec{r},\vec{v})$.

(2) On the other hand, we can evolve the microstate x_0 for a time t with the help of the Hamiltonian equations. That will give us $x_t = T_t x_0$. This later state x_t will then also determine a distribution of state $f^{[x_t]}(\vec{r},\vec{v})$.

It is a sensible question whether these two ways of obtaining a later distribution of state from an initial microstate are the same, i.e. whether the two time evolutions are consistent. In other words, the problem is whether the diagram below commutes:

$$\begin{array}{ccc} x_0 & \overset{\text{Hamilton}}{\longrightarrow} & x_t \\ \downarrow & & \downarrow \\ f^{[x_0]} & \overset{\text{Boltzmann}}{\longrightarrow} & f_t^{[x_0]} \overset{?}{=} f^{[x_t]} \end{array} \tag{136}$$

The first issue that has to be resolved here is the precise relation between a microstate and the distribution of state f. It is obvious that, in so far as this function represents the physical property of a gas system, it should be determined by the momentary microstate x. It is also clear, that in so far as it is assumed to be continuous and differentiable in time in order to obey the Boltzmann equation, this cannot be literally and exactly true.

So let us assume, as Boltzmann did, that the gas consists of N hard spheres, each of diameter

d and mass m, contained in some fixed bounded spatial region Λ with volume $|\Lambda| = V$. Given a microstate x of the system one can form the 'exact' distribution of state:

$$F^{[x]}(\vec{r},\vec{v}) := \frac{1}{N}\sum_{i}^{N}\delta^3(\vec{r}-\vec{q}_i)\delta^3(\vec{v}-\frac{\vec{p}_i}{m}). \tag{137}$$

This distribution is, of course, not a proper function, and being non-continuous and non-differentiable, clearly not a suitable object to plug into the Boltzmann equation. However, one may reasonably suppose that one ought to be able to express Boltzmann's ideas in a limit in which the number of particles, N, goes to infinity. However, this limit clearly must be executed with care.

On the one hand, one ought to keep the gas dilute, so that collisions involving three or more particles will be rare enough so that they can safely be ignored in comparison to two-particle collisions. On the other hand, the gas must not be so dilute that collisions are altogether too rare to contribute to a change of f. The appropriate limit to consider, as Lanford argues, is the so-called Boltzmann-Grad limit in which $N \longrightarrow \infty$, and:[67]

$$\frac{Nd^2}{V} = \text{constant} > 0. \tag{138}$$

Denote this limit as "$N \xrightarrow{\text{BG}} \infty$", where it is implicitly understood that $d \propto N^{-1/2}$. The hope is then that in this Boltzmann-Grad limit, the exact distribution $F^{[x^N]}$ will tend to a continuous function that can be taken as an appropriate initial condition for the Boltzmann equation. For this purpose, one has to introduce a relevant notion of convergence for distributions on the μ-space $\Lambda \times \mathbb{R}^3$. A reasonable choice is to say that an arbitrary sequence of distributions f_n (either proper density functions or in the distributional sense) converges to a distribution f, $f_n \longrightarrow f$, iff the following conditions hold:

$$\text{For each rectangular parallelepiped } \Delta \subset \Lambda \times \mathbb{R}^3 \ :$$

$$\lim_{n\longrightarrow\infty}\int_{\Delta} f_N d^3\vec{r}d^3\vec{v} = \int_{\Delta} f d^3\vec{r}d^3\vec{v}, \tag{139}$$

$$\text{and} \qquad \lim_{n\longrightarrow\infty}\int \vec{v}^2 f_n d^3\vec{r}d^3\vec{v} = \int \vec{v}^2 f d^3\vec{r}d^3\vec{v}, \tag{140}$$

where the second condition is meant to guarantee the convergence of the mean kinetic energy.

It is also convenient to introduce some distance function between (proper or improper) distributions that quantifies the sense in which one distribution is close to another in the above sense. That

[67] The condition can be explained by the hand-waving argument that Nd^2/V is proportional to the 'mean free path', i.e. a typical scale for the distance traveled by a particle between collisions, or also by noting that the collision integral in the Boltzmann equation is proportional to Nd^2/V, so that by keeping this combination constant, we keep the Boltzmann equation unchanged.

is to say, one might define some distance $d(f, g)$ between density functions on $\Lambda \times R^3$ such that

$$d(f_n, f) \longrightarrow 0 \Longrightarrow f_n \longrightarrow f. \tag{141}$$

There are many distance functions that could do this job, but I won't go into the question of how to pick out a particular one.

The hope is then, to repeat, that $F^{[x^N]} \longrightarrow f$ in the above sense when $N \overset{BG}{\longrightarrow} \infty$, where f is sufficiently smooth to serve as an initial condition in the Boltzmann equation, and that with this definition, the Boltzmannian and Hamiltonian evolution become consistent in the sense that the diagram (136) commutes. But clearly this will still be a delicate matter. Indeed, increasing N means a transition from one mechanical system to another with more particles. But there is no obvious algorithm to construct the state x^{N+1} from x^N, and thus no way to enforce convergence on the level of individual states.

Still, one might entertain an optimistic guess, which, if true, would solve the consistency problem between the Boltzmann and the Hamiltonian evolution in an approximate fashion if N is very large.

> OPTIMISTIC GUESS: If $F^{[x_0^N]}$ is near to f then $F^{[x_t^N]}$ is near to f_t for all $t > 0$, and where f_t is the solution of the Boltzmann equation with initial condition f.

As Lanford (1976) points out, the optimistic guess cannot be right. This is an immediate consequence of the reversibility objection: Indeed, suppose it were true for all $x \in \Gamma$, and $t > 0$. (Here, we momentarily drop the superscript N from x^N to relieve the notation.) Consider the phase point Rx obtained from x by reversing all momenta: $R(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = (\vec{q}_1, -\vec{p}_1; \ldots, ; \vec{q}_N, -\vec{p}_N)$. If $F^{[x]}(\vec{r}, \vec{v})$ is near to some distribution $f(\vec{r}, \vec{v})$, then $F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But as x evolves to x_t, Rx_t evolves to $T_t R x_t = R T_{-t} x_t = Rx$. Hence $F^{[T_t R x_t]}(\vec{r}, \vec{v}) = F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But the validity of the conjecture for Rx_t would require that $F^{[T_t R x_t]}(\vec{r}, \vec{v})$ is near to $f_t(\vec{r}, -\vec{v})$ and these two distributions of state are definitely not near to each other, except in some trivial cases.

But even though the optimistic guess is false in general, one might hope that it is 'very likely' to be true, with some overwhelming probability, at least for some finite stretch of time. In order to make such a strategy more explicit, Lanford takes recourse to a probability measure on Γ, or more precisely a sequence of probability measures on the sequence of Γ_N's.

Apart from thus introducing a statistical element into what otherwise would have remained a purely kinetic theory account of the problem, there is a definite advantage to this procedure. As mentioned above, there is no obvious algorithm to construct a sequence of microstates in the Boltzmann-Grad limit. But for measures this is different. The microcanonical measure, for example is not just a measure for the energy hypersurface of one N-particles-system; it defines an algorithmic sequence

of such measures for each N.

In the light of this discussion, we can now state Lanford's theorem as follows (Lanford 1975, 1976):

> LANFORD'S THEOREM: Let $t \mapsto f_t$ be some solution of the Boltzmann equation, say for $t \in [0, a) \subset \mathbb{R}$. For each N, let Δ_N denote the set in the phase space Γ_N of N particles, on which $F^{[x^N]}$ is near to f_0 (the initial condition in the solution of the Boltzmann equation) in the sense that for some chosen distance function d and for tolerance $\epsilon > 0$:
>
> $$\Delta_N = \{x^N \in \Gamma_N \ : \ d(F^{[x^N]}, f_0) < \epsilon\}. \tag{142}$$
>
> Further, for each N, conditionalize the microcanonical measure μ_N on Δ_N:
>
> $$\mu_{\Delta,N}(\cdot) := \mu_N(\cdot|\Delta_N). \tag{143}$$
>
> In other words, $\mu_{\Delta,N}$ is a sequence of measures on the various Γ_N that assign measure 1 to the set of microstates $x^N \in \Gamma_N$ that are close to f_0 in the sense that $d(F^{[x^N]}, f_0) < \epsilon$. Then: $\exists \tau$, $0 < \tau < a$ such that for all t with $0 < t < \tau$:
>
> $$\mu_{\Delta,N}(\{x^N \in \Gamma_N \ : \ d(F^{[x_t^N]}, f_t) < \epsilon\}) > 1 - \delta \tag{144}$$
>
> where $\delta \longrightarrow 0$ as both $\epsilon \longrightarrow 0$ and $N \xrightarrow{BG} \infty$.

In other words: as judged from the microcanonical measure on Γ_N restricted to those states x^N that have their exact distribution of state close to a given initial function f_0, a very large proportion $(1-\delta)$ evolve by the Hamiltonian dynamics in such a way that their later exact distribution of state $F^{[x_t^N]}$ remains close to the function f_t, as evolved from f_0 by the Boltzmann equation.

6.4.1 Remarks

Lanford's theorem shows that a statistical and approximate version of the Boltzmann equation can be derived from Hamiltonian mechanics and the choice of an initial condition in the Boltzmann-Grad limit. This is a remarkable achievement, that in a sense vindicates Boltzmann's intuitions. According to Lanford (1976, p. 14), the theorem says that the approximate validity of the Boltzmann equation, and hence the H-theorem, can be obtained from mechanics alone and a consideration of the initial conditions.

Still the result established has several remarkable features, all of which are already acknowledged by Lanford. First, there are some drawbacks that prevent the result from having practical impact for

the project of justifying the validity of the Boltzmann equation in real-life physical applications. The density of the gas behaves like N/d^3, and in the Boltzmann-Grad limit this goes to zero. The result thus holds for extremely rarified gases. Moreover, the length of time for which the result holds, i.e. τ, depends on the constant in (138), which also provides a rough order of magnitude for the mean free path of the gas . It turns out that, by the same order of magnitude considerations, τ is roughly two fifths of the mean duration between collisions. This is a disappointingly short period: in air at room temperature and density, τ is in the order of microseconds. Thus, the theorem does not help to justify the usual applications of the Boltzmann equation to macroscopic phenomena which demand a much longer time-scale.

Yet note that the time scale is not trivially short. It would be a misunderstanding to say that the theorem establishes only the validity of the Boltzmann equation for times so short that the particles have had no chance of colliding: In two fifths of the mean duration between collisions, about 40 % of the particles have performed a collision.

Another issue is that in comparison with Boltzmann's own derivation no explicit mention seems to have been of the *Stoßzahlansatz*. In part this is merely apparent. In a more elaborate presentation (cf. Lanford 1975, 1976), the theorem is not presented in terms of the microcanonical measure, but an arbitrary sequence of measures ν_N on (the sequence of phase spaces) Γ_N. These measures are subject to various assumptions. One is that each ν_N should be absolutely continuous with respect to the microcanonical measure μ_N, i.e. ν_N should have a proper density function

$$d\nu_N(x) = n_N(x_1, \ldots x_N)dx_1 \cdots x_N \tag{145}$$

where $x_i = (\vec{q_i}, \vec{p_i})$ denotes the canonical coordinates of particle i. Further, one defines, for each N and $m < N$, the reduced density functions by

$$n_N^{(m)}(x_1, \ldots x_m) := \frac{N!}{(N-m)!}\frac{1}{N^m}\int n_N(x_1, \ldots x_N)dx_{m+1} \cdots dx_N \tag{146}$$

i.e. as (slightly renormalized) marginal probability distributions for the first m particles. The crucial assumption is now that

$$\lim_{N \xrightarrow{\text{BG}} \infty} n_N^{(m)}(x_1, \ldots x_m) = n^{(1)}(x_1) \cdots n^{(1)}(x_m) \tag{147}$$

uniformly on compact subsets of $(\Lambda \times \mathbb{R}^3)^m$. This assumption (which can be shown to hold for the microcanonical measures) is easily recognized as a measure-theoretic analogy to the *Stoßzahlansatz*. It demands, in the Boltzmann-Grad limit, statistical independence of the molecular quantities for any pair or m-tuple of particles at time $t = 0$. As Lanford also makes clear, it is assumption (146) that

would fail to hold if we run the construction of the reversibility objection; (i.e. if we follow the states x in Δ_N for some time t, $0t < \tau$, then reverse the momenta, and try to apply the theorem to the set $\Delta'_N = \{Rx_t \ : \ x \in \Delta_N\}$).

But another aspect is more positive. Namely: Lanford's theorem does not need to assume explicitly that the *Stoßzahlansatz* holds *repeatedly*. Indeed a remarkable achievement is that once the factorization condition (146) holds for time $t = 0$ it will also hold for $0 < t < \tau$, albeit in a weaker form (as convergence in measure, rather than uniform convergence). This is sometimes referred to as "propagation of chaos" (Cercignani, Illner &Pulvirenti 1994).

But the main conceptual problem concerning Lanford's theorem is where the apparent irreversibility or time-reversal non-invariance comes from. On this issue, various opinions have been expressed. Lanford (1975, p. 110) argues that irreversibility is the result of passing to the Boltzmann-Grad limit. Instead, Lanford (1976) argues that it is due to condition (146) plus the initial conditions (i.e.: $x_N \in \Delta_N$).

However, I would take a different position. The theorem equally holds for $-\tau < t < 0$, with the proviso that f_t is now a solution of the anti-Boltzmann equation. This means that the theorem is, in fact, invariant under time-reversal.

6.5 The BBGKY approach

The so-called BBGKY-hierarchy (named after Bogolyubov, Born, Green, Kirkwood and Yvon) is a unique amalgam of the description of Gibbs and the approach of Boltzmann. The goal of the approach is to describe the evolution of ensembles by means of reduced probability densities, and to see whether a Boltzmann-like equation can be obtained under suitable conditions —and thereby an approach to statistical equilibrium.

First, consider an arbitrary time-dependent probability density ρ_t. The evolution of ρ is determined via the Liouville-equation by the Hamiltonian:

$$\frac{\partial \rho_t}{\partial t} = \{H, \rho\}. \tag{148}$$

Central in the present approach is the observation that for relevant systems in statistical mechanics, this Hamiltonian will be symmetric under permutation of the particles. Indeed, the Hamiltonian for a system of N indistinguishable particles usually takes the form

$$H(\vec{q}_1, \vec{p_1}; \ldots; \vec{q}_N, \vec{p}_N) = \sum_{=1}^{N} \frac{\vec{p_i}^2}{2m} + \sum_{i}^{N} V(\vec{q_i}) + \sum_{i<j}^{N} \phi(\|q_i - \vec{q_j}\|) \tag{149}$$

S-C. ZHANG

To see a world in a grain of sand

Shou-Cheng Zhang

Department of Physics, Stanford University, Stanford, CA 94305

Throughout John Wheeler's career, he wrestled with big issues like the fundamental length, the black hole and the unification of quantum mechanics and relativity. In this essay, I argue that solid state physics – historically the study of silicon, semiconductors and sand grains – can give surprisingly deep insights into the big questions of the world.

Contents

I. INTRODUCTION

Modern physics is built upon three principal pillars, quantum mechanics, special and general relativity. Historically, these principles were developed as logically independent extensions of classical Newtonian mechanics. While each theory constitutes a logically self-consistent framework, unification of these fundamental principles encountered unprecedented difficulties. Quantum mechanics and special relativity were unified in the middle of the last century, giving birth to relativistic quantum field theory. While tremendously successful in explaining experimental data, ultraviolet infinities in the calculations hint that the theory can not be in its final form. Unification of quantum mechanics with general relativity proves to be a much more difficult task and is still the greatest unsolved problem in theoretical physics.

In view of the difficulties involved with unifying these principles, we can ask a simple but rather bold question: Is it possible that the three principles are not logically independent, but rather there is a hierarchical order in their logical dependence? In particular, we notice that both relativity principles can be formulated as statements of symmetry. When applying non-relativistic quantum mechanics to systems with a large number of degrees of freedom, we sometimes find that symmetries can emerge in the low energy sector, which are not present in the starting Hamiltonian. Therefore, there is a logical possibility that one could start from a single non-relativistic Schrödinger equation for a quantum many-body problem, and discover relativity principles emerging in the low energy sector. If this program can indeed be realized, a grand synthesis of fundamental physics can be achieved. Since non-relativistic quantum mechanics is a finite and logically self-consistent framework, everything derived from it should be finite and logically consistent as well.

The Standard Model in particle physics is described by a relativistic quantum field theory and is experimentally verified below the energy scale of $10^3 GeV$. On the other hand, the Planck energy scale, where quantum gravitational force becomes important, is at $10^{19} GeV$. Therefore, we need to extrapolate 16 orders of magnitude to guess the new physics beyond the standard framework of relativistic quantum field theory. It is quite conceivable that Einstein's principle of relativity is not valid at Planck's energy scale, it could emerge at energies much lower compared to the Planck's energy scale through the magic of renormalization group flow. This situation is analogous to one in condensed matter physics, which deals with phenomena at much lower absolute energy scales. The "basic" laws of condensed

matter physics are well-known at the Coulomb energy scale of $1 \sim 10eV$; almost all condensed matter systems can be well described by a non-relativistic Hamiltonian of the electrons and the nuclei(Laughlin and Pines, 2000). However, this model Hamiltonian is rather inadequate to describe the various emergent phenomena, like superconductivity, superfluidity, the quantum Hall effect (QHE) and magnetism, which all occur at much lower energy scales, typically of the order of $1meV$. These systems are best described by "effective quantum field theories", not of the original electrons, but of the quasi-particles and collective excitations. In this lecture, I shall give many examples where these "effective quantum field theories" are relativistic quantum field theories or topological quantum field theories, bearing great resemblance to the Standard Model of elementary particles. The collective behavior of many strongly interacting degrees of freedom is responsible for these striking emergent phenomena. The laws governing the quasi-particles and the collective excitations are very different from the laws governing the original electrons and nuclei(Anderson, 1972). This observation inspires us to construct models of elementary particles by conceptually visualizing them as quasi-particles or collective excitations of a quantum many-body system, whose basic constituents are governed by a simple non-relativistic Hamiltonian. This point of view is best summarized by the following diagram:

Planck energy at $10^{19}GeV$	$\Leftrightarrow$	Coulomb energy at $10eV$
$\uparrow$?		$\downarrow$
Standard Model at 10^3GeV	$\Leftrightarrow$	Superconductivity, QHE, Magnetism etc at 1 meV
Relativistic quantum field theory of elementary particles		Effective quantum field theory of quasi-particles

The conceptual similarity between particle physics and condensed matter physics has played a very important role in the history of physics. A crucial ingredient of the Standard Model, the idea of spontaneously broken symmetry and the Higgs mechanism, first originated from the BCS theory of superconductivity. This example vividly shows that the physical vacuum is not empty, but a condensed state of many interacting degrees of freedom. Another fundamental concept is the idea of renormalization group transformation, which was simultaneously developed in the context of particle physics and in the study of critical phenomena. From the theory of renormalization group, we learned that symmetries can emerge at the low energy sector, without being postulated at the microscopic level. Today, as physicists face unprecedented challenges of unifying quantum mechanics with relativity, and tackling the problem of quantum gravity, it is useful to look at these historic successes for inspiration. A new era of close interaction between condensed matter physics and particle physics could shed light on these grand challenges of theoretical physics.

II. EXAMPLES OF EMERGENCE IN CONDENSED MATTER SYSTEMS

In this section, we review some well-known examples in condensed matter physics, where one starts from a quantum many-body system at high energies and arrives at a relativistic or topological field theory of the low energy quasi-particles and elementary excitations. The high energy models often look simple and innocuous, yet the emergent low energy phenomena and their effective field theory description are profound and beautiful.

A. 2+1 dimensional QED from superfluid helium films

Let us first start from the physics of a superfluid film. The mean velocity of the helium atoms are significantly lower compared to the speed of light, therefore, relativistic effects of the atoms can be completely neglected. The basic non-relativistic Hamiltonian for this system of identical bosons can be expressed in the following close form:

$$H = \frac{1}{2m}\sum_n \vec{p}_n^2 + \sum_{n<n'} V(x_n - x_{n'}) \tag{1}$$

where V is the inter-atomic potential, whose form depends on the details of the system. However, for a large class of generic interaction potentials, the system flows towards a universal low energy attractive fix point, namely the superfluid ground state. At typical inter-atomic energy scales of a few eV's, helium atoms are the correct dynamic variables, and the Hamiltonian (1) is the correct model Hamiltonian. However, at the energy scale characteristic of the superfluid transition, which is of the order of $1K \sim 10^{-4}eV$, the correct dynamical variables are sound wave modes and the vortices of the superfluid film. (See fig. (1). for an illustration).

The remarkable thing is that the effective field theory model for these low energy degrees of freedom is exactly the relativistic quantum electrodynamics (QED) in $2+1$ dimensions! This connection was established by the work of Ambegaokar, Halperin, Nelson and Siggia(Ambegaokar *et al.*, 1980) and derived from the point of view of vortex duality(Fisher and Lee, 1989). To see how this works, let us recall that the basic hydrodynamical variables of the superfluid film are the density $\rho(x)$ and the velocity $v_i(x)$ fields, $(i = 1, 2)$, satisfying the equation of continuity

$$\partial_t \rho + \partial_i(\bar{\rho} v_i) = 0 \tag{2}$$

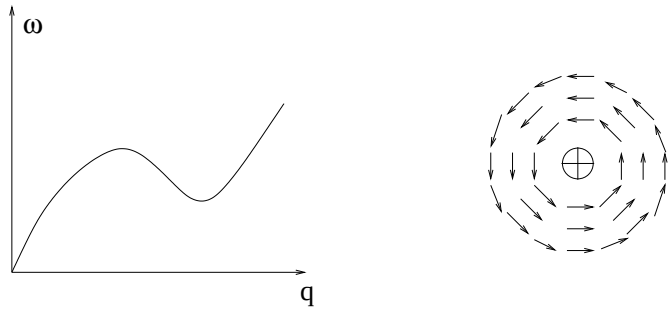


FIG. 1 Collective excitations of a neutral 2D superfluid film are the sound waves and the vortices. In the long wave length limit, the sound wave maps onto the Maxwell fields, while vortices map onto electric charges.

where $\bar{\rho}$ is the average density of the fluid. Now let us recall that in $2+1$ dimensions, the electric field E_i has two components while the magnetic field B has only one component, which can therefore be identified as a scalar. Faraday's law of induction is given by the Maxwell equation:

$$\frac{1}{c}\partial_t B + \epsilon_{ij}\partial_i E_j = 0 \tag{3}$$

where ϵ_{ij} is the antisymmetric tensor in two dimensions. Therefore, if we make the following identification,

$$B \quad \Leftrightarrow \quad -c\frac{\rho}{\bar{\rho}} \qquad\qquad E_i \quad \Leftrightarrow \quad \epsilon_{ij}v_j \tag{4}$$

we see that the equation of continuity of the superfluid film agrees exactly with Faraday's law as expressed in the Maxwell's equation (3). Next we examine the fluid velocity in the presence of a vortex, with unit vorticity, located at the position x_n. The superfluid state has a well defined $U(1)$ order parameter, and the velocity field can be expressed in terms of the phase, ϕ, of the $U(1)$ order parameter:

$$v_i = \frac{\hbar}{m}\partial_i\phi \tag{5}$$

Because of the single valuedness of the quantum mechanical wave function, $e^{i\phi}$ must be single valued. Therefore, the superflow around a vortex is *quantized*:

$$\int \vec{v}\cdot d\vec{l} = 2\pi\frac{\hbar}{m}q \tag{6}$$

where q is an integer. For elementary vortices, $q = \pm 1$. The differential form of this integral equation is

$$\epsilon_{ij}\partial_i v_j = 2\pi\rho_v(x) \tag{7}$$

where $\rho_v(x) = \sum_n q_n\delta(x - x_n)$ is the density of the vortices and $q_n = \pm 1$ is the vorticity. If we identify the vortex density with the electric charge density in Maxwell's equations, we see that equation (7) is nothing but Gauss's law in $2+1$ dimensions:

$$\partial_i E_i = 2\pi\rho_v(x) \tag{8}$$

Next let us investigate the dynamics of the superfluid velocity v_i, through the Josephson equations of superfluidity. The first Josephson equation relates the superfluid velocity to the gradient of the superfluid phase, ϕ, as expressed in (5). The second Josephson equation relates the time derivative of the phase to the chemical potential $\hbar\partial_t\phi = -\mu$. Combining the two Josephson equations, we obtain,

$$\partial_t v_i = \frac{\hbar}{m}\partial_t\partial_i\phi = -\frac{1}{m}\partial_i\mu = -\frac{\kappa}{m\bar{\rho}}\partial_i\rho \tag{9}$$

where we use the compressibility $\kappa = \bar{\rho}\frac{\partial\mu}{\partial\rho}$ to express the chemical potential μ in terms of the density ρ. This equation agrees exactly with the source-free Maxwell equation

$$c\epsilon_{ij}\partial_j B = \partial_t E_i \tag{10}$$

provided one identifies the speed of light as $c^2 = \kappa/m$. This equation needs to be modified in the presence of the vortex flow J_i^v, which unwinds the $U(1)$ phase by 2π each time a vortex passes. The vortex current satisfies the equation of continuity

$$\partial_t \rho_v + \partial_i J_i^v = 0 \tag{11}$$

Therefore, the source free Maxwell equation (10) acquires a additional term, in order to be compatible with both (11) and (8):

$$c\epsilon_{ij}\partial_j B = \partial_t E_i + 2\pi J_i^v \tag{12}$$

This is nothing but Ampere's law, supplemented by Maxwell's displacement current.

This proves the complete equivalence between the superfluid equations and Maxwell's equations in $2+1$ dimensions. Interestingly enough, we seem to have completed a rather curious loop. Starting from the relativistic Standard Model of the quarks and leptons, one arrives at an effective non-relativistic model of the helium atoms (1). However, as one reduces the energy scale further, the effective low energy degrees of freedom become the sound modes and the vortices, which are described by the field theory of $2+1$ dimensional quantum electrodynamics, very similar to the model we started from in the first place! A "civilization" living inside the helium film would first discover the Maxwell's equations, and then, after much harder work, they would establish equation (1) as their "theory of everything".

Superfluid ^4He films are relatively simple because the ^4He atom is a boson. The superfluidity of ^3He is much more complex, with many competing superfluid phases. In fact, Volovik(Volovik, 2001) has pointed out that many phenomena of the superfluid phase of ^3He share striking similarities with the Standard Model of elementary particles. These similarities inspired him to pioneer a program to address cosmological questions by condensed matter analogs.

B. Dirac fermions of d wave superconductors

Having considered the low energy properties of a superfluid, let us now consider the low energy excitations of a superconductor, with d wave pairing symmetry. In this case, there are low energy fermionic excitations besides the bosonic excitations. This system is realized in the high T_c superconductors. The microscopic Hamiltonian is the two dimensional (2D) Hubbard model, or the $t-J$ model, expressible as

$$H = -t\sum_{\langle ij\rangle,\sigma} c_{i\sigma}^\dagger c_{j\sigma} + J\sum_{\langle ij\rangle} \vec{S}_i \cdot \vec{S}_j \tag{13}$$

where $c_{i\sigma}^\dagger$ is the electron creation operator on site i with spin σ, $\vec{S}_i$ is the electron spin operator and $\langle ij\rangle$ denotes the nearest neighbor bond on a square lattice. Double occupancy of a single lattice site is forbidden.

This model is valid at the energy scale of $150meV$, which is the typical energy scale of the antiferromagnetic exchange J. When the filling factor x lies between 10% and 20%, the ground state of this model is believed to be a d wave superconductor. There is indeed overwhelming experimental evidence that the pairing symmetry of the high T_c superconductor is d wave like. Remarkably, the elementary excitations in this case can be described by the $2+1$ dimensional Dirac Hamiltonian. In contrast to the $t-J$ model, which is valid at the energy scale of $100meV$, the effective Dirac Hamiltonian for the d wave quasi-particles is valid at much lower energy, typically of the order of $30meV$, which is the maximal gap. While the connection between the $t-J$ model and d wave superconductivity still needs to be firmed established, the connection between the d wave BCS quasi-particle Hamiltonian and the Dirac equation is well-known in the condensed matter community(Balents *et al.*, 1998; Franz *et al.*, 2002; Simon and Lee, 1997; Volovik, 1993). Here we follow a pedagogical presentation by Balents, Fisher and Nayak(Balents *et al.*, 1998).

The BCS mean field Hamiltonian for a d wave superconductor is given by

$$H = \sum_{k\alpha} \epsilon_k c_{k\alpha}^\dagger c_{k\alpha} + \sum_k [\Delta_k c_{k\uparrow}^\dagger c_{-k\downarrow}^\dagger + \Delta_k^* c_{-k\downarrow} c_{k\uparrow}]. \tag{14}$$

where ϵ_k is the quasi-particle dispersion relation, and Δ_k is the d wave pairing gap, given by

$$\epsilon_k = -2t(cosk_x + cosk_y) \quad , \quad \Delta_k = \Delta_0(cosk_x - cosk_y) \tag{15}$$

One can introduce a four component spinor

$$\Upsilon_{a\alpha}(\vec{k}) = \begin{bmatrix} \Upsilon_{11} \\ \Upsilon_{21} \\ \Upsilon_{12} \\ \Upsilon_{22} \end{bmatrix} = \begin{bmatrix} c_{k\uparrow} \\ c_{-k\downarrow}^\dagger \\ c_{k\downarrow} \\ -c_{-k\uparrow}^\dagger \end{bmatrix}. \tag{16}$$

which doubles the number of degrees of freedom. This can be compensated by summing over only half of the Brillouin zone, say $k_y > 0$. In terms of these variables, the BCS Hamiltonian becomes

$$H = \sum_{k,k_y>0} \Upsilon^\dagger_{a\alpha}(\vec{k})[\tau^z \epsilon_k + \tau^+ \Delta_k + \tau^- \Delta^*_k]_{ab} \Upsilon_{b\alpha}(\vec{k}), \tag{17}$$

where $\vec{\tau}_{ab}$ are the standard Pauli matrices acting in the particle/hole subspace.

The d-wave nodes are approximately located near the special wave vectors $\vec{K}_1 = (\pi/2, \pi/2)$, $\vec{K}_2 = (-\pi/2, \pi/2)$, $\vec{K}_3 = -\vec{K}_1$ and $\vec{K}_4 = -\vec{K}_2$. In order to obtain a long wave length and low energy description, we can expand around the nodal points $\vec{K}_1$ and $\vec{K}_2$, which satisfy the $k_y > 0$ constraint. The nodal points $\vec{K}_3$ and $\vec{K}_4$ are automatically taken into account in the Υ spinor.

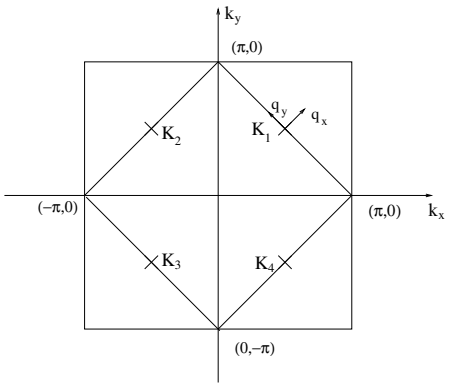


FIG. 2 A 2D d wave superconductor has four nodes, indicated by K_1, K_2, K_3 and K_4. Around these nodal points, BCS quasi-particles obey the massless Dirac equation.

Introducing the rotated coordinates q_x and q_y, as indicated in fig. (2), and the effective spinors

$$\Psi_{1a\alpha}(\vec{q}) = \Upsilon_{a\alpha}(\vec{K}_1 + \vec{q}) \quad , \quad \Psi_{2a\alpha}(\vec{q}) = \Upsilon_{a\alpha}(\vec{K}_2 + \vec{q}) \tag{18}$$

we obtain

$$H = \sum_{q \in K_1} \Psi^\dagger_{1a\alpha}(\vec{q})[\tau^z \epsilon_{K_1+q} + \tau^+ \Delta_{K_1+q} + \tau^- \Delta^*_{K_1+q}]_{ab} \Psi_{1b\alpha}(\vec{q}) + (1 \leftrightarrow 2) \tag{19}$$

Here $q \in K_1$ denotes a momentum sum near the vector K_1. Expansion near K_1 gives

$$\epsilon_{K_1+q} \approx v_F q_x \quad , \quad \Delta_{K_1+q} \approx \Delta q_y \tag{20}$$

A similar expansion applies for K_2. Going to the continuum limit, we obtain the Hamiltonian density

$$\begin{aligned} \mathcal{H} \;=\; & \Psi^\dagger_{1a\alpha}[v_F \tau^z i\partial_x + (\tilde{\Delta}\tau^+ + \tilde{\Delta}^* \tau^-) i\partial_y]_{ab} \Psi_{1b\alpha} \\ & + (1 \leftrightarrow 2; x \leftrightarrow y), \end{aligned} \tag{21}$$

which is exactly the Dirac Hamiltonian density in $2+1$ dimensions. Once again, we see the emergent relativistic behavior of a quantum many-body system. We start from a non-relativistic interacting fermion problem at higher energies, but recover a relativistic Dirac equation at low energies.

C. Emergence of a topological quantum field theory

When Einstein first wrote down his field equation of general relativity, he said that the left-hand side of the equation that had to do specifically with geometry and gravity was beautiful - it was as if made of marble. But the right-hand side of the equation that had to do with matter and how matter produces gravity was ugly - it was as if made of wood. Taking Einstein's aesthetic point of view one step further, one is tempted to construct a fundamental theory by starting with the description of the topology, or a topological field theory without matter and without even geometry from the start. Having demonstrated that the relativistic Maxwell equation and Dirac equation can emerge in the low

energy sector of a quantum many-body problem, I now give an example demonstrating how a topological quantum field theory, namely the Chern-Simons (CS) theory, can emerge from the matter degrees of freedom in the low energy sector of the QHE. The CS topological quantum field theory was derived microscopically by Zhang, Hansson and Kivelson(Zhang *et al.*, 1989), and reviewed extensively in ref. (Zhang, 1992).

The basic Hamiltonian of QHE is simply that of a two-dimensional electron gas in a perpendicular magnetic field.

$$H = \frac{1}{2m}\sum_n \left[\vec{p}_n - \frac{e}{c}\vec{A}(x_n)\right]^2 + \sum_n eA_0(x_n) + \sum_{n<n'} V(x_n - x_{n'}) \tag{22}$$

where $\vec{A}$ is the vector potential of the external magnetic field, which in the symmetric gauge can be expressed as

$$A_i = \frac{1}{2}B\epsilon_{ij}x_j \tag{23}$$

A_0 is the scalar potential of the external electric field, $E_i = -\partial_i A_0$, and $V(x)$ is the interaction between the electrons. For high magnetic fields, the electron spins are polarized along the same direction. Since the spin wave function is totally symmetric, the Hamiltonian (22) operates on orbital wave functions which are totally antisymmetric. This model is valid at the Coulombic energy scale of several eV's and has no particular symmetry or topological properties. Since the external magnetic field breaks time reversal symmetry, an invariant tensor ϵ_{ij} can be introduced into the response function, and in particular, one can have a current J_i, which flows transverse to the applied electric field E_j, given by

$$J_i = \rho_H^{-1}\epsilon_{ij}E_j \tag{24}$$

where ρ_H is defined as the Hall resistance. Since the electric field is perpendicular to the induced current, it does no work on the electrons, and the current flow is dissipationless. The 2D electron density n in a magnetic field B is best measured in terms of a dimensionless quantity called the filling factor $\nu = n/n_B$, where $n_B = B/\phi_0 = eB/hc$ is the magnetic flux density. QHE is the remarkable fact that the coefficient of the Hall response is quantized, given by

$$\rho_H = \nu^{-1}\frac{h}{e^2} \tag{25}$$

when the filling fraction is near a rational number $\nu = p/q$ with odd denominator q. QHE at fractional values of ν is referred to as the fractional QHE (FQHE).

FQHE is described by Laughlin's celebrated wave function. There is also an alternative way to understand this profound effect by the Chern-Simons-Landau-Ginzburg (CSLG) effective field theory(Zhang, 1992). The idea is to perform a singular gauge transformation on (22), and turn electrons into bosons. This is only possible in $2+1$ dimensions. Consider another Hamiltonian

$$H' = \frac{1}{2m}\sum_n \left[\vec{p}_n - \frac{e}{c}\vec{A}(x_n) - \frac{e}{c}\,\vec{a}(x_n)\right]^2 + \sum_n eA_0(x_n) + \sum_{n<n'} V(x_n - x_{n'}) \tag{26}$$

Every symbol in H' has the same meaning as in H, except the new vector potential $\vec{a}$, which describes a *gauge interaction* among the particles and is given by

$$\vec{a}(x_n) = \frac{\phi_0}{2\pi}\frac{\theta}{\pi}\sum_{n' \neq n} \vec{\nabla}\,\alpha_{nn'} \tag{27}$$

where $\phi_0 = hc/e$ is the unit of flux quantum and $\alpha_{nn'}$ is the angle sustained by the vector connecting particles n and n' with an arbitrary vector specifying a reference direction, say the $\hat{x}$ axis. The crucial difference here is while H operates on a fully antisymmetric fermionic wave function, H' operates on a fully symmetric bosonic wave function. One can prove an exact theorem which states that these two quantum eigenvalue problems are equivalent to each other when $\theta/\pi = (2k+1)$ is an odd integer. In this case, each electron is attached to an odd number of fictitious quanta of gauge flux (cause by a), so that their exchange statistics in $2+1$ dimensions becomes bosonic. These bosons, called composite bosons(Girvin and Macdonald, 1987; Read, 1989; Zhang *et al.*, 1989), see two different types of gauge fields: the external magnetic field A, and an internal statistical gauge field a. The average of the internal statistical gauge field depends on the density of the electrons. When the external magnetic field is such that the filling fraction $\nu = n_B/n = 1/2k+1$ is the inverse of an odd integer, we can always choose $\theta = (2k+1)\pi$ so that the net field seen by the composite bosons cancels each other on the average.

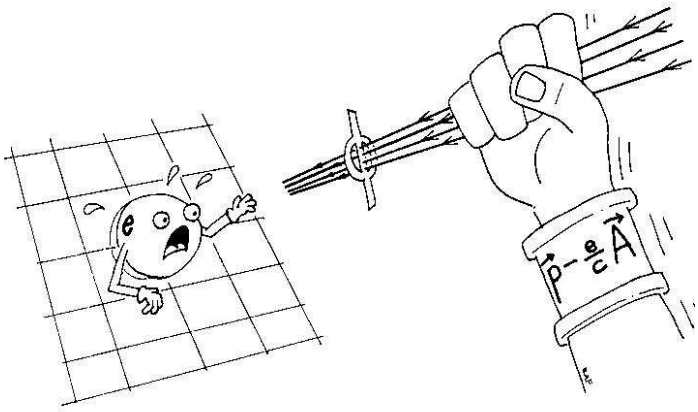


FIG. 3 An electron just before the flux transmutation operation. (taken from the PhD thesis of D. Arovas, illustrated by Dr. Roger Freedman).

The statistical transmutation from electrons to composite bosons can be naturally implemented in quantum field theory through the Cherm-Simons term. The Chern-Simons Lagrangian is given by

$$\mathcal{L} = \frac{1}{2}\left(\frac{\pi}{\theta}\right)\frac{1}{\phi_0}\,\varepsilon^{\mu v\rho}\,a_\mu\,\partial_v\,a_\rho - a_\mu\,j^\mu \tag{28}$$

here j^μ is the current of the composite boson, and $\mu = 0, 1, 2$ is the space-time index in $2+1$ dimensions. The equation of motion for the a_0 field is

$$\epsilon^{ij}\,\partial_i\,a_j(x) = \phi_0\,\frac{\theta}{\pi}\,\rho(x) \tag{29}$$

whose solution for $\rho(x) = \sum_n \delta(x - x_n)$ exactly gives the statistical gauge field in (27).

Now we can present the key argument of the CSLG theory(Zhang, 1992) of QHE. Even though course the statistical transformation can be performed in any $2+1$ dimensional systems, this does not mean that the low energy limit of any $2+1$ dimensional system is given by a CS theory, since the partition function also involves the integration over the matter fields j^μ in the second term of (28). The key observation is that at the special filling factors of $\nu = 1/2k+1$, the combined external and statistical magnetic field seen by the composite boson vanishes, therefore, composite bosons naturally condense into a superfluid state. This is the "magic" of the magic filling factors $\nu = 1/2k+1$. We already showed in section (II.A) that the effective field theory of a $2+1$ dimensional bosonic condensate is the $2+1$ dimensional Maxwell theory. Therefore, the integration over the matter fields in (28) gives the Maxwell Lagrangian, $f_{\mu\nu}^2$. In $2+1$ dimensions, the CS term contains one fewer derivative compared with the Maxwell term, it therefore dominates in the long-wave length and low-energy limit. Therefore, the effective Hamiltonian of FQHE is just the topological CS theory, without the matter current term in (28).

Matter degrees of freedom in the starting Hamiltonian (22) are magically turned into topological degrees of freedom of the CS field theory. Alchemy works! Wood is turned into marble! Many people argued that a quantum theory of gravity should be formulated independent of the background metric. The emergent CSLG theory starts from matter degrees of freedom in a background setting, but the resulting effective field theory is independent of the background metric. This demonstrates that in principle, background independent theory can indeed be constructed from non-relativistic quantum many-body systems. In fact, the CSLG theory also leads to a beautiful duality symmetry based on the discrete $SL(2, Z)$ group, very similar to the duality symmetry in the 4D Seiberg-Witten theory. This duality symmetry is again emergent, and it predicts the global phase diagram of the QH Hall system. The phase diagram has a beautiful fractal structure, with one phase nested inside each other, iterated *ad infinitum*(Kivelson *et al.*, 1992).

III. THE FOUR DIMENSIONAL QUANTUM HALL EFFECT

In the previous sections we saw that the collective behavior of quantum many-body systems often gives rise to novel emergent phenomena in the low energy sector, which are described in terms of relativistic or topological quantum field theories. Therefore, one can't help but wonder if the Standard Model could also work this way. The problem

is that the well-understood examples of emergent relativistic behaviors in quantum many-body systems work only for lower dimensions, and these models do not have sufficient richness yet. In order for the Stanford Model to appear as emergent behavior, we are led to study higher-dimensional quantum many-body systems, specially the higher-dimensional generalizations of QHE.

A. The model

Of all the novel quantum many-body systems, QHE plays a very special role: it is the only well understood condensed matter system whose low energy limit is a topological quantum field theory. Unlike most other emergent phenomena, like superconductivity and magnetism, QHE works only in two spatial dimensions. There are various ways to see this. First of all, the Hall current is non-dissipative. For the electric field to do no work on the current, the current must flow in a direction perpendicular to the direction of the electric field. In two spatial dimensions, given the direction of the electric field, there is an unique transverse direction for the Hall current, given by equation (24). Since the current and the electric field both carry spatial vector indices, the response must therefore be a rank-two tensor. But there are no natural rank-two antisymmetric tensors in higher dimensions! Secondly, both the single-particle wave function and Laughlin's many-body wave function make extensive use of complex coordinates of particles, which can only be done in two spatial dimensions. This suggests that the higher-dimensional generalization of QHE would necessarily involve a higher-dimensional generalization of complex numbers and analytic functions. In fact, both of these considerations lead to the same higher-dimensional structure, as we shall explain below.

In higher dimensions, given a direction of the electric field, there is no unique transverse direction for the Hall current to flow. However, this statement holds only if we consider the $U(1)$ charge current. If the underlying particles – and the associated currents – carry a non-abelian, *e.g.* $SU(2)$ quantum number, an unique prescription for the current can be given in *four* dimensions. In four dimensions, given a fixed direction of the electric field, say along the x_4 direction, there are three transverse directions. If the current carries a $SU(2)$ isospin label, it also has three internal isospin directions. In this case, the current can flow exactly along the direction in which the isospin is pointing. In this prescription, no preferential direction in space or isospin is picked. The system is invariant under a *combined* rotation of space and isospin. To be more precise, the mathematical generalization of (24) in four dimensions is

$$J^i_\mu = \sigma \eta^i_{\mu\nu} E_\nu \tag{30}$$

Here σ is the generalized Hall conductivity, $\eta^i_{\mu\nu}$ is the t' Hooft tensor, explicitly given by $\eta^i_{\mu\nu} = \epsilon_{i\mu\nu 4} + \delta_{i\mu}\delta_{4\nu} - \delta_{i\nu}\delta_{4\mu}$ and J^i_μ is the isospin current and E_ν is the electric field. Here $\mu, \nu = 1, 2, 3, 4$ label the spatial directions and $i = 1, 2, 3$ label the isospin directions. From (30), we see easily that if E_ν points along the x_4 direction, the current flows along the $x_{1,2,3}$ directions, explicitly determined by the direction of the isospin. Therefore, the t' Hooft tensor is exactly the rank-two antisymmetric tensor we were looking for! The occurrence of the t' Hooft tensor suggests that this problem must have something to do with the $SU(2)$ instanton(Belavin *et al.*, 1975), where the t' Hooft tensor was first introduced. It is not only an *invariant* tensor under combined spatial and isospin rotations, it also satisfies a self-duality condition:

$$\eta^i_{\mu\nu} = \epsilon_{\mu\nu\rho\lambda}\eta^i_{\rho\lambda} \tag{31}$$

Self-duality and anti-self-duality are the hallmarks of the $SU(2)$ Yang-Mills instanton.

Now let us motivate the problem from the point of view of generalizing complex numbers. The natural generalizations of complex numbers are quaternionic numbers, first discovered by Hamilton. A quaternionic number is expressed as $q = q_0 + q_1 i + q_2 j + q_3 k$, where i, j, k are the three imaginary units. This again suggests that the most natural generalization of QHE is from 2D to 4D, where quaternionic numbers can be interpreted as the coordinates of particles in four dimensions. Unlike complex numbers, quaternionic numbers do not commute with each other. In fact, the three imaginary units of quaternionic numbers can be identified with the three generators of the $SU(2)$ group. This suggests that the underlying quantum mechanics problem should involve a non-abelian $SU(2)$ gauge field.

Our last motivation to generalize QHE comes from its geometric structure. As pointed out by Haldane(Haldane, 1983), a nice way to study QHE is by mapping it to the surface of a 2D sphere S^2, with a Dirac mangnetic monopole at its center. (see Fig. 4). The Dirac quantization condition implies that the product of the electric charge, e, and the magnetic charge, g, is quantized, *i.e.* $eg = S$, where $2S$ is a integer. The number $2S + 1$ is the degeneracy of the lowest Landau level. The reason for the existence of a magnetic monopole over S^2 is a coincidence between algebra and geometry. In order for the monopole potential to be topologically non-trivial, the gauge potentials extended from the north pole and the south pole have to match non-trivially at the equator. Since the equator, S^1, and the gauge group, $U(1)$, are isomorphic to each other, a non-trivial winding number exists. Therefore, one may ask whether there are other higher-dimensional spheres for which a similar monopole structure can be defined. This naturally

leads to the requirement that the equator of a higher-dimensional sphere to be isomorphic to a mathematical group. This coincidence occurs only for the four sphere, S^4, whose equator, S^3, is isomorphic to the group $SU(2)$. This coincidence between algebra and geometry leads to the first two Hopf maps, $S^3 \to S^2$ and $S^7 \to S^4$.

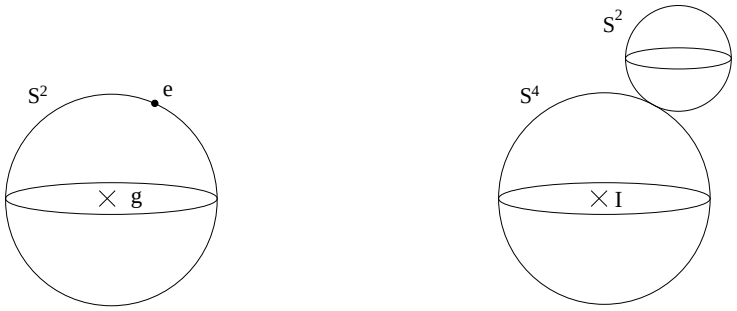


FIG. 4 The 2D QHE consists of electrons e on the surface of a 2D sphere S^2, with a $U(1)$ magnetic monopole g at its center. Similarly, the 4D QHE can be defined on the surface of a 4D sphere S^4, with a $SU(2)$ monopole I at its center. In the large I limit, the $SU(2)$ isospin degree of freedom is S^2.

Therefore, all three considerations – the physical motivation of the transverse current, the mathematical motivation of generalizing complex numbers to quaternionic numbers and the geometric consideration of non-trivial monopole structures – lead to the same conclusion: A non-trivial QHE liquid can be defined in four spatial dimensions (4D) with a $SU(2)$ non-abelian gauge group. Recently, Hu and I (ZH) indeed succeeded in constructing such a model for the 4D QHE(Zhang and Hu, 2001). The microscopic Hamiltonian describes a collection of N fermionic particles moving on S^4, interacting with a $SU(2)$ background isospin gauge potential A_a. It is explicitly defined by

$$H = \frac{\hbar^2}{2MR^2} \sum_{a<b} \Lambda_{ab}^2 \tag{32}$$

where M is the mass of the fermionic particle, R is the radius of S^4, and $\Lambda_{ab} = -i(x_a D_b - x_b D_a)$ is the gauge covariant angular momentum operator. Here x_a is the coordinate of the fermionic particle and $D_a = \partial_a + A_a$ is the gauge invariant momentum operator. The gauge potential A_a $(a = 1, 2, 3, 4, 5)$ is given by

$$A_\mu = \frac{-i}{1 + x_5} \eta^i_{\mu\nu} x_\nu I_i \quad , \quad A_5 = 0 \tag{33}$$

where I_i are the generators of the $SU(2)$ gauge group. An important parameter in this problem is I, the isospin quantum number carried by the fermionic particle. The eigenstates and the eigenvalues of this Hamiltonian can be solved completely, and the spectrum shares many properties with the Landau levels in the 2D QHE problem. In particular, when I becomes large, the ground state of this problem is massively degenerate, with the degenracy scaling like $D \sim I^3$. In order to keep the energy levels finite in the thermodynamical limit, one is required to take the limit $I \to \infty$ as $R \to \infty$, such that

$$R^2/2I = l^2 \tag{34}$$

is finite. l, called the magnetic length, defines the fundamental length scale in this problem. It gives a natural ultraviolet cut-off in this theory, without breaking any rotational symmetries of the underlying Hamiltonian.

While the 4D QH liquid can be elegantly defined on S^4, with the full isometry group as the symmetry of the Hamiltonian, it can also be defined on R^4, with more restricted symmetries. This construction has recently been given by Elvang and Polchinski(Elvang and Polchinski, 2002).

B. Properties of the model

The 2D QH liquid has many interesting properties including incompressibility of the quantum liquid, fractional charge and statistics of elementary excitations, a topological field theory description of the low energy physics, a realization of non-commutative geometry and relativistic chiral excitations at the edge of the QH droplet. Most of these properties also carry over to the QH liquid constructed by ZH. When one completely fills the massively degenerate lowest energy ground states with fermionic particles, with filling factor $\nu \equiv N/D = 1$, one obtains an incompressible

quantum liquid, with a finite excitation gap towards all excited states. FQH states can also be constructed for filling fractions $\nu = 1/k^3$, where k is a odd integer. Explicit microscopic wave functions, similar to Laughlin's wave function for the 2D QHE, can be constructed for these incompressible states. The elementary excitations of the FQH states also carry fractional charge $1/k^3$, providing the first direct generalization of fractional charge in a higher-dimensional quantum many-body system.

As discussed in section (II.C), the low energy physics of the 2D QHE can be described by a topological quantum field theory, the CSLG theory. A natural question is whether the QH liquid constructed by ZH can be described by a topological quantum field theory as well. This construction has indeed been accomplished recently, by Bernevig, Chern, Hu, Toumbas and myself(Bernevig *et al.*, 2002). As explained earlier, while the underlying orbital space for our QH liquid is four dimensional, the fermionic particles also carry a large internal isospin degree of freedom I. Since I scales like R^2, the internal space is 2D, which makes the total configuration space a six-dimensional (6D) manifold. Therefore, our QH liquid can either be viewed as a 4D QH liquid with a large internal $SU(2)$ isospin degrees of freedom, or equivalently, as a 6D QH liquid without any internal degree of freedom. The 6D manifold is CP_3, the complex projective space with three complex (and therefore six real) dimensions. This manifold is locally isomorphic to $S^4 \times S^2$. The deep connection between the four sphere S^4 and the complex manifold CP_3 was first introduced to physics through the twistor program of Penrose(Penrose and MacCallum, 1972) and has been exploited extensively in the mathematical literature. Sparling(Sparling, 2002) has recently pointed out the close connection between the twistor theory and the 4D QHE. Our recent work shows that the low energy effective field theory of our QH liquid is given by an abelian CS theory in $6+1$ dimensions

$$S = \nu \int dt d^6x A \wedge dA \wedge dA \wedge dA \tag{35}$$

where A is an abelian $U(1)$ gauge field over the total configuration space CP_3, and ν is the filling factor. This theory can also be dimensionally reduced to a $SU(2)$ non-abelian CS theory in $4+1$ dimensions, given by

$$S = \frac{4\pi\nu}{3} \int dt d^4x Tr \left(\mathbf{A} \wedge d\mathbf{A} \wedge d\mathbf{A} - \frac{3i}{2} \mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A} \wedge d\mathbf{A} - \frac{3}{5} \mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A} \right) \tag{36}$$

where $\mathbf{A}$ is a $SU(2)$ matrix-valued gauge field over the orbital space S^4. The precise equivalence of these two models parallels the two equivalent views of our QH liquid mentioned earlier.

An interesting property which arises from this field theory is the concept of duality. As discussed in section (II.C), there is a natural particle-flux duality in the 2D QHE problem: An electron can be represented as a boson with an odd number of flux quanta attached to it. In the new QH liquid, there are other extended objects, namely 2-branes and 4-branes besides the basic fermionic particle, which can be viewed as a 0-brane. Each one of these extended objects is dual to a generalized flux, according to the following table:

$$\begin{aligned} \text{Particle} &\iff \text{6-flux} \\ \text{Membrane} &\iff \text{4-flux} \\ \text{4-brane} &\iff \text{2-flux} \end{aligned}$$

In the 2D QH problem, the Laughlin quasi-particles obey fractional statistics in 2+1 dimensions. It is natural to ask how fractional statistics generalize in our QH liquid. It turns out that the concept of fractional statistics of point particles can not be generalized to higher dimensions, but fractional statistics for extended objects exist in higher dimensions(Tze and Nam, 1989; Wu and Zee, 1988). In our case, 2-branes have non-trivial statistical interactions which generalizes statistical interactions of Laughlin quasi-particles.

Extended objects like D-branes have been studied extensively in string theory, however, a full quantum theory describing their interactions still needs to be developed. The advantage of our approach is that the underlying microscopic quantum physics is completely specified. Since the extended topological objects emerge naturally from the underlying microscopic physics, there is hope that a full quantum theory can be developed in this case.

The study of 4DQHE is partially motivated by the possibility of emergent relativistic behavior in $3+1$ dimensions. There are several ways to see the connection. First of all, the eigenstates and the eigenfunctions of the Hamiltonian (32) have a natural interpretation in terms of the 4D Euclidean quantum field theory. If we consider a Euclidean quantum field theory as obtained from a Wick rotation of a $3+1$ dimensional compactified Minkowskian quantum field theory, one is naturally lead to consider the eigenvalues and the eigenfunctions of the Euclidean Dirac, Maxwell and Einstein operators on S^4. It turns out that the these eigenvalues and eigenfunctions coincide exactly with the eigenvalues and eigenfunctions of the 4DQHE Hamiltonian (32), where the spins of the relativistic particles are identified with the isospin quantum number, I. The eigenvalue problems of the Dirac, Maxwell and Einstein operators can be directly identified with the Hamiltonian eigenvalue problems for $I = 1/2, 1$ and 2. We mentioned earlier that the underlying fermionic particles constituting our QH liquid have high isospin quantum numbers. However, collective excitations

of this QH liquid, which are formed as composite particles, can have low isospin quantum numbers. It is therefore tempting to identify the collective excitations of the QH liquid with the relativistic particles we are familiar with. However, this equivalence is only established in Euclidean space. In order to consider the relationship to Minkowski space, we are naturally lead to the excitations at the boundary, or the edge of our QH liquid.

Let us first review the collective excitations at the edge of a 2D QH liquid. The 2D QH liquid can be confined by a one-body confining potential V. A density excitation is created by removing a particle from the QH liquid and placing it outside of the QH liquid. This way, we have created a particle-hole excitation. If the particle-hole pair moves along a direction parallel to the edge, with a center of mass momentum q_x, the Lorentz force due to the magnetic field acts oppositely on the particle-hole pair, and tries to stretch the pair in the direction perpendicular to the edge. This Lorentz force is balanced by the electrostatic attraction due to the force of the confining potential. Therefore, a unique dipole moment, or a finite separation y of the particle hole pair, is obtained in terms of q_x:

$$y = l^2 q_x \tag{37}$$

On the other hand, the energy of the dipole pair is simply given by $E = V'y$. Here V' is the derivative of the potential evaluated at the edge. Therefore, we obtain a relativistic dispersion relation for the dipole pair

$$E = V'y = l^2 V' q_x \tag{38}$$

with the speed of light given by $c = l^2 V'$. Since the cross product of the gradient of the potential and the magnetic field selects a unique direction along the edge, the excitation is also chiral. In this problem, it can also be shown that not only the dispersion, but also the full interaction is relativistic in the low energy limit. Therefore, the physics at the edge of a 2D QH liquid provides another example of emergent relativistic behavior(Stone, 1990; Wen, 1990).

The physics of the edge excitations of a 2D QH liquid *partially* carries over to our 4D QH liquid(Elvang and Polchinski, 2002; Hu and Zhang, 2002; Zhang and Hu, 2001). Here we can also introduce a confining potential, say around the north pole of S^4, and construct a droplet of the QH fluid. Since our QH liquid is incompressible, the only low energy excitations are the volume preserving shape distortions at the surface. These surface waves can be formed from the particle-hole excitations similar to the ones we described for the 2D QH liquid. A natural speed of light can be introduced, and is given by $c = l^2 V'$. Since our underlying particles carry a large isospin, I, the bosonic composite particle-hole excitations carry all isospins, ranging from 0 to $2I$. The underlying fermionic particles have a strong coupling between their orbital and isospin degrees of freedom. This coupling translates into a relativistic spin-orbital coupling of the bosonic collective excitations. Therefore, excitations with $I = 0, 1, 2$ obey the *free* relativistic Klein-Gordon, Maxwell and Einstein equations. This is an encouraging sign that one might be able to construct an emergent relativistic quantum field theory from the boundary excitations of our 4D QH liquid.

However, there are also many complications which are not yet fully understood in our approach. The most fundamental problem is that particles of our 4D QH liquid carry a large internal isospin, which makes the problem effectively a 6D one. This is the basic reason for the proliferation of higher-spin particles in our theory, an "embarrassment of riches". In addition, there is an incoherent fermionic continuum besides the bosonic collective modes. All these problems can only be addressed when one studies the effects of the interaction carefully. In fact, single particle states in the lowest-Landau-level (LLL) have the full symmetry of $SU(4)$, which is the isometry group of the six dimensional CP_3 manifold. In order to make the problem truly 4D, one needs to introduce interactions which breaks the $SU(4)$ symmetry to a $SO(5)$ symmetry, the isometry group of S^4. This is indeed possible. $SO(5)$ is isomorphic to the group $Sp(4)$. $Sp(4)$ differs from $SU(4)$ by an additional reality condition, implemented through a charge conjugation matrix R. Therefore, any interactions which involve this R matrix would break the symmetry from $SU(4)$ to $SO(5)$, and the geometry of S^4 would emerge naturally. In the strong coupling limit, low energy excitations are not particles but membranes. This reduces the entropy at the edge from $R^3 \times R^2$ to R^3, and is the first step towards solving the problem of "embarrassment of riches".

C. Space, time and the quantum

The 2D QH problem gives a precise mathematical realization of the concept of non-commutative geometry(Douglas and Nekrasov, 2001). In the limit of high magnetic field, we can take the limit of $m \to 0$, so that all higher Landau levels are projected out of the spectrum. In this limit, the equation of motion for a charged particle in an uniform magnetic field B and a scalar potential $V(x, y)$ is given by

$$\dot{x} = l^2 \frac{\partial V}{\partial y} \quad , \quad \dot{y} = -l^2 \frac{\partial V}{\partial x} \tag{39}$$

We notice that the equations for x and y look exactly like the Hamilton equations of motion for p and q. Therefore, this equation of motion can be derived as quantum Heisenberg equations of motion if we postulate a similar commutation relation:

$$[x, y] = il^2 \tag{40}$$

Therefore, the 2D QHE provides a physical realization of the mathematical concept of non-commutative geometry, in which different spatial components do not commute. Early in the development of quantum field theory, this feature has been suggested as a way to cut off the ultraviolet divergences of quantum field theory. In quantum mechanics, the non-commutativity of q and p leads to the Heisenberg uncertainty principle and resolves the classical catastrophe of an electron falling towards the atomic nucleus. Similarly, non-commutativity of space and time could cut off the ultraviolet space-time fluctuations in quantum gravity(Douglas and Nekrasov, 2001). However, the problem is that equation (40) can not be easily generalized to higher dimensions, since one needs to pick some fixed pairs of non-commuting coordinates. Our QH liquid provides a physical realization of non-commutative geometry in four dimensions. The generalization of equation (40) becomes

$$[X_\mu, X_\nu] = 4il^2\eta^i_{\mu\nu}n_i \tag{41}$$

where X_μ's are the four spatial coordinates and n_i is the isospin coordinate of a particle. This structure of non-commutative geometry is invariant under a combined rotation of space and isospin and treats all these coordinates on equal footing. It is tempting to identify l in equation (41) as the Planck length, which provides the fundamental cutoff of the length scale according to the quantization rule (41). In our theory, however, we know what lies beyond the Planck length: the degrees of freedom are those associated with the higher Landau levels of the Hamiltonian (32).

At this point, it would be useful to discuss the possible implications of (41) on the quantum structure of space-time. In the 4D QH liquid, there is no concept of time. Since all eigenstates in the LLL are degenerate, there is no energy difference which can be used to measure time according to the quantum relation $\Delta t = \hbar/\Delta E$. However, at the boundary of the 4D QH liquid, an energy difference is introduced through the confining potential. The left hand side of equation (41) involves four coordinates. Three of them are the spatial coordinates parallel to the boundary. The fourth coordinate, perpendicular to the boundary, measures the energy difference, and therefore measures time. The commutator among these coordinates implies a quantization procedure. The right hand side of this equation involves the Planck length and the spin. Therefore, this simple equation seems to unify all the fundamental physical concepts: space, time, the quantum, the Planck length and spin in a simple and elegant fashion. It would be nice to use it as a basis to construct a fundamental physical theory.

D. Magic liquids, magic dimensions, magic convergence?

So far our philosophical point of view and our model seem to be drastically different from the approach typical of string theory. However, after the discovery of the new QH liquid, a surprising pattern starts to emerge. Soon after the construction of the new 4D QH liquid, Fabinger(Fabinger, 2002) found that it could be implemented as certain solutions in string theory. Moreover, close examination of this pattern reveals remarkable mathematical similarities not only between these two approaches, but also with other fundamental ideas in algebra, geometry, supersymmetry and the twistor program on quantum space time. The following table summarizes the connections.

Division Algebras:	Real Namubers	Complex Numbers	Quaternions	Octonions
Hopf maps:	$S^1 \to S^1$	$S^3 \to S^2$	$S^7 \to S^4$	$S^{15} \to S^8$
QH liquids:	Luttinger liquid?	Laughlin liquid	ZH liquid	?
Random matrix ensembles:	Orthogonal	Unitary	Symplectic	?
Fractional statistics:	?	particles	membranes	?
Geometric phase:	Z_2	$U(1)$	$SU(2)$	?
Non-commutative geometry:	?	$[X_i, X_j] = il^2\epsilon_{ij}$	$[X_\mu, X_\nu] = 4il^2\eta^i_{\mu\nu}n_i$	?
Twistor transformation:	$SO(2,1) = SL(2,\mathbf{R})$	$SO(3,1) = SL(2,\mathbf{C})$	$SO(5,1) = SL(2,\mathbf{H})$	$SO(9,1) = SL(2,\mathbf{O})$
$N = 1$ SUSY Yang-Mills:	$d = 2 + 1$	$d = 3 + 1$	$d = 5 + 1$	$d = 9 + 1$
Green-Schwarz Superstring	$d = 2 + 1$	$d = 3 + 1$	$d = 5 + 1$	$d = 9 + 1$

The construction of the twistor transformation, the $N = 1$ supersymmetric Yang-Mills theory and the Green-Schwarz superstring rely on certain identities of the Dirac Gamma matrices, which work only in certain magic

dimensions. In these dimensions, there is an exact equivalence between the Lorentz group and the special linear tranformations of the real, complex, quaternionic and octonic numbers. Our work shows that QH liquids work only in certain magic dimensions exactly related to the division algebras as well! In fact the *transverse* dimensions $((D+1)-2)$ of these relativistic field theories match exactly with the *spatial* dimensions of the quantum liquids. The missing entries in this table strongly suggests that an octonionic version of the QH liquid should exist and may be deeply related to the superstring theory in $d = 9 + 1$. QH liquids exist only in magic dimensions, have membranes and look like a matrix theory. They may be mysteriously related to the M theory after all!

IV. CONCLUSION

Fundamental physics is faced with historically unprecedented challenges. Ever since the time of Galileo, experiments have been the stepping stones in our intellectual quest for the fundamental laws of Nature. With our feet firmly on the ground, there is no summit too high to reach. However, the situation is drastically different in the present day. We are faced with a gap of 16 orders of magnitude between the energy of our experimental capabilities and the summit of Mount Planck. Without experiments, we face the impossible mission of climbing up a waterfall!

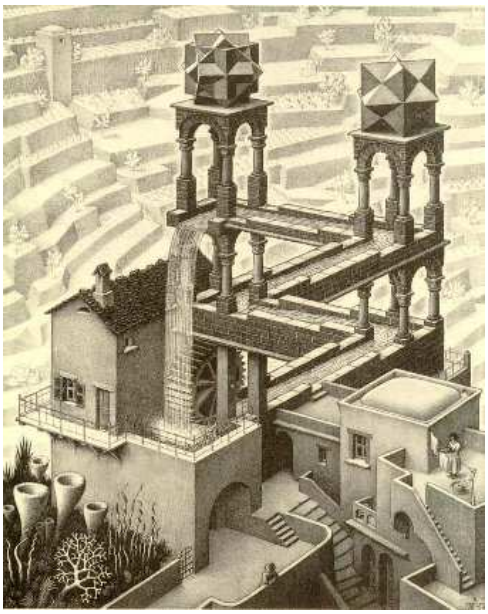


FIG. 5 Esher's waterfall: an alternate passage to Mount Planck?

But maybe there is an alternate passage to Mount Planck. The logical structure of physics may not be a simple one-dimensional line, but rather has a multiply connected or braided topology, very much like Esher's famous *Waterfall*. Instead of going up in energy, we can move down in energy! Atoms, molecules and quantum liquids are made of elementary particles at very high energies. But at low energies, they interact strongly with each other to form quasi-particles, which look very much like the elementary particles themselves! Over the past forty years, we have learned that the strong correlation of these matter degrees of freedom does not lead to ugliness and chaos, but rather to extraordinary beauty and simplicity. The precision of flux quantization, Josephson frequency and quantized Hall conductance are not properties of the basic constituents of matter, but rather are emergent properties of their collective behavior. Therefore, by exploring the connection between elementary particle and condensed matter physics, we can use experiments performed at low energies to understand the physics at high energies. By carrying out the profound implications of these experiments to their necessary logical conclusions, we may learn about the ultimate mysteries of our universe.

Throughout John Wheeler's life, he tackled the big questions of the universe with an unorthodox vision and a poetic flair. Lacking John's eloquence, I simply conclude this tribute to him by reciting William Blake's timeless lines:

To see a World in a Grain of Sand
And a Heaven in a Wild Flower,
Hold Infinity in the palm of your hand
And Eternity in an hour.

ACKNOWLEDGMENTS

I would like to thank A. Bernevig, C.H. Chern, J.P. Hu, R. B. Laughlin, J. Polchinski, P. SanGiorgio, L. Smolin, L. Susskind, N. Toumbas and G. Volovik for stimulating discussions. This work is supported by the National Science Foundation under grant number DMR-9814289.

References

Ambegaokar, V., B. I. Halperin, D. R. Nelson, and E. D. Siggia, 1980, Physical Review B **21**(5), 1806.
Anderson, P. W., 1972, Science **177**(4047), 393.
Balents, L., M. P. A. Fisher, and C. Nayak, 1998, International Journal of Modern Physics B **12**(10), 1033.
Belavin, A. A., A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, 1975, Physics Letters B **59**(1), 85.
Bernevig, B. A., C.-H. Chern, J.-P. Hu, N. Toumbas, and S.-C. Zhang, 2002, Annals of Physics **300**, 185.
Douglas, M. R., and N. A. Nekrasov, 2001, Reviews of Modern Physics **73**(4), 977.
Elvang, H., and J. Polchinski, 2002, hep-th/0209104 .
Fabinger, M., 2002, hep-th/0201016 .
Fisher, M. P. A., and D. H. Lee, 1989, Physical Review B **39**(4), 2756.
Franz, M., Z. Tesanovic, and O. Vafek, 2002, cond-mat/0203333 .
Girvin, S. M., and A. H. Macdonald, 1987, Physical Review Letters **58**(12), 1252.
Haldane, F. D. M., 1983, Physical Review Letters **51**(7), 605.
Hu, J. P., and S. C. Zhang, 2002, Physical Review B **66**, 125301.
Kivelson, S., D. H. Lee, and S. C. Zhang, 1992, Physical Review B **46**(4), 2223.
Laughlin, R. B., and D. Pines, 2000, Proceedings of the National Academy of Sciences of the United States of America **97**(1), 28.
Penrose, R., and M. MacCallum, 1972, Physics Reports **6**(4), 241.
Read, N., 1989, Physical Review Letters **62**(1), 86.
Simon, S. H., and P. A. Lee, 1997, Physical Review Letters **78**(8), 1548.
Sparling, G., 2002, preprint .
Stone, M., 1990, Physical Review B **42**(13), 8399.
Tze, C. H., and S. Nam, 1989, Annals of Physics **193**, 419.
Volovik, G. E., 1993, Jetp Letters **58**(6), 469.
Volovik, G. E., 2001, Physics Reports-Review Section of Physics Letters **351**(4), 195.
Wen, X. G., 1990, Physical Review Letters **64**, 2206.
Wu, Y., and A. Zee, 1988, Phys. Lett. B **207**, 39.
Zhang, S. C., 1992, International Journal of Modern Physics B **6**(1), 25.
Zhang, S. C., T. H. Hansson, and S. Kivelson, 1989, Physical Review Letters **62**(1), 82.
Zhang, S. C., and J. P. Hu, 2001, Science **294**(5543), 823.