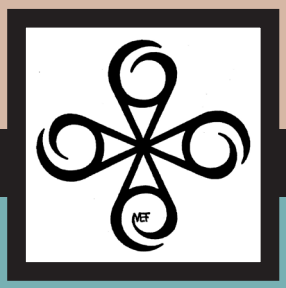
Doing Physics:

A Festschrift For **Thomas Erber**



Edited by *Porter Wear Johnson*

Doing Physics:
A Festschrift for Tom Erber

Edited by
Porter Wear Johnson

Illinois Institute of Technology
IIT Press

Doing Physics: A Festschrift For Thomas Erber
edited by Porter Wear Johnson

Published by:
IIT Press
3300 S. Federal St., 301MB
Chicago, IL 60616

ISBN: 1-61597-000-2

Doing Physics:
A Festschrift for Tom Erber

Preface

> "Äußerten wir oben, daß die Geschichte des Menschen den Menschen darstelle, so läßt sich hier auch wohl behaupten, daß die Geschichte der Wissenschaft die Wissenschaft selbst sei."
>
> Goethe, Zur Farbenlehre: Vorwort (1808)

It is widely asserted that the great physicists who grasped the full unity of physics are all dead, having been replaced in this age of specialization by scientists who have a deep understanding only for issues of rather limited scope. Indeed, it is difficult to refute such a viewpoint today, at the end of the first decade of the twenty-first century. The unifying principles of the quantum theory and relativity are part of the ethos of physics, but fragmented development in various disjointed areas has characterized the past several decades of progress.

Therefore, it has been a refreshing experience to interact with Tom Erber, who has sought to find unity by working in many areas — basic and applied, experimental and theoretical, practical and fanciful. Tom has published about one hundred articles in refereed journals over a period of about five decades. While that number is respectable but not extraordinary, the diversity of arenas in which he has made important contributions is quite striking. Here is a litany of fields that have been significantly influenced by his work:

- classical electrodynamics (radiation reaction, Cerenkov radiation)
- quantum electrodynamics (photoelectric effect, Compton scattering synchrotron radiation, vacuum polarization)
- random processes (randomness in quantum mechanics)
- cooperative systems (hysteresis)

- structural stability (fatigue)

- magnetism (flux compression, micromagnetics, piezomagnetism)

Thomas Erber has been my colleague at Illinois Institute of Technology for more than four decades. I have enjoyed his enthusiasm, his fresh and original approach to substantive issues, and his frequently heretical but always thoughtful assessments of current trends in physics. In addition, he has been a "cheerleader" to me and to quite a few other colleagues in offering timely advice, suggesting new approaches, and proposing extensions of published work.

He gave me one particularly cogent piece of advice, to avoid what he called "elsewhere physics" in written and oral presentations. Namely, be sure to explain ideas, principles, conclusions, and speculations as clearly as possible, and never, never, never to state outright or imply indirectly that a particular result has been explained "elsewhere". Instead, the avid listener or reader deserves to hear the "big picture", as well as the "detailed conclusions" in any scientific presentations.

Students fondly remember Tom Erber's lectures as being lucid, coherent, and elegant. He spoke clearly, and he almost never used written notes for reference. He took particular delight in pronouncing the names of physicists properly in the native language, and encouraging students to do the same. His faculty colleagues have long admired the chalkboards that he left behind after these lectures — with very clear writing, and almost "camera ready" copy on the boards. In fact, his lectures were almost theatrical performances, and surely this is a reflection of Tom's and his wife Audrey's lifelong involvement in amateur theater.

In advanced courses in mechanics, electrodynamics, and statistical mechanics, Tom frequently introduced modern topics in which he had been doing research. He often got students involved in these projects, providing them with their first introduction to the excitement of being engaged in basic research.

As a reflection of the breadth of Tom's interests in physics, this Festschrift in his honor contains articles by former students, colleagues at IIT, collaborators, and friends, which deal with a disparate variety of topics. These contributions are separated into categories, as follows:

Category I: Physics Students of Tom Erber at IIT

Category II: Current IIT Colleagues

Category III: Former IIT Colleagues

Category IV: Other Collaborators and Friends

The final article, by Tom Erber himself, is entitled "Eigenschrift: The End of the Classical Theories of Radiation Reaction". This article represents a survey of work done on this topic over his career, as well as prospects for future research. Tom definitely sees the big picture in science, and many who have known him or have worked with him have been infused with his zest for exploration. It has been an honor and a privilege to have initiated this project and even to have carried through the laborious technical details that have inevitably been involved in its creation. I have learned more than I ever wanted to know about layout, file conversion, and enforcing compability in the various versions of Word^{R} and LaTeX^{R} that are extant in the world. Still, it is important to me as well as to others to tell his story and to honor his career.

Good luck, Tom.

Porter Wear Johnson
Chicago Illinois USA

Email: Porter.Johnson@iit.edu
Website: http://mypages.iit.edu/~johnsonpo/

Contents

0

1

Quantum Cryptography

Bryan J. Field[1]
University of Durham[2]

Here are my recollections of meeting and working with Thomas Erber, a friend and colleague. I describe how we came to study quantum cryptography together in the spring of 1999 and what we discovered.

1.1 Introduction

During my time at IIT, I never had the pleasure of taking a traditional class with Tom Erber – but he had always impressed me as a man of great intellect in my encounters with him in the hallway and after colloquia. In many ways he represented to me an ideal physicist. He was well read, diverse in his research, and was always interested in sharing what he knew or, failing that, tackling a new subject together. We all try to live up to these standards and Tom made it look easy when I was an undergraduate.

That is how we came to study Quantum Cryptography together. IIT had begun the Interprofessional Projects (IPROs) a few semesters earlier and the impetus was on students to propose projects to faculty members who might be willing to sponsor a project. Although participating in an IPRO was not required for my class to graduate, I found them irresistible. I had just finished a project the previous semester with Dr. Harold Spector on the threat due to

[1]bryan.field@durham.ac.uk
[2]Institute for Particle Physics Phenomenology, Durham DH1 3LE, United Kingdom

Earth crossing asteroids when my good friend and colleague, John Katsoudas, suggested the idea of starting another project on cryptography. John and I had long played with several novel ideas on how to encrypt data as a hobby which gave us a good reason to study number theory which always had a kind of siren call for us. This had even lead us to part time jobs in the field of computer security and we often toyed with breaking conventional computer cyphers as a game.

However, the idea of using something brilliant like quantum mechanics to encrypt data was the brass ring. What we knew about conventional code breaking was very mathematical, which we liked, but it relied on knowing a great deal of math that is not typically taught to physicists and we often felt like we were at a disadvantage to someone with a broader mathematical background. We proposed that quantum cryptography might even the playing field. At that time, there was a great anticipation that quantum entanglement could be used as a method of key distribution to guarantee secure communication. We asked Dr. Spector who might help us with a project such as this and only one name came to mind – Thomas Erber.

John and myself along with two other students approached Tom and he thought it was a great project and he was eager to hear what we had to say. We would come to meet with Dr. Erber on Fridays and spent the rest of the week with the project consuming our lives. It was wonderful, thought provoking, a great deal of hard work, and extremely rewarding – in short, it was everything we came to expect of working with Dr. Erber.

1.2 Conventional Cryptography

Cryptography is the study of hiding information [1]. We should briefly note that the terms encrypt and decrypt can be offensive terms in certain cultures where they refer to burial rites. The proper terminology would then be encode (decode) or encypher (decypher), but I shall continue to use the standard American terms. To introduce some other terms, plaintext is encrypted into cyphertext and decrypted back to plaintext though the application of a cypher – an algorithm to obfuscate data.

The history of codes is a long and fascinating one, a wonderful history can be found in Ref. [2]. For our purposes, we can divide conventional codes into two categories – symmetric algorithms and public-key encryption. An important subclass of symmetric algorithms is known as a one-time pad which we will discuss shortly. Symmetric algorithms have been used in cryptography since

the very beginning of covert communication. However, simple symmetric algorithms are often labeled by the community as "security through obscurity" and can virtually always be decrypted to plaintext given a large enough sample of cyphertext, time, and computing power. This is because a symmetric algorithm is merely a mathematical operation – it may be a complicated set of functions that are history dependent (like the German Enigma cypher or the Japanese cypher known as Purple) but they are still just functions. These algorithms then have to be protected as perhaps even more valuable than any of the messages they encode. The modern approach to this problem is to make the algorithm public, but with an enormous key-space – more on this soon. To successfully use a symmetric code to transmit information, both parties have to be in close contact and agree on an algorithm or rely on a trusted messenger of some kind to pass the key. The problem of key distribution remains an open problem for symmetric algorithms and led to the development of public-key encryption.

However, one-time pads do not have these drawbacks. It can be said that if a one-time pad system is used correctly, it ensures perfect security – that is, no amount of mathematical ingenuity or computational strength can decrypt a message encoded with a one-time pad [3]. The concept of a one-time pad is simple, Alice and Bob want to communicate secretly and (for this simplified example) they share a favorite book – let's say a very specific book to ensure no differences between printings will arise, we'll use Hamlet for our example. When Alice wants to send a message to Bob, she writes her message and then compares the message letter by letter to the Hamlet text and performs an operation on the two letters (on a computer this is usually an XOR operation) and sends the result to Bob. Bob uses his copy of Hamlet to decode the message and crosses out all the used letters and sends his reply starting from where the message ended. The message is protected because without knowing that Alice and Bob are using Hamlet, the message cannot be decrypted or even attacked. Each character is essentially encrypted independently from one another, an idea that will come back later. It is easy to see how to make the practice stronger – use random strings of letters instead of Hamlet, and it is easy to see it's weaknesses – if a message is lost in transmission, no more messages can be decrypted, i.e. messages must be decrypted in the order they were written. Also, Alice and Bob have to agree on what pad to use before they begin passing messages which could be very cumbersome for long messages and random pads. In practice, one-time pads have several serious drawbacks – they require perfectly random pads, the pads must be at least as long as the message it is encoding, the method is only as secure as its ability to exchange pads between the sender and the receiver, and the pads must be kept a secret forever and never reused. For now, we will

leave our discussion of one-time pads.

To end our discussion of symmetric algorithms, let us look at how symmetric cyphers are compared. Cyphers are "rated" by the size of their key-space – the number of distinct keys that are possible and are usually written in terms of computer bits. The bits referred to here are the bits in the binary representation of the encryption key, so an n-bit algorithm has 2^n keys. It is easy to see that a 256-bit key-space is 2^{128} times the size of a 128-bit key (about 3×10^{38} times bigger). Obviously, knowing the nature of the algorithm here is of little help if one has to look through a key-space of this size to decrypt a message by force. However, exhaustive searches are rarely the most efficient way to break a code. In the community, a cryptographic "break" is anything faster than a brute force search but requires knowing something of how a message was encrypted. Without getting bogged down in specifics here, most symmetric cryptosystems have known breaks that produces results faster than an exhaustive search, albeit, typically only marginally so.

The near universal use of symmetric algorithms changed forever in the twentieth century with the invention of public-key cryptography. Public-key cryptosystems are based on the idea of a trapdoor function as a way of solving the outstanding problem of key distribution. The strength of public-key cryptography is that it does not require any initial exchange of information. Here one uses two keys – one public and one private. Let's say Alice wants to send a message to Bob. Alice generates a pair of keys and makes the public-key available to Bob (and anyone else for that matter) in a public channel. Bob uses the public-key to encrypt a message and sends it to Alice who uses her private key to decrypt the message. The two keys are related in a way that makes it unfeasible (even in principle) to discover the private key from the public key. This is the so-called trapdoor function.

A trapdoor function (first publicly disclosed in Ref. [4]) is a function that is simple to calculate only in one direction, but its inverse is exceedingly difficult without privileged information or "key". An everyday example of a trapdoor function is the difficulty involved in reassembling a bowl of guacamole back into an avocado. Most modern implementations of this idea revolve around factoring the product of two large prime numbers. We should make the distinction here that even though symmetric algorithms have a key-space that can only be solved by a brute force trial of the key-space, public-key systems also have a key-space that can be attacked by brute force but the complexity of exhausting the public-key key-space dwarfs the complexity of attacking the factorization problem directly even when the known cryptographic breaks are taken into account.

When we say a pair of numbers, we mean that two large prime numbers were picked at random (which is a process that takes time to ensure it is truly random). Since the density of prime numbers to choose from is small compared to the regular integers (which includes the prime numbers), the number of bits used for public keys must be very large to match the key-space of symmetric algorithms. For instance, a public-key of 1024-bits has the equivalent symmetric key-space of 80-bits and is only recommended (by RSA Security) to keep data secure until 2010. An asymmetric 2048-bit key should keep data safe until 2030. NIST claims that a 15260-bit key is needed to match a symmetric 256-bit key. Clearly, the advantage of public-key encryption is not that it is "stronger" than a symmetric scheme (at least not with current trap door functions) but their strength lies in the computationally secure nature of the key exchange for the transmission of data.

1.2.1 Conventional Computing

When we started our IPRO on quantum cryptography, we spent a fair amount of time learning conventional cryptography to see what the state of the art was at the time and what improvements were possible in the immediate future. As we have seen, a big part of that equation has to do with conventional computational speed so perhaps a recap of the progress in computing power over the past decade will not be out of order.

The state of the art computing in 1999 was not all that different from today. Chips have smaller transistor structures and clock speeds are faster than ten years ago. It was not hard to see even then that conventional computer chips would hit a physical roadblock in how many transistors could be packed onto a chip without having quantum tunneling effects make conventional electron switching difficult if not impossible. The then state of the art Intel Pentium III sported a transistor feature size of 250 nm and ran at 500 MHz. Now in 2009, the Intel Core i7 has a feature size of 45 nm and runs at 3.5 GHz with 32 nm structures expected early next year. Quantum tunneling will become destructive when the barrier inside a transistor approaches 5 nm which is projected in 2018. However, in the last ten years the computing paradigm has shifted away from simply packing more transistors onto a single chip with the introduction of multicore chips. It also appears that some complex computing power may be shifted to Graphical Processing Units (GPUs) taking some of the pressure off of CPUs. GPUs have specialized instruction sets which make them better at many mathematical operations. In the end, it may turn out that the key constraint to conventional computing speed will be the time for an electron to flip spins

[5].

Encryption key sizes can always be kept a step ahead of rising computational speed when it comes to brute force attacks. All algorithms can be attacked in an intelligent way to avoid these worst case scenarios. However, looming on the horizon, there is always the possibility that an advance in algorithms could kill conventional cryptography as we know it. The time is ripe to ask ourselves if we can do better by going down a new path.

1.3 Quantum Key Distribution

The limitations of conventional cryptography leads us to the idea, what if we could combine the strength of symmetric encryption codes with a better form of key distribution, a distribution system based on quantum mechanics that would guarantee secure communication? The idea here is to set up a system where if a third party wants to eavesdrop he is forced to measure a quantum state which would introduce an anomaly in the system that could be detected and the communication would be aborted. It should be said that in this simple scenario, it is possible to prevent any coded communications via a denial of service attack (by making all channels insecure) even without an intention of intercepting an encryption key. The real advantage here is that this system does not rely on a trapdoor function, but rather security is a physical property of the communication. If one truly had a secure system of communication, most of the conventional paradigms of cryptography would need to be reexamined.

Information can be encoded into a quantum state, or qubit. Since the current communications network has a backbone of fiber optics, we will talk about the exchange of photons prepared into quantum states. Generically speaking, there are two main classes of quantum key distribution – prepare and measure protocols and entanglement based protocols.

An example of the prepare and measure system is known at the BB84 protocol [6, 7]. Let us imagine Alice wants to send a message to Bob while Eve tries to intercept the message. Alice begins with two strings of bits, a and b, each n-bits long and constructs n qubits,

$$|\psi\rangle = \bigotimes_{i=1}^{n} |\psi_{a_i b_i}\rangle, \tag{1.1}$$

where a_i and b_i are the i^{th} bits of a and b. The basis is given independently for each element in the state vector (reminiscent of the one-time pad) depending

Alice's random bits	0	1	1	0	1	0	0	1
Alice's random basis	+	+	−	+	−	−	−	+
Photon polarization	↑	→	↘	↑	↘	↗	↗	→
Bob's random basis	+	−	−	−	+	−	+	+
Measured polarization	↑	↗	↘	↗	→	↗	→	→
Reveal bases								
Secret key	0		1			0		1

Table 1.1: Illustration of BB84 Quantum Key Distribution

on $a_i b_i$ yielding four qubit states,

$$|\psi_{00}\rangle = |0\rangle, \qquad |\psi_{10}\rangle = |1\rangle,$$
$$|\psi_{01}\rangle = |+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \qquad |\psi_{11}\rangle = |-\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \quad (1.2)$$

such that the b_i bit is the basis the a_i bit. These are not mutually orthogonal states, so it is impossible to determine all of a without knowing b. Alice transmits the state $|\psi\rangle$ to Bob (and perhaps Eve). If Eve intercepts $|\psi\rangle$, Bob would know and would ask for the message to be aborted. Once Bob has a copy of the state he generates a string of random bits b' with the same length as b and compares it to $|\psi\rangle$. Bob may now publicly tell Alice he received the message without it being overheard. Alice publicly announces b and Bob announces b'. When Alice and Bob discard the qubits in a and a' where b and b' do not match, they are left with a private key that they can both use. The key has been securely communicated. This is illustrated in Table 1.1. We can see that this is useful in generating a key, but would not be very helpful in sending a message itself. However, this key could then be used in a symmetric cryptosystem and the message could be passed by conventional means.

An example of a scheme that utilizes entangled states is known as the E91 protocol. This system is based on the idea that the polarization of two photons can become entangled. Alice and Bob would observe a common source of entangled photons that are created at random. Any eavesdropping on this common source of photons would destroy the entanglement and could be detected. A common secret key can be created by this method.

1.4 Conclusion

This is still a relatively young field in the physics community. Working implementations have been demonstrated using the BB84 protocol by several different groups and even more commercial companies. A select handful of financial transactions have been processed through such systems and the results of one election were secured via quantum cryptosystems. It is not yet the age of quantum cryptography, but many of the pieces are in place.

Our IPRO project consisted of myself, John Katsoudas, Jonathan David, and Tony Di Lallo. Although the project did not end with a working quantum cryptosystem as we had hoped, it did give us a solid foundation in the state of the art. Even now, ten years later, I was able to pick up where I left my research to write this review. Even though I did not continue any serious research in cryptography, the experience was fantastic. I learned a great deal from Tom, and I hope he feels the same.

References

[1] Bruce Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C, Second Edition*, John Wiley and Sons, ISBN-13: 978-0471117094

[2] David Kahn, *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*, Scribner, ISBN-13: 978-0684831305

[3] Claude Shannon "Communication Theory of Secrecy Systems," Bell System Technical Journal **28** (4) 656715 (1949).

[4] W. Diffie and M. Hellman, "New Directions in Cryptography", IEEE Trans. Info. Theory 22(6) 644654 (1976)

[5] Lev B. Levitin and Tommaso Toffoli, "Fundamental Limit on the Rate of Quantum Dynamics: The Unified Bound Is Tight", Phys. Rev. Lett. **103** 160502 (2009)

[6] C. H. Bennet and G. Brassard, "Quantum Cryptography: Public key distribution and coin tossing", in Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, p. 175 (1984)

[7] P. W. Shor and J. Preskill, "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol," Phys. Rev. Lett. **85**, 441 (2000), [arXiv:quant-ph/0003004].

2

Well Now Professor Erber

Dave Gavelek[1]
Lockheed Martin[2]

"Invest more in the systematics of eikonal expansions" is representative of the advice provided by Professor Tom Erber and preserved in my collection of old exam papers. I begin with this as an example of one of the most noteworthy truths of Tom's attitude of always looking deeper, "peeling the next layer on the onion." A perfect mark on the examination may be awarded, but there is more to be understood and other approaches that yield additional insight. There is the contextual background and formative events in the lives of the scientists who laid the foundations of the material under study. Tom's lectures included all of this and he expected his students to apply a corresponding effort.

I studied under the guidance of Professor Erber through the 80s, first as an undergraduate, and then as a graduate student and Doctoral candidate. Professor Erber served as my advisor for both my Masters and Ph.D. work. I was awarded a Ph.D. in Physics in 1990 and accepted a job performing data analysis and signal processing for military radars. I have worked all of the succeeding years in this field in a variety of roles including individual contributing engineer, team lead, project lead and functional (people) manager and I am currently a Department manager of a group of about sixty engineers located across three states.

My goal in writing this is to convey the lasting impact that Professor Erber

[1]Email: david.r.gavelek@lmco.com
[2]Department Manager, Sensor Systems Engineering

has had upon me. I might have written about the research we pursued during the course of my graduate studies and how that work has influenced my subsequent career. I considered dusting off some of our unpublished work and offering that as part of this collected tribute. I quickly dismissed these and other options for this collection of recollections of snippets from Professor Erber's lectures. Now, almost twenty years later, these insights have become trusted reference points aiding navigation through a career. Other contributions to this collection will note Tom's contributions to our understanding of the physical world around us. I hope to recognize how Tom also contributed to his student's understanding of the world they were to emerge into and work within.

I must admit that I failed to follow the particular recommendation concerning the investment in eikonal expansions. I suspect that my career, as it has played out, has probably not been adversely affected by this particular failure. In this tribute I intend to focus on a few of the anecdotes and adages that Tom so liberally sprinkled into his lectures and conversations. It is these, at least some small subset of them, which have stayed with me over the years. It is these that I have frequently drawn upon to interpret the events, politics, motivations, successes and failures that I have witnessed and experienced. It is these that I have so frequently found myself drawing upon as I teach. I hope to try to describe how these snippets of wisdom have become so embedded in my worldview that I am certain they are a lifelong influence.

Just as Science is understood through its framework of hypothesis and laws, human experience is understood through our adages and anecdotes. Tom provided his students a deeper understanding of the working of Nature through his explanation of Physics. He also prepared his students for the world they would enter through his insight into the social dynamics of the working scientist.

In preparing to describe these vignettes, I turned to my collection of course notes in search of examples of the insights and phrasing that were interspersed in Tom's lectures. I found that I failed to capture these in my notes. Frustrated in my futile search, I realized that this offering is the realization of something I wished I had done back in school, all those years ago. I distinctly recalled listening to a lecture in one of the last courses I took with Professor Erber and thinking that I should have captured all of the stories and anecdotes that Tom had used during the many courses he had a taught me.

Each of Tom's lectures was packed with so much information that I could only capture in my notes a small fraction of the presentation and I was focused on the Physics much more than the supporting stories and interpretations. Writing this now, rather than as a student, suffers from all that I have forgotten. I am certain that this is much. Alternately, that which has stuck has survived the

crucible of experience and what follows is offered with respect and thanks to my teacher.

As for the title, it was Tom's typical way of beginning a conversation with me, "Well now Mr. Gavelek" What better way to begin this recollection?

2.1 Peeling the Onion

I have already mentioned, "peeling the onion." This phrase is not original to Tom, but I do associate it with him. It is so characteristic of his approach to science, and he certainly used this metaphor. I always think of this metaphor when approaching an aspect of the engineering process that appears uninteresting. Many steps of a solid engineering process are unappealing to many engineers; documentation, test planning, requirements management, and documentation are examples.

"I understand that these activities are necessary, but they're not for me" is a prevalent attitude. Most of us find the interesting and challenging work to lie in the realm of design. I have found that although activities such as documentation and requirements management may not offer as obvious a platform for creative problem solving as design, they do offer these attractions if you look deeper; i.e. "peel the onion". Writing a useful product, or assembly guide is not easy. Why? Certainly there are challenges and problems to be solved to create accessible, useful documentation. But the interesting questions lurk beneath the surface and are not explored when the task is treated as a required activity. Will the documentation make sense to the typical user? How will it help the novice or expert user? How will it be maintained as the product evolves? How can the experience of the early adopters get effectively incorporated into the guides and how are these experiences relevant to the late adopters? How best can the documentation be translated and adapted to different languages and different cultures? Peeling away at the layers of these questions reveals challenges as difficult as any posed to the design engineer. It is more about the attitude you bring to the task, the deeper you dig the more interesting it will get. If you assume the task is trivial or boring, it almost certainly will prove to be so. If you peel the onion you will be rewarded with a challenge.

As an engineering manager I see that one of my most valued employees is the one I can ask to address a wide variety of concerns. The engineer I can trust in diverse roles has innate adaptability but also an attitude of finding interest in every task. I see this as an aspect of the "onion peeler" personality. Another manifestation of the onion peeler personality is associated with the

most innovative engineers. These are the people, who like Tom, keep digging and looking for deeper insight. They are continually searching for and making associations between problems, approaches, and finding the innovative solutions that take the work forward. The onion peeler may be an innovator or a trusted generalist, but either way, the onion peelers are among the most valuable players to have on your team.

2.2 First with the Least or Next with the Best

I cannot remember a specific context in which Tom used this gem. I do recall that it was one that he used frequently and I seem to recall that he would say it in Latin, "You can be firstus with the leastus or nextus with the bestus," but that may just be a trick of memory. Although, Tom used this to characterize the process of discovery and the development of physical theories, it also captures a primary dynamic of the engineering development cycle. Let me reduce this cycle to two, primary modes 1) creation of a new capability or product, and 2) refinement of that capability or product.

The scientist or engineer working on a new product or concept is motivated to quickly provide the basic capability or describe the essence of the idea. Being first allows the flag to be set on the shore, capturing a strategic position for continued work. Delay risks the claim jumper, loosing recognition of being first and arguably, more importantly, loosing the strategic position that First provides. The engineer working at the leading edge is conflicted by the desire to stake out as much territory as possible before filing with the claims office and by the understanding that the claim will be worthless if someone else files first. These concerns seem to have been one of the factors that lead to the long delay between Newton's insight into the nature of Mechanics and his eventual publication of the Principia. I will return to this later. The psychology associated with making the decision to continue to protect the advancement or announce the achievement seems related to business decisions such as keeping a technology proprietary versus releasing as an open standard.

Once the new is established it will be built upon and developed. The next, and all subsequent contributions, must be better than what came before. The challenges faced by the First include convincing others of the promise and overcoming resistance to the new. The challenge of the Next is demonstrating the better. Of course, there will be political and financial hurdles complicating and confusing the analytical/technological claims of improvement. Most of our work as engineers and scientists is devoted to this process of building upon, making

improvements to that which came before, and providing the best as next.

All engineers understand that each product release must be better than the prior. The most successful engineers are those most in tune with the decision makers, cost analysts, customers, competitors and peers who will play a large role in the acceptance and utilization of their work. These are the people that will establish the ground rules of the competition, the aspect(s) of the product that need to be improved, and the metrics that will be used to establish the improvement.

To be successful working at the cutting edge, the engineer must be able to convey the vision and excitement of the new while being grounded enough to know and understand what has come before. It is not easy to be first, but if you are, you just need enough. Not being first is more typical, but not necessarily easier; you must be better. It can be daunting to the new engineer to survey the accomplishments of their more senior engineers and understand that they will need to provide better to be next.

2.3 A Good Experiment

A good experiment is one where you learn something either way it turns out. If the outcome of an experiment is that something has been learned, it will contribute to the eventual solution to the problem. The resolution will be revealed through the systematic application of experiments designed to inform independent of the outcome. This is a criterion that Tom proposed and that I have frequently drawn upon, both as a working engineer and as an engineering manager.

The fast pace of engineering development drives several types of behavior. Using a baseball analogy, I characterize one of these behavior types as the "Slugger". The Slugger is the engineer who likes to work for the quick break through. Sometimes successful, but as in baseball, the Slugger will often strike out. The slugger tends not to use a methodology that guarantees progress with each experiment, and characteristically does not take the time to document what is learned in the pursuit of a solution. Often, this type of engineer will work through the available schedule and then report that "I worked hard, but did not solve the problem." My response typically includes, "Yes, I understand that the problem is difficult and you worked hard, but what did you learn and how did you advance the cause? What kind of experiments did you conduct?"

In contrast, the "Utility Hitter" will achieve few overnight breakthroughs, instead setting a pace for continual progress. In baseball, the utility hitter will

do whatever it takes to advance the runner. In engineering it is the utility hitter that applies a systematic program of good experiments to advance the research. If the problem is not yet solved, the utility hitter does not describe the effort expended on the search for the solution, but rather on all that has been learned and how the solution space has been narrowed. It is this type of engineer that designs experiments that provide insight which ever way the results fall.

Working as an engineer I have been guilty of "swinging for the fences". I have been fortunate however; my third base coach was there to remind me what I was doing before the third strike. As an engineering manager I have been challenged to get the most out of my team. Sometimes I'll let my Sluggers take a couple swings, and sometimes they will be asked to lay down a sacrifice bunt; at least that's how it sometimes feels when I insist on a good experiment.

2.4 You Have to Live Long Enough

In his Statistical Mechanics course, Tom delved into Boltzmann's work. The mathematics and physical relevance of the ergodic hypothesis, the H-theorem, the Boltzmann equation, etc. were covered in detail. In addition to the physics, Tom described the human drama surrounding Boltzmann's achievements. This backdrop made the physical ideas all the more compelling. Boltzmann's work was challenged throughout his lifetime. Most famously, there were debates on the reality of atoms with Ostwald and Mach; the recurrence (Wiederkehreinwand) objection of Zermelo; and the reversibility (Umkehreinwand) objection to the H-theorem. The reversibility objection was raised by Boltzmann's teacher, Johann Loschmid, and subsequently haunted him for more than 20 years, if not until his death. And, his depression and death by suicide, just as the atomic view of nature, and Boltzmann's own accomplishments, were becoming firmly established. Yes, Boltzmann was widely known and respected during his life, but he quit too soon to realize the full impact of his work. Tom said, "You have to live long enough."

As working scientists and engineers, we are all subject to, and guilty of the same emotions of ego that affected the giants. Before our work is judged by time, it is judged by our teachers and others who have worked the problem before us. They are prideful of their efforts and eager to critique ours. We work without the advantage of the clarity time eventually bestows on our concerns. For all the reasons the human condition demands, our work will be a struggle, but deserving work will be recognized if we continue the fight and live long enough.

This lesson has stayed with me because of the multiple times my work has not been immediately accepted, because of the objections that have been leveled by the gatekeepers, the ensuing debates, and the occasional victories. The work is always a struggle (In fact I believe that if you are not struggling and causing some turmoil, then you really aren't trying.) and success requires good work, perseverance and "living long enough." Few are fortunate enough to succeed in tackling a problem as significant as Boltzmann's, but all of us must face similar political challenges. I cannot say that Tom's lectures on Boltzmann's endeavor for acceptance prepared me for the human factor challenges I have faced I imagine that only experience can do that but his lessons did allow me to recognize the struggle and understand that mine is not unique.

2.5 Locked in a Closet

Another historical story conveyed during one of Tom's lectures concerned Newton's writing and publication of the Principia. Newton had formulated the main part of what was to become the Principia by the late 1660s, but did not publish this work for almost 20 years. Newton suffered from reluctance to share his work and only did publish under strong encouragement, and financial backing, from his friend Haley and the threat that his rival, Hooke, was working on a mathematical proof of the inverse square law. The lessons of this story was that for twenty years one of Physics most important works was "locked in a closet", of no use to anyone during that time, and that this work might easily have been lost altogether. Publication, marketing, engagement and other forms of spreading the word is necessary if your work is to be known, appreciated and used by others. Ultimately, the value of your work is measured by how it is used, not how prettily it fills a closet.

I have witnessed many examples of the Newton syndrome. An engineer will "complete" a strong piece of technical work and give a presentation, write a technical memo, or even publish in a refereed journal. Some time later the engineer will complain that no one has read the memo or observe that the work has not been incorporated into the ongoing effort or object that the contribution was not appreciated. I almost always use the example of Newton in my response. Doing the technical work is a necessary but not nearly sufficient condition of success. Completing the technical work is only the first step, and often the easiest part of the path to success.

The most successful engineers and scientists are those that are willing and able to explain their work using a variety of techniques, and able to tailor the

message to their audience. They have their elevator pitch ready in the event of a chance encounter with a senior decision maker. They have the cost-benefit analysis ready for those motivated by cost. They have the detailed equations clearly described and motivated for those who will accept the challenge of understanding the details once you capture their attention. They have a high level overview with clear, illustrative diagrams and relevant examples available for the large segment of their customer base that do not have the time to understand every nuance of the project. Does it answer the requirement? Does it provide too much capability? Is it too risky, or too expensive? It may be elegant but is it practical? The most successful engineers create opportunities to explain their work and refine their sales pitch and are able to speak to the concerns of the decision makers.

One of my colleagues refers to a paper she read describing that, typically, a new idea must be explained seven times before it is understood. It is certainly consistent with my experience that new ideas must be explained multiple times before they are grasped. I was again reminded of this reality this week working as part of a proposal development team. Prior to the team's first meeting I provided a description of the capabilities that my modeling and simulation team could provide for the technical baseline. During the first meeting of the proposal team the leads identified a gap related to target modeling; one of the capabilities I had identified as available. (The team leads' expertise lie in other aspects of the system, and the models used to describe the RF response of the targets is equivalent to a new idea.) I interjected that we did in fact have an in house capability to address this aspect of the emerging design and provided additional detail on how the model had been previously used and why it was appropriate for the current pursuit. In the next day's meeting it was as if the prior day's conversation had never occurred; the lead began by identifying the target models as the primary gap in our existing capability. During the course of this meeting I twice described the relevance of our target models and then arranged a conversation between the proposal leads and my target modeling lead. After that meeting my modeling lead reported that the proposal leads did not yet understand how our capability fit their need. I understand that we still need a couple more explanations before we achieve breakthrough.

In this example, the fact that a solution is available is of no value until the proposal team understands this and works the capability into our plan. The solution has no value sitting in the closet or, more aptly, in the server. For value to be realized the work needs to be understood and used by others and that means that much work remains after the technical work is completed.

For many engineers the technical work is easy compared to the marketing

and sales that are required to get the work out of the closet. For many engineers it is difficult to accept that solving the hard problem is not enough for success and that your added value is measured by your impact on the product not the quality or volume of your work stored in the file cabinet. For me, Tom's interpretation of the publication of Newton's great work is my navigation beacon out of the treacherous closet.

The preceding sayings and anecdotes have been discussed in the context of an engineering research and development workplace. The next two have made their impact on me outside the workplace and contribute to my overall worldview.

2.6 Good and Bad Things About Everyplace

As I write this I find myself in California finishing a business trip and beginning a family vacation. I recall a conversation I had with Tom as I was nearing completion of graduate studies and preparing to relocate to Los Angeles. I was somewhat nervous about leaving Chicago, the city in which I had lived my entire life, and I mentioned my apprehension to him. My only exposure to Los Angeles was that obtained through TV, and I thought it unlikely that I would easily acclimate to the LA scene. Los Angeles, home of UCLA, was one of Tom's sabbatical favorites and I expected he would help sell me on the merits of the city.

Tom told me, "There are good and bad things about every place. You need to seek out and take advantage of the good things and avoid the rest." I was surprised by this neutral advice, having expecting a whole-hearted advocacy for a place I knew he very much enjoyed. Needless to say, I did move to Los Angeles and, consistent with his advice, found much to like. One example stands out.

One of the influences Chicago imparted on me was an appreciation of Blues music. In fact, I wrote most of my Ph.D. thesis with the overnight Blues-format radio station playing in the background. Very good Blues music was (and is) readily available in the Windy City not so in Los Angeles but I was determined to find a club that offered a good dose of the Blues. I tried multiple venues, dragging my wife and new friends along without any success in finding the music that would relieve my homesickness. The music may have been good and it may even have been bluesy, but it was not like back home.

One Friday night we found ourselves in a small club in Hollywood or North Hollywood, I am not quite sure. It quickly became clear once the band started playing that it was to be more of the same more like rock than blues and no sign of any soul. The band played just two songs when the lead stated that he

was going to take a break and invited the audience to get-up and sing. Without any hesitation a well-dressed (LA club style) lady sitting a couple of tables away stood up and approached the microphone. She spoke a moment with the band, they began to play and she began to sing. Wow! She was good! The band caught the fever and for the first time since leaving, I was back in Sweet Home Chicago. I thought that we had been tricked, that the guy who I thought was the band lead was just the warm-up act and this was the feature, at least that's what I thought until the song came to its bittersweet conclusion and another young lady from the other side of the room took the microphone for her chance to impress. Yes, she was, as I recall, even better than the previous singer. Too good I think, because the bandleader returned to his position and did not relinquish it again until after we had lost patience and left the club.

Tom's advice to me on this matter proved to be true. In this case it was clear that LA might not offer Chicago style Blues, but the talent in the audience cannot be topped. Similarly, Tom's advice has always proven as I have relocated and changed employers and, for that matter, as applied to most all life events. The grass may "look to be greener on the other side of the fence" but most certainly there is a price to be paid for the green; i.e. there are good and bad things about everyplace. There are good and bad things about being married/single and bringing/not bringing children into your life. To me, this perspective provides a more grounded philosophy than the more common attitude of "my team is better" - where "team" can be replaced by any type of group that people associate themselves with.

I feel obliged to note that after 13 years living in the Boston area, I have still not bothered to find a venue for the Blues, but still make it a point to visit a club when I'm back in Chicago. Each destination has something good to offer. Our challenge is to seek it out and appreciate it. I now also see Tom's advice to me as a reflection of his approach to life - certainly his approach to science. Each problem holds something of interest to one who seeks it out and works to find the unexpected implications and interrelations. The unexpected advice Tom provided me that late fall afternoon has proven to be much more valuable than the encouragement I sought.

The forms were Complete when Marie Antoinette went to the Guillotine How often have we heard or seen the unjust but legal or the crazy consequence of strict application of policy. Whenever I do, I hear Professor Erber saying, " ... you can bet that when they took Marie Antoinette to the guillotine they had all the forms complete." Unfortunately I do not recall which injustices were the objects of this retort, certainly there is no shortage of examples, and correspondingly, this was one of the sayings he used more than once. Reflecting

on this, two things come to mind. First, the breadth of coverage within each of Tom Erber's lectures; imagine, references to the injustices of the French Revolution in a Physics lecture. Second, Tom's penchant for the poetic; he didn't bother with some mundane comment about legal injustice, but rather used a thought provoking turn of phase that said it all.

2.7 The Sincerest Form of Flattery

In preparing my first set of lecture notes for the undergraduate E&M course I eventually taught several times, I drew heavily from the notes I took as Professor Erber's student through the full set of undergraduate and graduate E&M course offerings. After the first lecture of a new term I was preparing to leave the classroom when Tom entered to prepare for his class which happened to be in the same classroom. I had not erased the boards and as Tom entered he glanced at the remnants of the lecture I had just presented and immediately reflected, "Imitation is the sincerest form of flattery." I recall feeling a little awkward having been caught red handed in my imitation, but I also recognized that the interpretation was right; my lectures were a form of flattery to my teacher. I also recall that this incident taught me to erase the boards after my lectures to avoid providing an overdose of flattery.

Certainly Tom did not originate this saying; but I first heard it on this occasion and I will always associate it with him. Just as Tom's influence was reflected in my E&M lectures, his influence can be found, for example, in how I'll often inquire about a person's progress on a task. The phrasing he used with me was "How is your knitting coming?" I tend to ask, "How's your soup coming?" but have also used Tom's question verbatim. Certainly, my frequent referral to his sayings and anecdotes is an affirmation of Professor Erber's influence on me; as well as my recognition of the thorough incorporation of these insights into my worldview. Evidence of Tom's influence must also lie in my technical work, since he has always sat beside me critiquing the effort, "Would he be satisfied with this explanation?" "Would he peel away another layer of the onion?" (And he has been especially vocal in the preparation of this reflection.) I have but infrequently managed to achieve the quality of his work, but offer the flattery of my imitation.

3

Magnetism Erber Alles

Bruce Harmon
Distinguished Profesor of Physics
Iowa State University
Deputy Director, Ames Laboratory
Director, Center for Physical and Computational Mathematics

Abstract

This is a selected history of my time at IIT when I overlapped with Tom Erber. Also mentioned are some selected 'adventures in magnetism,' which followed.

I grew up in Chicago, on the Northwest side. Early on I discovered the electrical system of our house and blew quite a few fuses doing experiments. That was also a time when chemistry sets were not so restricted by liability concerns, and one could get kits to build radios and other devices (with vacuum tubes). Neither of my parents, nor any adult acquaintances, had gone to college, but they encouraged me to explore. Nonetheless, my knowledge outside the neighborhood was limited. Taft High School was about six blocks north, but I went to Lane Tech, which at the time had about 6000 students – all boys. It became coed a few years after I graduated in January 1965. I went directly to IIT, which admitted a fairly large midyear class. I remember getting a fairly generous scholarship, and found out later it was based on my checking a box on some application indicating that I was interested in metallurgy. I really had no idea what that was. At IIT, one of my early physics classes used the

second volume of the Feynman Lectures in Physics, a book I still consider one of the greatest in the English language. I declared my major as physics, and soon learned that metallurgists (at least students) were better compensated. I graduated from IIT in June 1968, and the last two years I worked with Tom Erber. I don't remember how the first connection was made, but the adventures with him and his group were directly responsible for launching me on my future and continuing endeavors in physics.

The first and many later projects involved the little magnets. Each small cylindrical magnet needed to be drilled dead center. There were questions about the drilling rig and the quality of the hole bottom, which rested on a pin and required minimum friction. I remember bringing in my own microscope, cutting a small circular hole in the column and inserting a half silvered mirror so that one could illuminate the bottom of the hole while viewing the hole bottom through the eyepiece. It worked well, and confirmed that the holes and pin settings were 'clean.' By itself, each magnet could spin on a pin for minutes; but when inserted into the big array, the rather strong horizontal magnetic forces caused additional friction. This rather quickly slowed the collective magnets to a stop. One could reduce the horizontal forces and friction by increasing the distance between magnets, but the complexity of the patterns and interactions were reduced, even if the spinning and fluctuations went on for more time. Professor Erber and his frequent collaborator at the time, Dr. H. G. Latal, worked out the distance dependence of the octupole forces between the little magnets, and showed how they were instrumental in causing a very complex energy landscape. This landscape could be studied ... that became my territory.

Coils were placed underneath the array of little cylindrical magnets and were pulsed rapidly with strong currents. This got the magnets 'boiling.' When the approximately random driving forces were stopped, the kinetic energy of the magnets dropped quickly, the energy dissipation being caused by magnetic hysteresis and friction. With a 30×30 and larger arrays of magnets the number of possible patterns was quite large. My job for some time was to sit and push the 'on' button to boil the magnets, then identify and tabulate the static pattern the magnets had fallen into. Some patterns were common, while others were so infrequent that they were like 'who ordered that?' occurrences. Professor Erber joked that with enough time (or wine?) the pattern would spell 'GOD.' Boring? Yes! But also interesting how one 'becomes one with the machine.' The machine was an insightful model for other phenomena I would encounter later, such as protein folding, or just fast energy quenches in large multidimensional spaces. The State-Area Principle developed by Professor Erber to describe the quenching process was quite general and provided a wonderful heuristic picture.

The little magnets provided more insight in the area of stress strain hysteresis. Another undergraduate summer student and I were playing with the little magnets and noticed that observing the magnets with two extreme lattice geometries gave quite different and incompatible patterns. Imagine four magnets on pins at the corners of a square, then slowly change the angle of one corner from 90^o down to 10^o, keeping the sides of fixed length. We knew the patterns had to be different, but how did the pattern evolve? We went to lunch at the IITRI cafeteria with the idea, and it being a hot summer day, we had popsicles. That was it! We ran back to the lab, pushed pins through the popsicle-sticks and changed the angle continuously. The directions of the magnets changed continuously with angle, but at a critical angle the magnets jumped discontinuously to a lower energy surface and a new pattern. That turned into a larger project with a flexible array shielded from the earth's field by a large mu-metal cylinder . With more magnets, quite complex stress strain hysteresis patterns emerged. Professor Erber, seeing generalizations in other directions beyond magnetism collaborated with Professor Guralnick on a number of papers and experiments dealing with real materials. For magnetic materials, the Barkhausen effect sensitively measured by a squid magnetometer was used to signal the energy jumps.

As a diversion one summer day I was invited to accompany the Erber group for the attempt to achieve a megagauss magnetic field via explosive compression. The experiment took place at an ordnance facility in Indiana. The 'apparatus' consisted of putting the explosives around a low mass pulsed electromagnet. I was some distance from the bunkers when the countdown commenced, and the explosion shook the ground. Running to the bunker and opening the door, I was greeted with dust and darkness, and someone saying "Thanks for opening the door." The lights had been dislodged from the ceiling. After some searching in the chaos, victory was declared when the oscilloscope trace showed more than one megagauss had indeed been obtained.

I might mention another minor episode during this period that has continued to influence me. Some Russian scientists were visiting, and Professor Erber asked if I wanted to come along for lunch in Greektown. The Russians were indeed lively, and I ended up consuming quite a bit of Retsina. I've been a wine enthusiast ever since, although even my Greek friends sometimes wonder about my continued fondness for strongly resinated wine, which nowadays is very difficult to find even in Greece.

I graduated from IIT in June 1968 and started graduate school at Northwestern in the fall. I thought I would become an experimentalist, but at that particular period the exciting research I found was with the new department

chairman, Art Freeman, who got me into computers and electronic structure. My enthusiasm for magnetism never waned, and my thesis was the first spin-polarized calculation for a rare earth material, hcp-Gadolinium. That got me a postdoc at Ames Laboratory, and I've been at Ames ever since. I might have some more titles, but the most important thing in my life has been riding the advancement of computer speed. Problems that I could only dream about solving quickly became feasible, and then routine, as computers followed Moore's law to millions of times more capability. In high school I programmed in machine language, and at IIT I used Fortran and cards. It just keeps getting better, and later this summer I will help organize a national workshop on computational materials discovery, pointing to the anticipated development of exascale computing (10^{15} floating point operations per second). I will be championing magnetism and the need to develop better permanent magnets, not by theory alone, but by using computational insights to finally join our experimental metallurgist colleagues as equal partners. It has taken some time for me to return to my 'metallurgy roots' that got me to IIT so long ago.

4

A Few Memories of Prof Erber

Christopher Merrill
Department of Mathematics
University of Chicago

I came to IIT in the spring of 1983 as a freshman intending to major in engineering. That summer I read two books: a biography of Einstein and an introduction to quantum mechanics. That fall I decided to change my major to physics because I was interested in its deep conceptual and philosophical issues in particular, those of quantum theory. I've never regretted this choice.

Now I was warned at that time, though, that most physicists take a "Shut Up And Calculate" approach when it comes to quantum mechanics. After all, the theory is very successful at generating correct predictions for a wide assortment of experiments – without having to worry about "what is really going on" beneath the mathematical formalism. But I couldn't shake the feeling that somehow Einstein was right to think quantum mechanics was incomplete, or maybe a kind of statistical mechanics of an objective, local reality. So when I met Professor Erber I was pleased to find out he not only thought about foundations of physics, he actively worked on them and was available for lively discussion on the issues.

For instance, Prof Erber introduced me to the work of Hans Dehmelt at the University of Washington, who at that time was attempting to isolate and observe a single electron in an ion trap. I remember Erber telling me about his experiment on a few occasions; because he thought it was possible (but not likely) we'd learn something about the nature of quantum randomness if Dehmelt succeeded in controlling the environment of a single ion, sufficiently so to observe discrete quantum state changes.

Of course, Dehmelt did succeed and won the 1989 Nobel Prize in Physics for this work. But I'm not sure any progress was made on understanding exactly why jumps occur where/when they do. Is it uncertainty about future boundary conditions as well as present boundary conditions, as the traditional Heisenberg uncertainty principle stipulates? I've always wanted to ask Erber more about the interpretation of Dehmelt's results.

And Prof Erber introduced me to conceptual conundrums involving quantum mechanics far beyond anything I'd read, after he learned of my interest. I remember he once quipped: "If you think the story of Schrödinger's cat is weird, you should read about Wigner's friend."

I took all of my required Electromagnetism courses from Prof Erber. I remember his beautiful, Shakespearean-like elocution in lectures. He would always fluently pronounce the names of non-US born physicists in their native tongue ("Al-bert INE-shtine"). He always came to lecture impeccably dressed and well-prepared with an unusual array of handwritten notes very professorial. And his notes were written not on normal-sized notebook paper, but on much larger sheets of paper that he would unfold in class as he drew on the blackboard.

His diagrams were not simply copied from our E&M textbook, Corson & Lorrain, but were usually better and more elaborate. He would diligently use colored chalk to make it easier to distinguish the fields involved (e.g. the E field is green; the H field is blue, etc.). E&M theory leans heavily on vector fields, so it made sense to use large diagrams for illustrating Maxwell's equations, Poynting vectors, div grad curl and all that. Prof Erber knew this and his lectures were always high quality.

He once admonished me in class for not knowing the answer to a question posed in lecture, when I had a copy of Misner Thorne & Wheeler's book on Gravitation sitting in front of me on my desk.[1] He motioned toward the book and said "The real physics is up here". I knew this wasn't a slam on general relativity, but a suggestion that I prioritize my studies toward the fundamentals first – which is always good advice. He pointed out that deep mysteries about space and time exist even within the realm of classical electrodynamics. For instance, how the third time derivative of an electron's position affects its radiative reaction force (self-reaction). And, whether or not a field can have momentum – if it doesn't, then the Lorentz force law violates Newton's Second Law.

Prof Erber suggested I not worry about action at a distance until I learned QED and quantum field theory, where the relation between particles, forces and

[1] The book rather conspicuously adds to the local spacetime curvature.

fields is made explicit. This again was sound advice.

I eventually got my B.S. degree in physics and took a job in finance, but my mind never left certain open questions and conceptual problems in physics. And my conversations with Prof Erber, now over 20 years old, are a big reason why.

If it were possible I'd ask him, with the benefit of additional years of accumulated wisdom, this question:

> In your opinion, what are the most important open questions in theoretical physics?

5

Two and Three-Dimensional Hysteresis: Simple Magnetic Cooperative System

Richard P. Olenick[1]
Department of Physics
University of Dallas

Abstract

The interactions of two magnets (cylinders) described to up to octupole terms interacting with a homogeneous external field leads to two-dimensional hysteresis. The inclusion of the octupole term leads to discontinuities which are not present when only the dilatationally invariant dipole terms are considered. We consider a simple two magnet system and investigate the hysteresis with two (spacing and external field strength) and three hysteresis coordinates (spacing, external field strength, and external field direction). Experimental and analytical studies reveal discontinuous transitions.

5.1 Introduction

Although James A. Ewing [1] coined the word hysteresis during his study of ferromagnetism in 1881, the special properties of magnetite seem to have been known and utilized by ancient Greek and Chinese civilizations at least in the first and second millennia BC. The ancients constructed compasses that take

[1]E-mail: olenick@udallas.edu

advantage of the remanence in magnetite, which is a consequence of hysteresis –or lag – in the applied magnetic field versus the resulting magnetization. Although ferromagnetism offers a classical example, hysteresis also appears in diverse areas ranging from material science to mechanical engineering, physics to biology, and from electronics to economics.

Weiss [2] introduced the concept of a magnetic domain based on spontaneous magnetization and postulated that atoms in ferromagnetic materials had permanent magnetic moments which were aligned parallel to one another over extensive regions of a sample. This was later refined into a theory of 'domains' of parallel moments [3] in which the overall magnetization of a slab of material is the vector sum of the domain magnetizations. In the demagnetized state, this magnetization is zero but as an external field is applied, changes in the domain configuration such as the relative widths of domains or orientations, allow a net magnetization in the field direction. Weiss' hypothesis was later confirmed by direct observation by Bitter [4].

Studies by Erber et al. [5] for two dimensional n x n arrays of finite dipoles (dipoles + octupoles + 32nd poles + ...) revealed that the effective range of the weak, *i.e.*, higher order, forces increases with extreme rapidity as the complexity of the system is augmented. These results have implications for lattice calculations that rely on nearest neighbor approximations, micromagnetic models of ferromagnetic fluids [6], and the origin of domain formation [7]. Indeed, the collective effects of the strongest forces are not necessarily the controlling forces in such situations such as Ewing lattices.

Ewing lattices consist of two-dimensional arrays of pivoted magnets that can rotate in a plane. The potential produced by a single magnetized right circular cylinder may be approximated by the standard expansion of terms for a finite dipole and consists of terms varying as$1/r^{l+1}$, where the $l = 1, 3, 5, \ldots$ terms represent, respectively, the dipole, octupole, 32-pole, etc. contributions. If the distance between the equivalent poles of the cylinder is denoted by d then the expansion can be expressed in terms of the coupling constant $\varepsilon = d^2/(2r^2)$, where r is the distance from the center to the field point.

5.2 Two-Dimensional Hysteresis: Spacing and External Field

The simplest Ewing lattice is a 2×1 lattice and this system can be used to explore simple hysteresis when the spacing between magnets and the strength of an external field are varied. To second order the potential of a magnetized

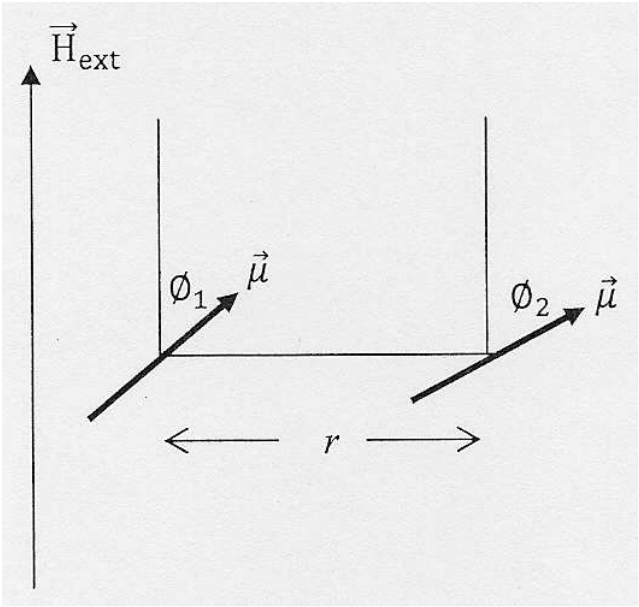


Figure 5.1: Variables for two magnet system.

cylinder may be expressed as

$$V(r,\theta) = \frac{\mu \cos\theta}{r^2}\left[1 + \varepsilon\left(5\cos^2\theta - 3\right)\right] \tag{5.1}$$

where μ is the dipole moment. The interaction energy of a two magnet system is

$$\begin{aligned} U &= -\frac{\mu^2}{2r^3}\left\{\cos\alpha - 3\cos\beta - \varepsilon\left[13\cos\alpha + 5\cos\beta + 5\cos^2\alpha\cos\beta\right.\right. \\ &\quad \left.\left. -35\cos\alpha\cos^2\beta\right]\right\} \end{aligned} \tag{5.2}$$

where $\alpha = \varphi_1 - \varphi_2$ and $\beta = \varphi_1 + \varphi_2$, as shown in Figure 1, where φ_1, φ_2 are measured from the normal to the line connecting the two magnets.

When a homogeneous external field is present, the interaction energy of the magnets with the external field is given by

$$U_{ext} = -2\mu H \cos\left(\frac{\alpha}{2}\right)\cos\left(\frac{\beta}{2}\right). \tag{5.3}$$

The total Hamiltonian for the system therefore is

$$\begin{aligned} H &= -\frac{\mu^2}{2r^3}\left\{\cos\alpha - 3\cos\beta - \varepsilon\left[13\cos\alpha + 5\cos\beta + 5\cos^2\alpha\cos\beta\right.\right. \\ &\quad \left.\left. -35\cos\alpha\cos^2\beta\right]\right\} - 2\mu H\cos\left(\frac{\alpha}{2}\right)\cos\left(\frac{\beta}{2}\right) \end{aligned} \tag{5.4}$$

The physically realizable states are a subset of the extremals which are determined by solutions of the equations

$$\frac{\partial H}{\partial \alpha} = 0, \tag{5.5}$$

$$\frac{\partial H}{\partial \beta} = 0. \tag{5.6}$$

If we assume that $\alpha = 0$, that is, at all times the magnets make equal angle with respect to the direction of the external field, then Eq.(5.5) yields the trivial case whereas Eq.(5.6) requires

$$3\sin\beta - \varepsilon\left[-10\sin\beta + 70\cos\beta\sin\beta\right] - \frac{2r^3}{\mu}H\sin\left(\frac{\beta}{2}\right) = 0. \tag{5.7}$$

Introducing the natural dimensionless field strength $h = H\,r^3/\mu$ and writing $\beta = 2\varphi$, the extrema are defined by

$$\cos\varphi\left[3 - \varepsilon\left(70\cos 2\varphi - 10\right)\right] = h. \tag{5.8}$$

For the ideal dipole case, $\varepsilon = 0$, Eq.(5.8) implies that

$$\cos\varphi = \frac{h}{3}, \tag{5.9}$$

in other words perfect alignment with the external field ($\varphi = 0$)occurs when the dimensionless field strength reaches the critical value $h = 3$, which was first noted by Gallop [8].

The physical states correspond to solutions which are local minima and are determined by the necessary and sufficient condition that the Hessian,

$$\mathcal{H}(\alpha,\beta;\varepsilon,H)=\begin{bmatrix} \frac{\partial^2 H}{\partial\alpha^2} & \frac{\partial^2 H}{\partial\beta\,\partial\alpha} \\ \frac{\partial^2 H}{\partial\alpha\,\partial\beta} & \frac{\partial^2 H}{\partial\beta^2} \end{bmatrix}, \tag{5.10}$$

is a positive definite matrix.[2]

Under the assumptions $\alpha = 0$ and

$$\frac{\partial^2 H}{\partial\beta\,\partial\alpha}=0\,,$$

the condition becomes

$$\frac{\partial^2 H}{\partial\beta^2}>0\,,$$

so that stable states must satisfy the condition

$$h>\frac{3\cos\varphi-\varepsilon\left[-10\cos 2\varphi+70\cos^4\varphi\right]}{\cos\varphi}. \tag{5.11}$$

By writing $\cos 2\varphi = 2\cos^2\varphi - 1$ we may cast the condition of Eq. (5.11) in the form of a cubic equation

$$\cos^3\varphi+p\cos^2\varphi+q\cos\varphi+r=0, \tag{5.12}$$

where the coefficients are

$$p=0,\quad q=-\frac{3+80\varepsilon}{140\varepsilon},\quad r=\frac{h}{140\varepsilon}.$$

For appropriately limited values of h and ε there exist three real roots of which one root is greater than one and hence physically impossible. As h increases, another root becomes physically unrealizable. Finally, the discontinuity is manifest when all solutions ϕ are no longer real: that is, the solution consists of two complex roots and one real root great than one.

Experimental measurements of the critical angles were made with magnets of equal dipole moments of 7 G-cm^3 on supports that could be dilated from 1.5 to 10.0 cm separation. A Helmholtz coil pair provided a uniform magnetic field over the region with intensities ranging between $-5 \le H \le +5$ G. Table

[2] All eigenvalues of a positive definite matrix are positive.

1 provides a comparison between theoretical and experimentally determined angles for values of ε and h.

ε	r (cm)	H (G)	h	$\Phi^{theor}(0)$	$\Phi^{exp}(0)$	h_{crit}^{theor}	h_{crit}^{exp}
0.001	9.93	0.22	2.935	4.5	3.0	2.940	2.935
0.01	3.14	0.643	2.712	0.0	0.0	2.410	2.712
0.01	3.14	0.597	2.518	22.5	26.0	2.400	2.518
0.02	2.22	1.618	2.412	0.0	0.0	2.269	2.412
0.02	2.22	1.544	2.301	-58.5	-42.0	1.800	2.301
0.03	1.81	2.673	2.159	58.9	-2.0	2.356	2.159
0.03	1.81	2.577	2.081	-64.8	-50.5	1.200	2.082

Table 1 Experimental and theoretical values of critical points for the two magnet system.

The values in Table 1 for $\varepsilon = 0.001$ represent nearly pure dipole interactions and verify Gallop's result. The smallest value of the coupling constant for which hysteresis is observed is $\varepsilon = 0.0088$.

Figure 2 illustrates the onset of hysteresis for the case $\varepsilon = 0.04$ in which there is a crossover between extremals at $h_c = 2.510$ and again at 0.600. The abrupt transition to complete alignment and unit magnetization is completely unexpected since all forces vary smoothly. The variation of the external field hysteresis coordinate prods the system along a phase space until singular points are reached. The data also indicate that as the lattice spacing is decreased the energy loss in the hysteresis, which is proportional to the area enclosed by the magnetization versus applied field curve also increases.

5.3 Three-Dimensional Hysteresis: Spacing, External Field, Rotation

More complexity IN the 2×1 system is introduced when the external magnetic field is not in a direction perpendicular to the axis containing the centers of the two magnets but rather at any angle, as shown in Figure 3.

With the introduction of the external field rotated at an angle ψ with respect to the perpendicular to the line joining the two centers of the magnets, the Hamiltonian becomes

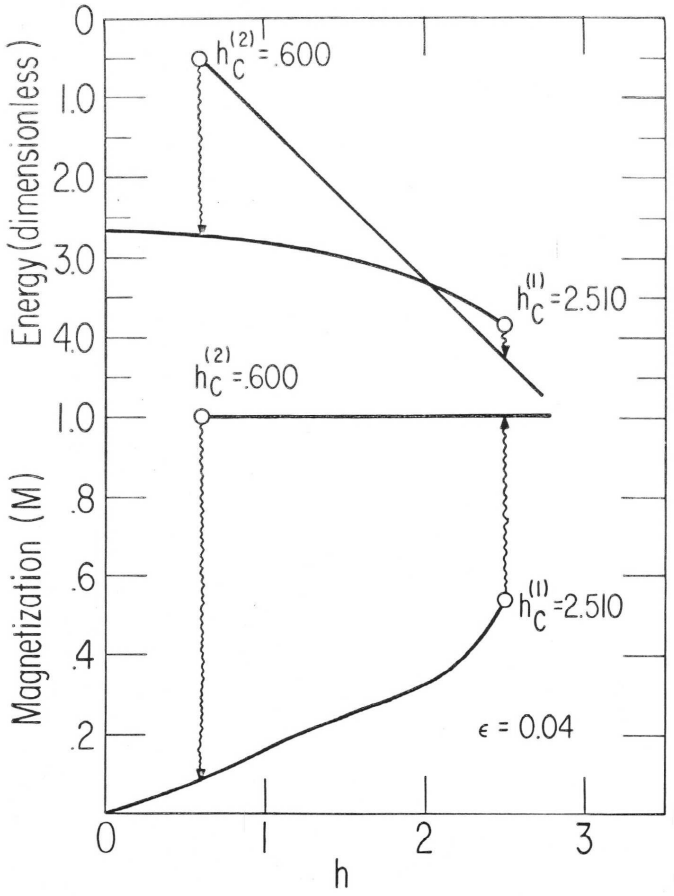


Figure 5.2: Magnetization and energy for the case $\epsilon = 0.04$ illustrating discontinuities

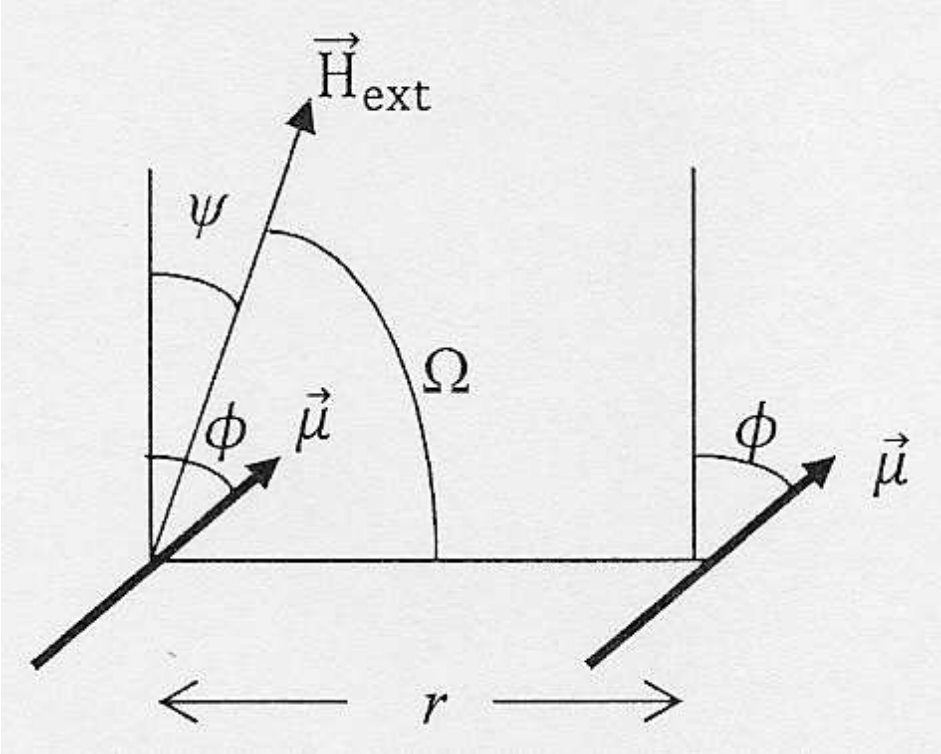


Figure 5.3: Parameters for three-dimensional hysteresis

$$H = -\frac{\mu^2}{2r^3}\left\{1 - 3\cos 2\varphi - \varepsilon\left[13 + 10\cos 2\varphi - 35\cos^2 2\varphi\right] + 4h\left(\cos\varphi\cos\psi + \sin\varphi\sin\psi\right)\right\}, \tag{5.13}$$

where we assume that $\alpha = 0$. The extremal condition leads to

$$h = \frac{\cos\varphi\left[3 - \varepsilon\left(70\cos 2\varphi - 10\right)\right]}{\cos\psi\left[10 + \cot\varphi\tan\psi\right]} \tag{5.14}$$

and stability requires

$$h > \frac{3\cos 2\varphi - \varepsilon\left[-10\cos 2\varphi + 70\cos 4\varphi\right]}{\cos\psi\left[\cos\varphi + \tan\psi\sin\varphi\right]}. \tag{5.15}$$

Table 2 compares the theoretical predictions and experimental results for $\psi = -50$.

ε	r (cm)	H(G)	h	$\Phi^{theor}(0)$	$\Phi^{exp}(0)$	h_{crit}^{theor}	h_{crit}^{exp}
0.01	3.14	0.513	2.164	-52.6	-60.0	2.083	2.164
0.02	2.22	1.435	2.139	-58.0	-55.5	2.081	2.139
0.03	1.81	2.460	1.987	-66.4	-60.5	2.196	1.987
0.03	1.81	3.345	2.702	-9.4	-17.5	2.560	

Table 2 Comparison of experimental and analytical values for critical points for $\psi = -50$.

The predicted behavior of the dipole angle as a function of the applied field for the case of $\psi = -50$ is shown in Figure 4 for several values of the coupling constant.

Experimental checks of the validity of the assumption that $\alpha = 0$ shows that for the case of large lattice spacing and small fields that the percent difference in the angles of the two magnets is approximately 10%. For smaller lattice spacings ($\varepsilon \approx 0.1$) the percent difference decreases to 3% and for stronger fields the value is even lower, being approximately 2%.

Figure 5 shows a cross section of the energy diagram indicating the development of localized "dips" in the overall energy curve due to the octupole interactions. The instabilities occur when the dips spill out and prod the system to a lower energy on the curve. The analysis and results presented here show that for the simplest possible system the origin of hysteresis is the existence

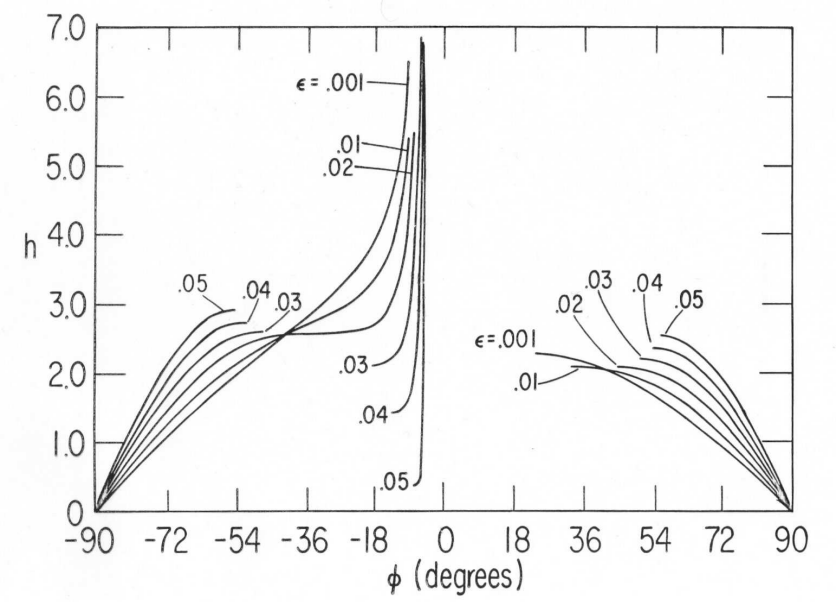


Figure 5.4: Magnet angle for stable configurations as a function of the applied field h.

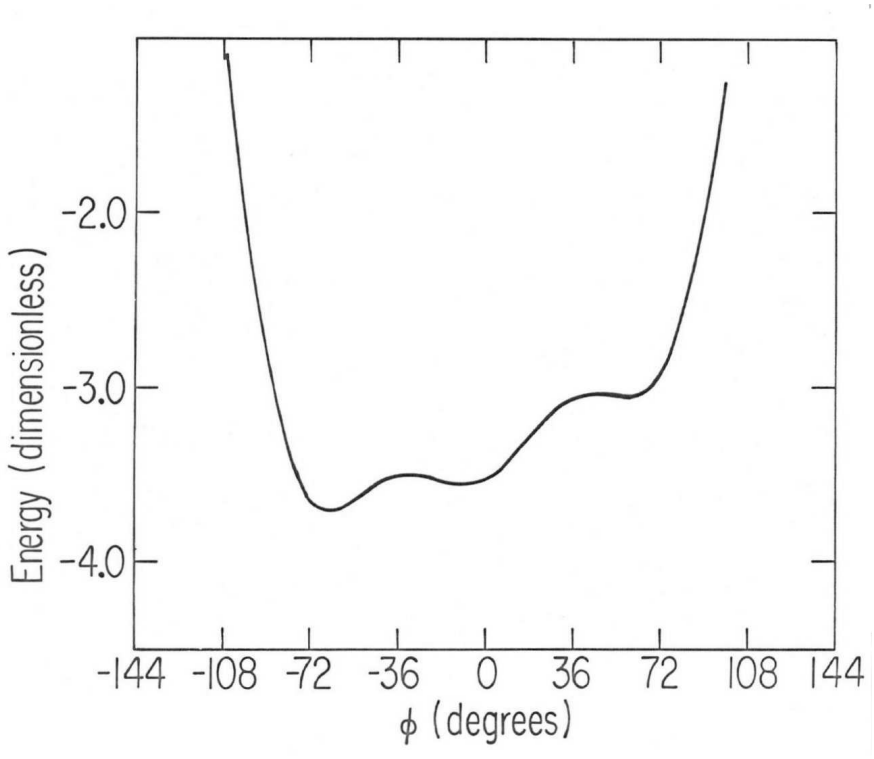


Figure 5.5: Energy surface cross section for $\psi = -5.0, h = 2.20, \epsilon = 0.035$.

of multiple metastable equilibria associated with the system dynamics under consideration.

5.4 Discussion

This investigation shows how relatively simple systems can exhibit discontinuous behavior, further illustrating how the saturation of forces in a complex system can lead to behavior dictated by weaker forces. In addition, the study may have application to a basic understanding of domain wall behavior. In general, two grains meeting at a grain boundary are at an arbitrary crystallographic

orientation to one another, and their easy magnetization directions are not parallel [9]. If the applied field is not sufficient to rotate the grain magnetizations out of their easy directions, there will be a discontinuity in the component of the magnetization normal to the grain boundary, and free poles will be present. The free poles may interact through higher-order forces from nearest neighbors and suffer jump discontinuities.

Kittel and Galt [10] proposed that rigid domain wall motion could be modeled by such interaction and that model has been widely used as a qualitative description of wall energetics and dynamics. Defects, such as dislocations, locally modify the 'constants' characterizing the interaction and magnetocrystalline anisotropy The resulting potential energy wells act as pinning sites, holding the walls in place until sufficient energy is supplied to free them and cause a sudden jump and irreversible behavior.

5.5 Tribute to Professor Thomas Erber

Each of us has the opportunity to shape the lives of those around us. For most of us, we impact the lives of our brothers and sisters, parents, and friends. Occasionally, an individual can influence the lives of a greater number of people through their hard work and persistence. Truly exceptional, however, is the individual who can influence an entire community or a city. Even rarer is an individual who can impact humanity's understanding in fundamental areas of research. Such an individual is Dr. Tom Erber.

I vividly recall my Thermodynamics course that he taught during my undergraduate days at IIT. He is the consummate professor who brought knowledge of the field with presentations that made everyone yearn to learn more. He always had interesting stories to draw us in and then hook us with amazingly clear derivations and discussions. Thermodynamics quickly was my favorite course as an undergraduate. Sadly, my graduate course in it was the antithesis of Tom's and I grew to dislike that area.

I also was privileged to work with him on magnetic cooperative systems. After explaining a project he wanted me to undertake, he let me loose in his lab. I spent every free hour and weekends in the lab tinkering with magnets. He even had Mr. Duda, a master machinist, create a device to dilate arrays of magnets so that I could explore the effects of dilation on systems. Well, when charged with the simple task of painting the aluminum surfaces so that there wouldn't be glare in photos, which back then were used to make angular measurements, I foolishly spray painted the device and got paint on the bearings

and ended up ruining the device. Professor Erber with great calm told me, "If that is the only mistake you ever make, you will be alright." Mr. Duda, on the other hand, fumed and had harsher words for me, and rightly so. Although I leaned more towards theoretical physics, I never lost my love for tinkering in the lab trying to find novel ways to improve undergraduate labs.

This small paper presents unpublished findings of that undergraduate research some 37 years ago. That experience sold me on physics. And Tom Erber's mentoring made me want to be as good a teacher as he was and aspire to walk in his footsteps. Tom seems to have understood the potential within an individual and prompted it to become real. An influential poet of the late 20th century, William Stanley Merwin, expressed this notion in a poem entitled The Unwritten.

> Inside this pencil
> crouch words that have never been written
> never been spoken
> never been taught
> they're hiding

Tom prodded those ideas crouching within us to move from shadow to reality as we began professional careers.

One day in the lab I complained of a headache and he said, "Mr. Olenick if physics gives you a headache then you should get out of physics." Since that day I have never had a headache from physics. On the contrary, I have been very fortunate that my curiosity in physics was piqued by such a great, caring, and respected scientist.

I'll also never forget when he took me to the University of Chicago to hear Werner Heisenberg speak. I still proudly tell my students in quantum mechanics that I heard Heisenberg speak and over a dozen more Nobel Laureates since then. Hearing the founders of areas of physics is something that undergraduates nowadays rarely get to do. Even more fortunate are those whose careers are launched by great physicists and I thank Tom Erber for launching mine.

Recently I saw a program on The Science Channel concerning futuristic shields for spacecraft. The producers visited a high magnetic field lab identical to the one Tom had in the basement of an old building at IIT forty years ago when he was investigating high magnetic fields, implosion, and synchrotron radiation. Tom was always a step ahead of everyone else!

References

[1] J.A. Ewing, Magnetic Induction in Iron and Other Metals, London: Electrician Printing and Publishing Co. (1894).

[2] P. Weiss, "Hypothesis of the molecular field and ferromagnetic properties," J. Phys. 4, 661 (1907).

[3] P. Weiss and G. Foex, Le Magnetisme, Armand Colin (1926).

[4] F. Bitter, Physical Review, 38 1903 (1931).

[5] T. Erber, S.A. Guralnick, and H.G. Latal, Ann. Phys. 69, 161 (1972).

[6] M. Seul and D. Andelman, Science 267, 476 (1995).

[7] T. Erber, H.G. Latal, and B.N. Harmon, Adv. in Chem. Phys. 20, 71 (1971).

[8] E.G. Gallop, Messenger Math. 27, 6 (1897).

[9] J.B. Goodenough, Physical Review 93(4), 917 (1954).

[10] C. Kittel and J.K. Galt, Solid State Physics #3, 437 (1956).

[11] W. S. Merwin, The Unwritten. See
http://www.journaltherapy.com/poem/archive/2003/poem0308.htm

6

The Surface Coulomb Problem: Energy Minima and Hausdorff Metrics

Liam Coffey
Illinois Institute of Technology

Abstract

Professor Tom Erber studied the surface Coulomb problem with collaborators in the late 1990's. This paper provides a brief overview of his contributions to this topic.

The surface Coulomb problem consists of finding the minimum and metastable energy states of N equal charges on the surface of a unit sphere. Using a computer search for the case $2 \leq N \leq 112$, Erber and Hockney [1] identified 2054 such states. For $N \leq 15$, a single minimum energy value was found. However, starting at $N = 16$, metastable energy minima were detected with energies very closely spaced above the minimum energy, with as many as 60 metastable states being identified for $N = 112$, for example.

Furthermore, starting at $N = 15$, equilibrium charge configurations often occur in chiral, enantiomorphic, pairs, with simple ordered arrangements of charges for low values of N, such as $N = 15, 16, 23, ..$, but increasingly irregular, non-random, patterns for larger values of N. The Hausdorff-Mislow metric [2, 3] was used by Coffey, Drapala, and Erber [4], to quantify the structural similarities between these chiral, enantiomorphic, pairs.

6.1 The Surface Coulomb Problem

The set of N unit vectors $[\vec{r}_i, 1 \leq i \leq N]$ denote the positions of N point charges constrained to lie on the surface of a sphere. The Coulomb energy of the system is given by

$$E \;=\; \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|} \;=\; \sum_{i=1}^{N} E_i(N) \tag{6.1}$$

where the term $E_i(N)$ is the partial energy associated with the i^{th} charge. The charge positions can also be described by angular variables on the sphere: the colatitudes $0 \leq \phi_i \leq \pi$ and longitudes $-\pi \leq \theta_i \leq \pi$. The Coulomb energy is then written as

$$\begin{aligned} &E(\phi_1, \theta_1, \cdots, \phi_N, \theta_N) \quad = \\ &\frac{1}{2}\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{\sqrt{[\sin\phi_i \sin\phi_j \sin^2[(\theta_i - \theta_j)/2] + \sin^2[(\phi_i - \phi_j)/2]]}} \end{aligned} \tag{6.2}$$

The function $E(\phi_1, \theta_1, ..., \phi_N, \theta_N)$ describes a $2N$ dimensional energy landscape.

Another quantity of interest is the dipole moment of a configuration which is given by

$$\vec{d}(N) \;=\; \sum_{i=1}^{N} \vec{r}_i \tag{6.3}$$

This quantity should have a magnitude of zero for a spherically symmetric charge distribution, but will be non-zero for the complex, non-random, equilibrium charge configurations in the surface Coulomb problem.

6.2 Results

For $2 \leq N \leq 14$, computer searches lead to a unique ground state energy $E(N)$ for each value of N. Erber and Hockney stored the charge configurations in a standard configuration in which a configuration is rotated so that the charge with the lowest partial energy is at $\phi = 0$ and $\theta = 0$, and the charge with the second lowest partial energy is at zero longitude $\theta = 0$ [1].

To illustrate the nature of the results, consider the $N = 15$ case. At $N = 15$, a unique ground state energy $E(N) = 80.67024411$ exists, but the two charge

configurations associated with this energy form an eniantiomeric pair. Out of 10^4 initial, randomized, configurations with which the $N = 15$ search is initiated, 4958 terminate in one of the two charge patterns, which is labelled $C^L(15)$, and 5042 with the other of this chiral pair, which is labelled $C^R(15)$. The *capture basin* is essentially evenly divided between the two configurations.

The two $N = 15$ configurations consist of five parallel rings, each containing 3 charges. The three charges with the lowest partial energies form equilateral triangles about a great circle. The other four triangles are obtained by imparting a twist to the successive rings. Since these twists can be made clockwise or anti-clockwise, the two resulting charge configurations of the enantomeric pair cannot be superimposed by rotations or proper isometries. The angular co-ordinates of the two $N = 15$ configurations are shown in radians in the accompanying table to three decimal place accuracy. The results of Erber and Hockney determine the two angles ϕ and θ to a much higher degree of precision of 10 decimal places.

$C^L(\phi_1^L, \theta_1^L, ..., \phi_{15}^L, \theta_{15}^L)$			$C^R(\phi_1^R, \theta_1^R, ...\phi_{15}^R, \theta_{15}^R)$		
Charge	Colatitude	Longitude	Charge	Colatitude	Longitude
i	ϕ_i^L	θ_i^L	j	ϕ_j^R	θ_j^R
1	0	0	1	0	0
2	2.094	0	2	2.094	0
3	2.094	3.142	3	2.094	-3.142
4	1.792	2.19	4	1.915	-1.074
5	1.915	-2.068	5	1.915	2.068
6	1.792	-1.012	6	0.979	1.653
7	0.979	1.489	7	1.792	-2.129
8	1.915	1.074	8	0.979	-1.489
9	0.979	-1.653	9	1.792	1.012
10	2.712	-1.354	10	1.010	-2.640
11	1.010	2.640	11	1.010	0.501
12	2.712	1.787	12	2.712	1.354
13	1.184	-2.687	13	2.712	-1.787
14	1.010	-0.501	14	1.184	-0.455
15	1.184	0.455	15	1.184	2.687

As another illustration of the results, consider the case $N = 16$. Three final configurations occur for $N = 16$. In the Erber/Hockney search, 75.7 % of their 10^4 minimizing runs end in a state with $E(16) = 92.91165530$ which is equally divided between two enantiomorphic configurations $C^L(16)$ and $C^R(16)$. The remaining 24.3 % of minimizing runs end up in a state with $E(16) = 92.92035396$, and with a symmetric arrangement of four sets of rings outlining a series of four relatively rotated squares with a charge at every corner.

As N increases, the number of states associated with each energy minimum increases significantly, along with their structural complexity, and the number of distinct chiral pairs for a given N.

If $M_C(N)$ denotes the number of charge configurations found for a fixed value of N, a good fit to the data for $70 \leq N \leq 112$ is obtained with

$$M_C(N) \; = \; A(e^{\nu N} - e^{\nu})/(1 - e^{-\nu}) \tag{6.4}$$

where $A \simeq 0.382$ and $\nu \simeq 0.0497$.

The distribution of values of the energy minima, and the metastable state energies, for $2 \leq N \leq 112$ were investigated in the Erber/Hockney work.

If the lowest energy state for a given N is denoted by $E_G(N)$, for example, the expression

$$E_G(N) \; = \; 0.5N^2 - 0.5513N^{3/2} \tag{6.5}$$

fits the data with error bounds of 0.1% at $N = 20$ and 0.01% at $N = 112$. This can be compared with the average energy of a set of random charge configurations for fixed N which would yield

$$< E^{rand}(N) > = \; N^2/2 \tag{6.6}$$

Lying above the lowest energy states given by Eq.(6.5) are the metastable states which, as was discussed earlier for the example case $N = 16$, are very close in energy with relative energy spacings of approximately

$$\Delta E(N)/E_G(N) \; \simeq 10^{-6} \rightarrow 10^{-4} \tag{6.7}$$

The properties of the energies of the states for a fixed N were studied further by comparing the distribution for partial energies $E_i(N)$ of Eq.(6.1)to the distribution of total energies $E(N)$. The ratio of these two distributions range in value all the way from 0.26 for $N = 32$ to 305 for $N = 45$. Another quantitative

measure of the the total and partial Coulomb energies is the *energy diversity ratio* given by

$$D_e(N,m) \;=\; 100\frac{n_e(N,m)}{N} \tag{6.8}$$

where $n_e(N,m)$ denotes the number of distinct partial charge energies that occur in the m^{th} state of N charges. $D_e(N,m)$ also ranges over a wide range of values from 0 to 100, with the larger values clustering at large values of $N > 70$, approximately.

Distributions of dipole moments, Eq.(6.3), were also studied for the lowest energy, and metastable, states for a fixed value of N. For example, for the 52 states found for $N = 100$, the magnitude of the dipole moment was found to be in the range

$$0 \;\leq\; |\vec{d}| \;\leq 0.0037 \tag{6.9}$$

Taking together the data for all N values, a distribution of non-zero dipole moments is found ranging in value from 0.01322 for $N = 11$ down to 10^{-6} for $N = 38$, with a clustering of values between approximately 10^{-4} and 10^{-3} for $80 \leq N \leq 112$.

6.3 Hausdorff Metrics for the Surface Coulomb Problem

The Hausdorff-Mislow metric was used by Coffey, Drapala and Erber to measure structural information in the complicated arrangements of point charges on the sphere for the charge configurations found by Erber and Hockney.

Distributions of Hausdorff distances were calculated between the two charge configurations of a chiral pair for a given N by fixing one of the enantiomorphic pair charge configurations, and rotating the other set of charges, corresponding to the other member of the pair, relative to it. The resulting distribution of Hausdorff distances yielded a type of "structural spectroscopy", measuring the degree of similarity between the spatial configurations of the two members of a chiral pair.

Furthermore, if $C^L(N)$ and $C^R(N)$ represent the two members of a chiral pair for a fixed N, the lowest saddle point on the energy landscape separating these two configurations was also investigated.

The Hausdorff distance between two sets of charges C^L and C^R is the smallest number $\delta(C^L, C^R)$ satisfying the two conditions: (i) a spherical ball with

radius δ centered at any point of C^L contains at least one point of C^R; and conversely (ii) a spherical ball of radius δ centered at any point of C^R contains at least one point of C^L. This definition ensures that δ satisfies all the criteria for a metric.

In applying this metric to the Coulomb problem, the angular separation between the i^{th} charge of the configuration $C^L(N)$ and the j^{th} charge of $C^R(N)$ is calculated using

$$\Psi_{ij} \;=\; cos^{-1}(\vec{r_i}.\vec{r_j}) \quad 1 \leq i,j \leq N \tag{6.10}$$

where $\vec{r_i}$ and $\vec{r_j}$ denote the unit vectors locating the i^{th} and j^{th} charges on the surface of the unit sphere.

If the resulting Ψij are tabulated in an $N \times N$ array, the smallest Ψ_{ij} values are located in each of the N rows of the table. The maximum value of these N values is then obtained

$$max_i \; min_j \; \Psi_{ij} \;=\; \delta(C^L \;\rightarrow\; C^R) \tag{6.11}$$

The second step is to now collect the minimum values of Ψ_{ij} in all of the N columns of the array. The maximum value of these N values is then found

$$max_j \; min_i \; \Psi_{ij} \;=\; \delta(C^R \;\rightarrow\; C^L) \tag{6.12}$$

Finally,the Hausdorff distance is found by

$$\delta(C^L, C^R) \;=\; max\;[\delta(C^L \;\rightarrow\; C^R), \delta(C^R \;\rightarrow\; C^L)] \tag{6.13}$$

If the C^L configuration is kept fixed, and the C^R configuration is rotated relative to it through a grid of angles, a distribution of Hausdorff distances will result which will have a minimum value: $\delta_{min}(C^L, C^R)$. The distribution of Hausdorff distances should be renormalized by a factor which changes with N value

$$\Psi_{max} \;=\; max_{R_P} \; min_l \; cos^{-1}(\vec{r_l}.\vec{R_P}) \tag{6.14}$$

where $\vec{R_P}$ is the position of an arbitrary point P on the surface of the unit sphere. Finally, the Hausdorff distance is defined as

$$d(C^L, C^R) \;=\; \delta(C^L, C^R)/\Psi_{max} \tag{6.15}$$

Table 3 of reference [4] illustrates the 15×15 table of Hausdorff distances obtained by this procedure for $N = 15$ case, for illustration. This procedure yields

a distribution of minimum Hausdorff distances for all the 1824 paired configurations for $15 \leq N \leq 112$, the 912 values of which range from approximately 0.1 to 0.8. The larger values of d_{min} in this range, between 0.6 to 0.8 approximately, cluster predominantly between the larger values of N from 80 to 112. This trend confirms Mislow's basic conjecture that complex configurations with the least symmetry have the greatest chirality. The minimum Hausdorff distances are shown in Figure 1 of reference [4].

The full distributions of Hausdorff distances for an icosahedral arrangement of 12 charges on a sphere, and for the $N = 15$ enantiomorphic pair of the surface Coulomb problem, are shown in Figures 3 and 4 of reference [4]. The equally spaced sharp peaks, and the broader peak at the Hausdorff distance 0.856 of Figure 3 of reference (4) derive from the ordered symmetrical charge configuration of the icosahedron. Similarly, Figure 4 encodes structural information of the less symmetrical $N = 15$ surface Coulomb chiral pair charge arrangements.

6.4 Saddle Point Heights on the Energy Landscape

The two $N = 15$ chiral pair charge configurations, which have been denoted by $C^L(15)$ and $C^R(15)$, are two degenerate energy minima on the energy landscape, corresponding to two different spatial point charge arrangements.

The variation in the Coulomb energy, $E(15)$, as the $C^L(15)$ charge arrangement is changed over to that of the $C^R(15)$ arrangement was studied in reference [4]. The goal was to find the path passing over the minimum saddle point height going from $C^L(15)$ energy minimum to that of $C^R(15)$. The interesting result was that this deformation path corresponded to a saddle point whose energy was only 0.2% higher than the degenerate energy minima of the $C^L(15)$ and $C^R(15)$ states. The result is depicted in Figure 5 of reference [4].

One of the conclusions from this study is that sequential, one at a time, movements of charges which evolve the $C^L(15)$ configuration into the $C^R(15)$ configuration follow lower energy paths on the energy landscape than collective displacements of charges would.

References

[1] T. Erber and G. M. Hockney, "Complex Systems: Equilibrium Configurations of N Equal Charges on a Sphere ($2 \leq N \leq 112$)", Advances in Chemical Physics, Vol. XCVIII, pp. 495-594, (1997).

[2] A. B. Buda, T. auf der Heyde, and K. Mislow, Angew. Chem. Int. Ed. Engl., Vol. 31, pp. 989-1007 (1992)

[3] F. Hausdorff, Mengenlehre (Berlin: de Gruyter) 1927.

[4] L. Coffey, J.A. Drapala, and T. Erber, "Chiral Hausdorff Metrics and Structural Spectroscopy in a Complex System", J. Phys. A: Math. Gen., Vol. 32, pp. 2263-2284, (1999).

7

A Mechanical Model for Simulating Fatigue Failure in Metals

Sidney A. Guralnick,[1] Jamshid Mohammadi, [2] [3]
and Amy M. Kephart[4]
Illinois Institute of Technology

Abstract

A model for simulating fatigue damage accumulation and the fatigue failure process in metals is presented. The simulation is achieved by modeling material behavior with a series of nonlinear mechanical springs with randomized behavior. With each applied stress, a certain number of springs rupture. The damage accumulation process is modeled by the number of springs that have ruptured during the entire stress application cycle. When a sufficiently large number of springs rupture, the entire system is considered to have failed. This constitutes fatigue failure. This paper follows two previous publications by the first two authors and extends the model further by incorporating additional random variables, investigating the significance of uncertainty in the spring behavior and simulation of the hysteresis behavior of metals during the fatigue damage accumulation process. Results similar to (1) the Wöhler $S - N$ curve and

[1]Department of Civil, Architectural and Environmental Engineering,
[2]Department of Civil, Architectural and Environmental Engineering
[3]Correspondence: E-mail: mohammadi@iit.edu
[4]PDC Engineers, Inc., Anchorage, Alaska, USA

(2) the hysteresis loss versus the number of stress cycle relationship, observed in laboratory testing of metal specimens, are presented.

Key Words: Fatigue Damage, Damage Accumulation, Damage Modeling

Nomenclature

- F : Applied stress
- k : Stiffness
- m : Number of springs
- n : Number of stress cycles
- p_n : Probability
- R : Resistance
- S : Length
- S_X : Standard deviation of X
- u : Strain
- ΔU : Strain energy

7.1 Introduction

The fatigue failure process in materials is a complex process and is affected by the material behavior at the macro level. The material behavior, the macrostructure arrangement in the material, applied stress level, size of the component, existence of prior cracks are among the factors that are known to affect the fatigue failure process. Uncertainties associated with these parameters are the main source of large scatter often observed in fatigue failure behavior of materials. There have been efforts in the past to develop models for describing fatigue behavior reported, for example, by Timoshenko [1], Jackel [2], Martin [3] and Wetzel [4]. However, to date experimental investigations have remained as the only reliable method to establish fatigue failure behavior in materials. Previous mechanical models developed to explain the interlinked set of phenomena needed to describe fatigue failure have been unsatisfactory to a greater

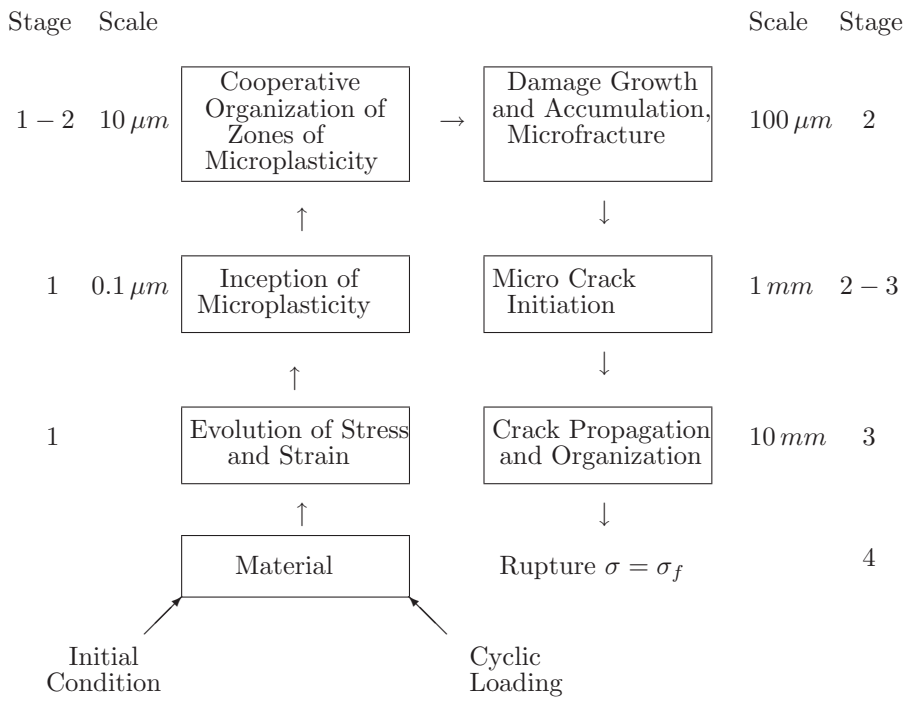


Figure 7.1: Stages in progression leading to fatigue fracture or rupture

or lesser degree because these models simulated only one or two of the many gross macroscopic aspects of metal fatigue.

As may be seen in Fig. 1, the evolution of fatigue failure is assumed to proceed through four stages prior to fracture or rupture. These are: (1) the inception of isolated microscopic zones of plasticity; (2) the "organization" of the microscopic zones of plasticity into macroscopic plastic regions; (3) the initiation of cracks; and (4) complete separation or rupture. Conceivably, stage 1 occurs at very low stress levels and, if stresses remain low, fatigue failure in steels with a definite endurance limit does not occur. At stress levels higher than those needed to initiate microscopic yielding, stage 2 occurs (curve marked P in Fig. 2). If the stresses are raised to still higher levels, then stage 3, or crack initiation occurs (curve F in Fig. 2). This curve is sometimes called "French's line of damage". When the stresses are raised to even higher levels, the cracks, which first appeared in stage 3 propagate and eventually cause rupture as illustrated by the uppermost curve in Fig. 2. This last curve is what is known as the $S-N$ (or Wöhler) curve in fatigue behavior of metals. As is evident from Fig.2, the entire fatigue process (stage 1 through 4) is quite different in the low-cycle region from that which occurs in the high-cycle region. Guralnick [5] describes fatigue phenomena based on incremental collapse of a simple portal frame composed of an elastic perfectly plastic material.

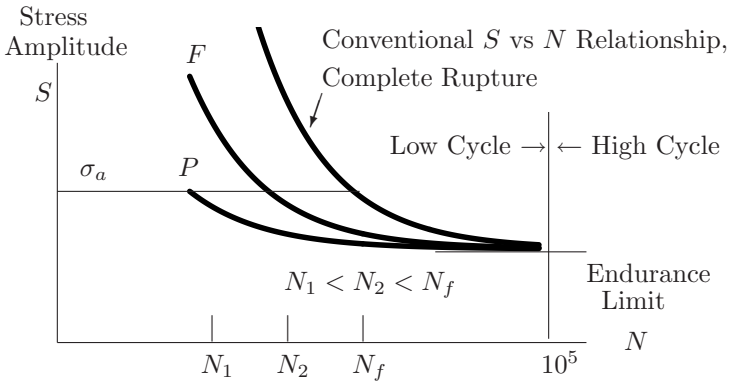


Figure 7.2: Conventional process.

Available models, at best, can only simulate the first two stages of the pro-

gression leading to fatigue failure depicted in Fig. 1. To overcome this shortcoming and to develop a model to simulate the fatigue behavior through the $S-N$ curve, such a model must be capable of describing the non-linear behavior of the material as well as the many sources of uncertainties that result in the large scatter observed in the $S-N$ behavior. This means that the model, more appropriately, requires a probabilistic formulation. Accordingly, Guralnick and Mohammadi [6] developed a simple model to demonstrate the capability of probabilistic formulations in providing a pathway through all four stages culminating in rupture or fatigue. The model was intended for demonstration and consisted of only a limited number of random variables. It was developed as an extension of the model originally proposed by Jenkin [7] to simulate the elastic-plastic mechanical behavior of materials. The model proposed by Guralnick and Mohammadi [6] introduced a probabilistic concept in Jenkin's model for simulating the highly uncertain behavior of materials in fatigue-causing load environments. This model is based on the behavior of a system of parallel springs, which undergo a random stress cycle. Failures among the springs occur at random and can be used as a means to simulate fatigue damage and fatigue behavior. When a sufficient number of springs rupture, fatigue failure is assumed to have occurred. The model used only a small number of random variables and was intended as an initial step to verify its capability in simulating fatigue damage and failure behavior. Additional developments in the model were considered by Guralnick and Mohammadi [8] and Kephart [9]. These new developments include (1) introducing additional random variables in the model; (2) investigating the significance of uncertainties in the spring deformation behavior; and (3) simulating the hysteresis loss versus the number of stress cycles exhibited by metals in a fatigue failure process.

When a material is subjected to a cyclically varying load beyond the elastic limit, the stress-strain behavior will be represented with a hysteretic loop indicating the amount of energy loss in each cycle. In fact, as is known, the area enclosed by the stress-strain curve is the strain- energy density called "hysteresis loss". Each time the material undergoes a cycle of stress, there will be a corresponding amount of hysteresis loss (ΔU_i), where i is the number of the stress cycle. The total hysteresis loss after the material has failed is called the "fatigue toughness" and it is the area under the hysteresis curve when each of the values of hysteresis loss is plotted against the number of stress cycles (N). Three distinct regions are noted in the $\Delta U_i - N$ curve. In the first region, a dramatic drop in the hysteresis loss is observed; while in the second region, a practically constant hysteresis loss is noticed. The third region, just before material failure, exhibits another dramatic drop in the hysteresis loss. When

the hysteresis loss eventually drops to zero, the strain-energy of the material has been depleted, and the material fails. Figure 3 shows these three zones of the hysteresis loss curve for a cyclically loaded specimen. Of course, if the loading stays below the elastic limit, presumably the stress-strain behavior in the unloading and loading cycles will always be represented by a straight line. However, most common steels will show the hysteresis type of behavior depicted in Fig. 3 even if the stress lies below the elastic limit of the material. Each time a fatigue test is done at a different level of stress, the hysteresis loss per cycle in comparison to the number of cycles to failure will be different. At higher levels of stress, the hysteresis loss will have a greater value, but take fewer cycles to dissipate. At lower levels of stress the hysteresis loss per cycle will have a lower value and endure for more cycles. The total area under the different curves will be approximately equal because the material has a certain amount of strain energy to exhaust (Kephart [9]).

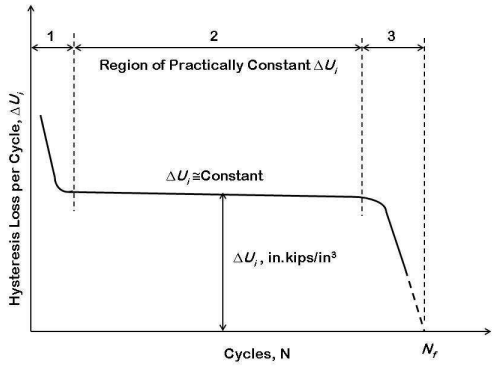


Figure 7.3: Three zones in evolution of hysteresis loss per cycle.

Of particular interest in the behavior shown in Fig. 3, is Region 2, where the hysteresis energy remains constant. In investigating the fatigue failure behavior through simulations, in addition to generating the SN relationship, development of the hysteresis energy loss curve versus the number of load cycles can also be used to test the capability of the model used for simulation as it is demonstrated later in this paper.

7.2 Description of the Model

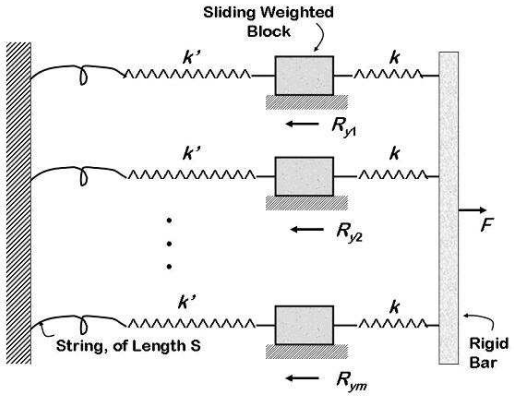


Figure 7.4: Model schematics.

Figure 4 illustrates the mechanical model used in this study to simulate fatigue behavior. Parameter F plays the same role as the stress range does in laboratory specimens used in fatigue failure investigations. The model consists of m parallel springs, which resist the applied stress F. Each spring is a non-linear element with the stress-strain behavior curve shown in Fig. 5. To be consistent with actual stress-strain behavior in metals, the term strain is used in defining the deformation capability of the springs in Fig. 5. To simulate the uncertainties inherent in metal behavior in resisting cyclic stresses and in exhibiting fatigue failure, several key parameters describing the behavior of the springs are treated as random variable. The uncertainty in the spring behavior can specifically be considered from two distinct sources. One is related to the randomness in the variables that describe the stress-deformation curve in the elastic region. The other is related to those variables that describe the stress-deformation in the post-yield region and, specifically, in the strain-hardening portion of the curve as depicted by variables u'_y and u_L and the slope of the curve in the post-yield region (k_e).

In the current formulation of the model, in addition to the spring behavior curve, the number of springs (m) making up the model and the percentage of ruptured springs (N_L) that constitute fatigue failure are also treated as random

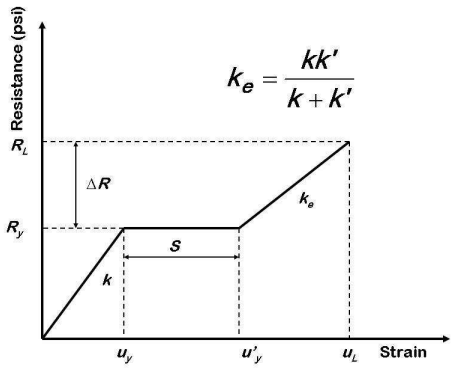


Figure 7.5: Response of a single element model (spring).

variables. All of the random variables in the model are described with their respective means and standard deviations and a prescribed distribution model (e.g., normal). The model can also include the randomness in parameter F to simulate the variations in the applied stress range in structural members such as in highway bridges (Mohammadi, Guralnick, and Polepeddi [10]). However, in most laboratory investigation, the variation in the applied stress is expected to be minimal or even non-existent.

7.3 Basic Formulation

Probabilistic Formulation of Spring Failure The applied stress (F) is distributed evenly among the m springs (see Fig. 4); and as such, the stress in any spring i is $\sigma_i = F/m$. Rupture in a spring will occur when the stress σ_i exceeds the resistance R_L (see Fig. 4). As the deformation and strain in a spring increases, three steps can be recognize as described below:

- Step 1: In the first step, spring k (attached to the rigid bar) will stretch until the stress in the spring has reached a level high enough to move the block (Ry_i). If the stress on the system is not greater than the elastic limit, then the system will never undergo a failure due to fatigue. This step corresponds to the initial linear portion in the stress-strain curve of Fig. 5.

- Step 2: In this step, the block will begin to slide and continue to slide until the string has stretched to its full length (S). This corresponds to the horizontal portion of the stress-strain curve of Fig. 5 and is similar to the plastic portion of the behavior of metals such as steel. When the spring (with constant k) is still in the plastic portion, the second spring (with the constant k') will not stretch at all.

- Step 3: In this step, the stress goes beyond the elastic limit, and the model will begin experiencing perfectly plastic behavior. Once this step has been reached, spring k will continue with its stretch; while spring k' will just begin to stretch. The combined behavior of the two springs will yield the spring constant k_e, in which,

$$k_e = \frac{k\,k'}{k + k'} \tag{7.1}$$

  Both springs will continue to deform until either the stress reverses or spring k' breaks. If spring k' has not yet broken, the behavior will follow the strain-hardening portion of the stress-strain curve of Fig. 5 as marked with a constant spring constant k_e.

- Step 4: If spring k' has not yet broken, then the block will slide backward and the stress-strain behavior will show a stress reversal pattern as shown in Fig. 6. The element will recover a strain in the amount of (Δu_x) which is the total stress on the element (R_x) divided by the spring constant (k):

$$\Delta U_x = \frac{R_x}{k} \tag{7.2}$$

In each cycle where the stress is greater than the yield stress (R_y), the initial length of the string (S) will be reduced to a new length (Sn), which will in turn reduce the ductility of the system. This is an important aspect of fatigue failure. The load history plays a large role in the load capacity of a member. In this model, the length of the string is proportional to the portion of the available plastic region being used:

$$S_n - S\,\frac{R_L - R_x}{R_L - R_y} \tag{7.3}$$

When the stress has returned to zero, the model returns to its original position, but with a residual strain causing the string to be shorter in the next cycle. When spring k' breaks, the behavior changes (as shown in Fig. 7). However,

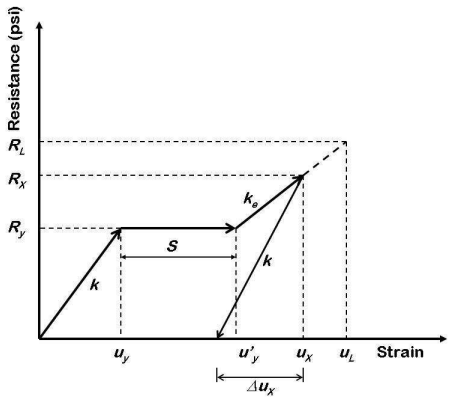


Figure 7.6: Deformation recovery response (springs k and k' intact).

after this stage, the broken element is removed from the system; and the burden of carrying the stress F will be borne by the surviving springs. In such a case, of course, each of the individual surviving springs will carry a larger stress value than was previously the case.

Considering again the system before any element has been removed and upon application of the first stress cycle, the probability of failure p_1 in any one spring is,

$$p_1 = P(R_L \leq \sigma_i) \tag{7.4}$$

Let $Z_1 = R_L - \sigma_i$, then $p_i = P(Z_1 \leq 0)$.

From Fig. 5, R_L and consequently Z_1 can be written as,

$$R_L = R_y + \Delta R = R_y + k_x\,(u_L - u'_x) \tag{7.5}$$

$$Z_t = R_y + k_e\,(u_L - j'_y) - \frac{F}{m} \tag{7.6}$$

where $k_e = k\,k'/(K + k')$, as described by Eq.(7.1) . Parameters R_y, k_e, u_L and u'_y are all treated as random variables. Using a first order approximation, the probability term in Eq.(7.4) can be written in the following form,

$$p_t = \Phi\left(\frac{0 - \bar{Z}_1}{S_1}\right) \tag{7.7}$$

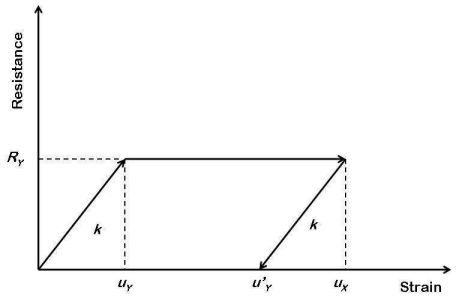


Figure 7.7: Strain recovery response (spring k' broken).

in which $\bar{Z}_1$ and S_1 are the mean and standard deviation of Z_1, respectively, and $\Phi(\cdot)$ represents the normal probability distribution function. They can be estimated using the first order approximation (Khisty and Mohammadi [11]) in terms of the mean and standard deviations of all random variables describing the spring behavior as depicted in Fig. 5. Upon the application of the first stress cycle, the estimated number of springs that survive the load is $(1-p_1)\,m$. After the second cycle, the probability of failure of a given spring is p_2, where,

$$p_2 = P(Z_2 \leq 0) = \Phi\left(\frac{0-\bar{Z}_2}{S_2}\right) \tag{7.8}$$

and

$$Z_2 = R_y + k_e\,(u_L - u'_y) - \frac{F}{m\,(1-p)} \tag{7.9}$$

Again $\bar{Z}_2$ and S_2 are the mean and standard deviation of Z_2, respectively. The number of springs that survive after the second stress cycle is $(1-p_1)\,(1-p_2)\,m$. In general, after the nth stress cycle, if p_n is the failure probability of a given spring, then,

$$p_n = P(Z_n \leq 0) = \Phi\left(\frac{0-\bar{Z}_n}{S_n}\right) \tag{7.10}$$

and

$$Z_n = R_y + k_e\,(u_L - u'_y) - \frac{F}{m\,(1-p_1)\,\cdots\,(1-p_n)} \tag{7.11}$$

Again $\bar{Z}_n$ and S_n are the mean and standard deviation of Z_n, respectively. Note that the product $(1-p_1)(1-p_2)\cdots(1-p_n)\,m$ is the number of springs surviving after the nth stress cycle. Using the first order approximation, a set of general equations for the mean and standard deviation of Z_n can be written as follows,

$$\bar{Z}_n = R_y + k_e\,(u_L - u'_y) - \frac{F}{m\,(1-p_1)\cdots(1-p_n)} \tag{7.12}$$

$$S_n^2 = S_{R_y}^2 + S_{k_e}^2\,(\bar{u}_L - \bar{u}'_y)^2 + \bar{k}_e^2\,(S_{u_L}^2 + S_{u'_y}^2) + \frac{S_F^2}{m\,(1-p_1)\cdots(1-p_n)} \tag{7.13}$$

in which a bar over each symbol indicates the mean value of the respective random variable; whereas, s term such as S_X indicates the standard deviation of a random variable X. Although, parameter F in deriving Eqs. (7.12) and (7.13) is treated as a random variable, when simulating fatigue failure of laboratory test specimens, the variation in F expected to be small or non-existent. Thus, the standard deviation of F in Eq.(7.13) can be taken equal to zero.

Simulation of Fatigue Failure Fatigue failure occurs when a significantly large number of springs have failed. Depending on the type of material, this number is also random; and, in this paper, we used a random variable describing the number of spring failures which constitute fatigue failure. In a given case of applied stress, a random value for N_L is selected. This random value can be generated using a mean value $\bar{N}_L$ and a standard deviation S_{M_L}. Solution Process Simulating Cyclic Damage Process Equations 10 and 11 can be solved in a step-by-step manner starting with $n = 1$ and continuing the computations until system failure occurs. Before each computation run, the mean and standard deviation of all parameters describing the spring behavior are selected. A random number is then generated for the required percentage of ruptured springs, that constitutes fatigue failure (i.e., N_L) and the total number of springs that make up the entire model (i.e, the parameter m). The computation of the terms p_n is then achieved using Equations 10-13 after each stress application. Upon each stress cycle, the percentage of ruptured springs is computed and compared against N_L to determine whether fatigue failure has occurred. Once the percentage of ruptured springs is equal to or exceeds N_L, then the cycle number corresponding to this stress application is recorded. This is the number of stress cycles causing failure.

7.4 Simulation of Hysteresis Loss vs. Number of Stress Cycles

With every cycle of loading that is beyond the plastic region, there will be an associated loss of recovered strain energy. This loss of recovered strain energy is the hysteresis loss per cycle (ΔU_r). The variable (r) is the number of cycles beyond the elastic limit of the element. If the stress is smaller than the elastic limit of the element, there will be no hysteresis loss and the elements can be loaded an infinite number of times without experiencing fatigue. The hysteresis loss for each element is the area under the stress-strain curve for each individual element. The hysteresis loss for the entire model will be the sum of the hysteresis for all elements. If spring k' has not yet broken and the system undergoes a stress reversal process as shown in Fig. 6, The hysteresis loss will be

$$\Delta U_1 = \frac{R_Y}{2}\, u_y + R_y\,(u'_Y - u_Y) + \frac{R_X + R_Y}{2}\,(u'_X - u_X) - \frac{R_Y}{2}\,\Delta U_X \qquad (7.14)$$

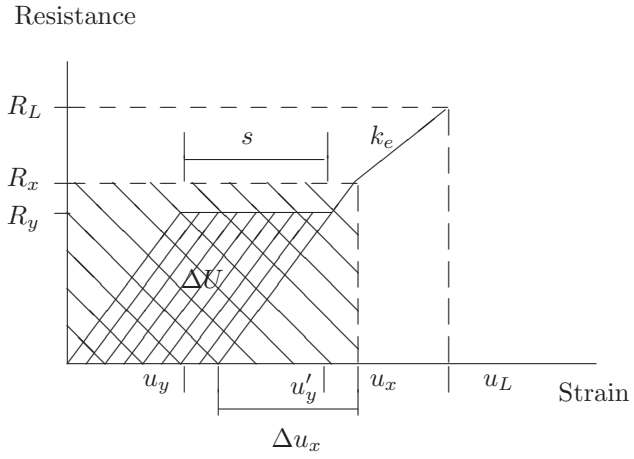
However, if spring k' fails, as was described earlier, from Fig. 7 the hysteresis loss is,

$$\Delta U_2 = R_y\,(u_X - u_Y) \qquad (7.15)$$

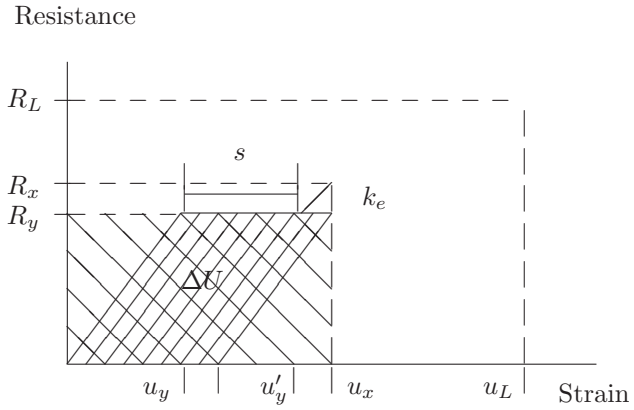
If upon each stress cycle, m_1 elements, of the total of m, survive, then the total hysteresis loss is

$$\Delta U_r = m_1\,\Delta U_1 + (m - m_1)\,\Delta U_2 \qquad (7.16)$$

The hysteresis loss per cycle at the beginning of the strain-hardening region, is shown in Fig. 8a (when spring k' has not yet broken). With each cycle of applied stress, the size of the available stretch in the string (see Fig. 4) will be decreased, and therefore the length of the plastic region will become shorter. This corresponds to a loss of ductility similar to the loss that occurs in metals when they are cyclically loaded above their elastic limit. This loss of ductility has been noted by Puskar and Golovin [12]. The hysteresis loss when spring k' breaks is shown in Fig. 8b for comparison. Figure 9 shows the hysteresis loss near the end of the strain-hardening region of the stress-strain curve. Of course, this is only possible if spring k' does not break. It is important to note that the strain-energy produced by a failed element is larger than that of a surviving element. When the hysteresis loss is no longer decreasing, a sufficient number of elements have failed; and as such, the model simulates the condition similar to fatigue failure in metals.



(a) spring k' unbroken



(b) spring k' broken

Figure 7.8: Hysteresis loss: beginning of strain hardening region.

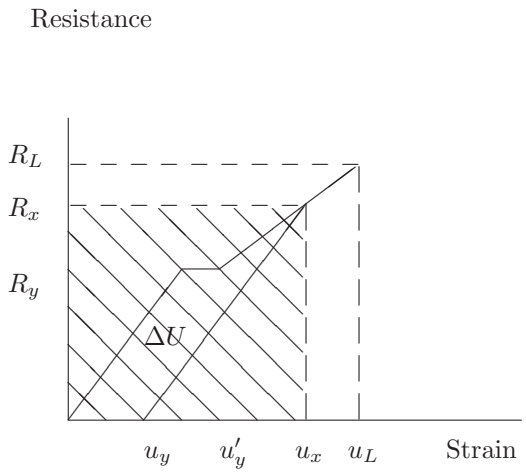


Figure 7.9: Hysteresis loss at the end of stress hardening region of the stress-strain curve.

7.5 Determination of Model Parameters

Spring parameters depicted in Fig. 5 can be determined from simple tests of materials. The standard deviations of individual parameters are the measures of uncertainty in materials behavior and can be estimated from published data on variations observed in metal yield and ultimate strain values. The strain-hardening behavior shown in Fig. 5 appears to play a crucial role in the overall behavior of the material in producing random fatigue failure results (as described later). This strain-hardening behavior can also be investigated by means of a limited number of simple tension tests. The limited number of tests needed to calibrate the spring behavior would require, by far, much less labor and cost compared with the numerous tests needed to investigate the fatigue failure behavior using conventional laboratory investigations. As is evident from the formulation, the parameter m plays an important role in the model and in the simulation process. Generally, m will be proportional to the inverse of the probability of failure p_n. One can begin with a large m as a starting value and revise it later if such revision becomes necessary.

7.6 Special Cases

In extreme conditions, where the applied stress is either very small or very large, special cases for fatigue failure are observed. When the stress is very small, the value of p_n does not change from one stress cycle to another. This case will especially occur when R_L is large (i.e., the ratio $m\,R_L/F$ is very large). And as such, Z_n will be large and result in a nearly zero failure probability. If after a very large number of cycles, there is still no or very little change in p_n, failure is unlikely. This case corresponds to a very low stress range smaller than the "endurance" limit in fatigue behavior of steel. In another extreme condition, when F is very large and R_L is small (i.e., the ratio $m\,R_L/F$ is very small), the value of p_n changes dramatically and approaches a large value rapidly after only a few stress cycles. This condition corresponds to low-cycle fatigue, where after a few stress cycles, failure occurs.

7.7 Example

Table 1: Variability in Model Parameters

Variable	Unit	Mean	Coefficient of Variation
R_y	psi	1	0.2
u_y	in	0.067	0.05
u'_y	in	0.183	0.05
u_L	in	0.283	0.05
k, k'	$\ell bs/in^3$	20	0.14
k_e	$\ell bs/in^3$	10	0.2

To demonstrate the model, the number of elements is taken to be $10,000$. All of the model parameters are treated as random variables and defined with their respective means and coefficients of variation. These values are shown in Table 1. The stress has a range of values between $3,000$ psi and $15,000$. The mean value of the parameter N_L is selected as 25 percent with a coefficient of variation equal to 0.20. Several random numbers for this parameter are also generated using a normal distribution. The random values range from 20.45 to 31.40 percent. The number of elements, m can also be treated as a random variable to further increase the variations to be expected in typical fatigue failures. Within each stress value, the results of the total number of cycles to failure are random and show a rather wide range. Within each stress level, 25 randomly-generated simulations were conducted. The number of cycles to failure and the percentage of surviving elements upon failure were obtained and averaged for the 25 simulations. Table 2 summarizes the total number of stress cycles obtained for various stress values applied to the system.

Table 2: Results of Simulated S-N Relation

Applied Stress (psi)	Average Number of Cycles to Failure	Average Percentage of Surviving Elements
15,000	3	65.1
14,000	5	62.6
13,000	10	52.1
12,000	20	49.0
11,000	42	49.0
10,000	119	49.0
9,000	317	49.2
8,000	1,063	40.5
7,000	4,139	41.4
6,000	17,190	42.1
5,000	26,659	21.6
4,000	71,913	18.2
3,000	326,365	10.9

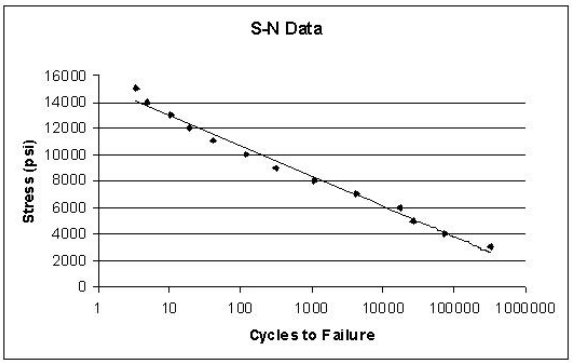


Figure 7.10: Simulation of Wöhler $S-N$ Diagram (stress vs. logarithm of cycles to failure).

As is evident from Table 2, at stresses below 3,000 psi, a very large number of stress cycles will be required to cause failure. A stress value below this constitutes the "endurance limit", at which failure will not occur. Figure 10, shows the results in graphic form using a semi-log scale. This figure clearly depicts the usual $S-N$ (or Wöhler-type) curve. One of the more interesting aspects of the strain-hardening region is the hysteresis loss per cycle. Because

(r) can have such a large range, the hysteresis loss for each of the stress levels must be analyzed for several different (r) ranges. Table 3 shows the values of the hysteresis loss per cycle at different stress levels. This table is just one example of the values that the hysteresis loss can take on at these stress levels. The randomness inherent in the model output also affects the hysteresis results. Upon repeating the same stress levels in a new run, different results will be generated within the uncertainties implemented in the model parameters. When the (r) values of hysteresis were plotted at stress levels of $13,000$ psi, $11,500$ psi, and $10,000$ psi the behavior is similar to what would be seen for an actual metal under different stress levels. Figure 11 shows this relationship. This further shows that the model is capable of simulating the fatigue behavior in metals.

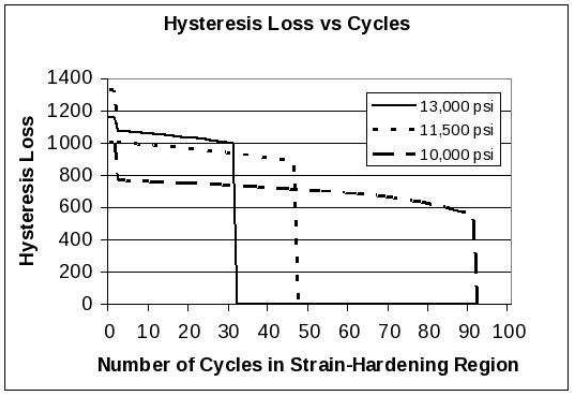


Figure 7.11: Simulated hysteresis loss vs. cycles for various stress levels.

Table 3: Hysteresis Loss at Different Stress Levels

Stress → Cycles ↓	13,000 psi	11,500 psi	10,000 psi
1	1159	1326	999
5	1072	994.199	762.179
10	1061	984.844	757.384
15	1049	974.66	752.327
20	1035	963.469	746.974
25	1017	951.033	741.286
30	999.208	937.039	735.214
35	0	921.127	728.699
40	0	903.238	721.666
45	0	887.25	714.018
50	0	0	705.628
55	0	0	696.327
60	0	0	685.877
65	0	0	673.934
70	0	0	659.974
75	0	0	643.141
80	0	0	621.922
85	0	0	593.448
90	0	0	556.209

7.8 Sensitivity of Results to Variation in Material Properties

An important feature of the model presented herein is its ability to portray the scatter in fatigue life, which is often observed in experiments with metals. To further explore the main sources of this scatter, the influence of the variation in model parameters on the final outcome for fatigue life was investigated. Specifically, three sources of parameter variations were evaluated. These were: (1) variation in the parameters describing the elastic portion of the spring behavior curve; and (2) variation in the parameters describing the post-yield region (i.e., strain-hardening region and the onset of plastic behavior) in the spring behavior curve. The results indicate that the variability in the parameters describing the material behavior in the post-yield region has the most significant effect on the variation observed for fatigue life. The variability in the parameters describing the elastic region of the spring behavior curve is not a major factor in the

fatigue life scatter. This variability will be shown to have negligible influence on fatigue life. Figure 12 presents the significance of variability in the parameters describing the post yield region of the spring behavior curve compared with the variability in parameters describing the elastic region of the curve. As may be seen in this figure, as the variability in the elastic parameters increases, the change in the variability in fatigue life is only marginal. However, when the variability in post-yield parameters increases, there is a dramatic change in the variation obtained for fatigue life. These results are consistent with actual fatigue behavior in metals. As explained earlier (see Fig. 1), the inception of fatigue damage starts by the formation of microscopic zones of plasticity and the organization of these zones into macroscopic plastic region which lead to crack initiation. To a great extent, the scatter inherent in fatigue life is affected by randomness in the plastic regions which is predominately affected by material non-linearity. In our model, this non-linearity is represented by parameters describing the strain-hardening and the onset of the plasticity region in the spring behavior.

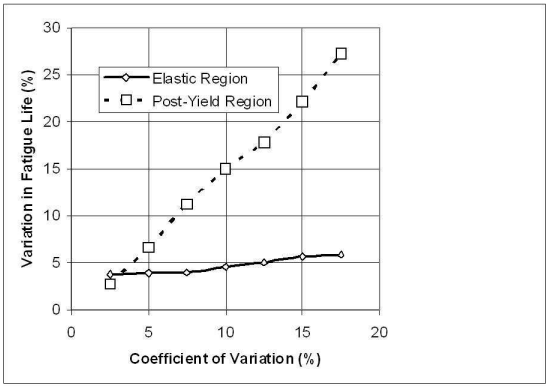


Figure 7.12: Effect of parameter uncertainty on variation in fatigue life.

7.9 Summary and Conclusions

The model presented in this paper can be used to study fatigue process in metals without an excessive appeal to empiricism. The model's capability is demon-

strated in developing the $S-N$ relationship with characteristics similar to those associated with analogous curves obtained in actual experiments with metals. To apply the model to a broader fatigue study, several additional developments will be necessary. Specifically, an approach must be developed that can be used in calibrating the parameters describing the spring behavior in the model for a given material. We believe that this can be achieved with a limited number of simple laboratory tests on material behavior and calibration of the spring model parameters with a few cyclic load test results (as described earlier). This area is under further investigation by the first two authors in a series of comprehensive studies on understanding the inception and origin of fatigue behavior in metals. Specifically, a methodology based on the minimization of the difference between the strain energy of the spring model and a specific metal specimen is being investigated to obtain realistic values for the model parameters.

The following summarizes the main conclusions:

- A model made of a large number of nonlinear parallel springs can be used to simulate fatigue failure in metals.
- The results from simulating fatigue behavior from this model show that most of the variation in fatigue life is inherent in the strain-hardening and plastic behavior of the non-linear springs.
- Specific capabilities of the model are in simulating (1) the $S-N$ relationship; and (2) observed mechanical hysteresis behavior in metals subject to cyclic stresses and fatigue failure.

References

[1] Timoshenko, S.P. (1956). Strength of materials, Vol. II, Third edition, Van Nostrand Company, Inc., Princeton, NJ, p. 413 and 514.

[2] Jackel, H.R.(1970). "Simulation, duplication and synthesis of fatigue load histories", Sound and Vibration, Vol. 4, p. 18-29.

[3] Martin, J.F., Topper, T.H., and Sinclair, G.M. (1971). "Computer-based simulation of cyclic stress-strain behavior with applications to fatigue", Materials Research and Standards Journal, ASTM, Vol. 11, No. 2.

[4] Wetzel, R.M. (1971). A model of fatigue damage analysis, Ford Motor Company Scientific Research Staff Technical Report, Privately published.

[5] Guralnick, S.A. (1975). "An incremental collapse model for metal fatigue", Publ. Intern. Association of Bridge and Structural Engineers, Vol. 35, Part II, Zurich, Switzerland.

[6] Guralnick, S.A., and Mohammadi, J. (2002). "The origin and inception of fatigue in steel – a probabilistic model", in Recent Advances in Experimental Mechanics, E.E. Gdoutos (ed.), Kluwe Academic Publishers, Netherlands, p. 187-196.

[7] Jenkin, C.F. (1922). "A mechanical model illustrating the behaviour of metals under static and alternating loads", Engineering, Vol. 114, p. 603, 1922.

[8] Guralnick, S.A., and Mohammadi, J. (2005). "A probabilistic model to simulate the origin and inception of fatigue failure in metals", Proc., 11th International Conference on Fatigue, Department of Structural Engineering, Politecnicodi Turin, Italy.

[9] Kephart, A. M. (2006). A probabilistic model for fatigue failure in metals, MS thesis, Illinois Institute of Technology, Chicago, Illinois.

[10] Mohammadi, J., Guralnick, S.A. and Polepeddi, R. (1998). "Bridge fatigue life estimation from field data", Practice Periodical on Structural Design and Construction, ASCE.

[11] Khisty, C.J., and Mohammadi, J. (2001). Fundamentals of systems engineering, with economics, probability and statistics, Prentice Hall.

[12] Puskar, A., and Golovin, S.A. (1985). Fatigue in materials: cumulative damage processes, Elsevier Science Publishing, Inc., New York, NY p. 187.

8

Energy Conservation: Science or Ideology?

Porter W. Johnson[1]
and David Atkinson[2]
Illinois Institute of Technology

"I no longer believe in the existence of – nor the conservation of – energy."

- Thomas Erber (October 1976)

8.1 Introduction

One topic of long-standing interest to Tom Erber has been his exploration of the nature of energy conservation and its limitations in practice. We shall make a few comments on this topic here.

Energy conservation appears to be a direct and inevitable consequence of Newtonian mechanics, in that, for systems which interact through conservative potential fields or which undergo elastic collisions, it leads directly to conservation of energy. The work-energy theorem accounts for the energy balance, the net work being either stored as potential energy, or else presumably converted into other forms of energy. Indeed, in the context of the special theory of relativity, energy and and momentum conservation are placed upon the same

[1]Professor of Physics, Emeritus. E-mail: johnsonpo@iit.edu
[2]Professor of Physics, Emeritus, University of Groningen (NL)

footing, since for either of them to be satisfied in every inertial frame, they must both be met. Although the connection between conservation of energy and conservation of momentum is more tenuous in Newtonian mechanics, the relativistic extension of this theory treats these requirements on a unified basis, and tied to basic symmetries of space-time invariance. How could one imagine a "physical reality" without these cornerstone principles?

8.2 Zeno Balls

Actually, there has been a substantial amount of recent exploration of this question – by philosophers (see References [1] - [3]) rather than by physicists[3], who apparently have been pre-occupied with the "Zen" aspects of such things as the Higgs boson, superstrings, and dark matter. These analyses have pointed out certain difficulties, which can be encapsulated in the classic elementary demonstration apparatus known as Newton's balls.

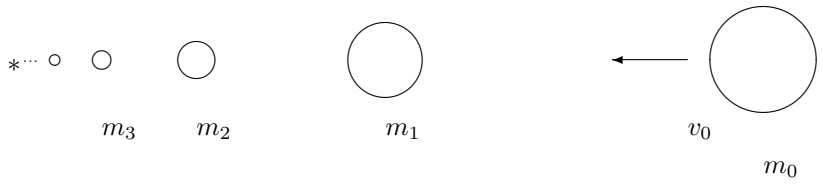


Figure 1. Collision of an infinite number of progressively smaller balls

Here one places a series of identical masses precisely in a row, and gives the first mass a velocity v_0 toward its nearest neighbor. After a sequence of collisions, the last mass in the row proceeds off with velocity v_0 in the direction of the initial motion, with all other masses being at rest. This demonstration is a classic and concise demonstration of energy and momentum conservation in classical mechanics. However, what happens if there is no "last mass"? That is, what occurs if there is an infinite sequence of masses in the row. Do the energy and the momentum simply disappear into the continuum of mass?

[3]It is an interesting coincidence that the mathematician Anatole Beck, who attended Peter Stuyvesant High School in New York City at the same time as Tom Erber, has written an analysis of the related problem of the hare and the (fixed) tortoise. He made the point that the standard specification of the velocity of the hare at a discrete set of locations that accumulate at the location of the tortoise is not sufficient to guarantee that the hare actually catches the tortoise. Furthermore, the result is still correct if the position of the hare is continuous and infinitely differentible. For details see Reference [4].

Within a few heartbeats, a physicist might point out that such a system is "unphysical", inasmuch as it would require an infinite amount of mass, and that it would be impossible to align the masses so as to maintain collinearity of the collisional process, without introducing friction or other dissipative effects. We will set aside the second objection for the present, since within the context of classical mechanics there is no intrinsic level of uncertainty or misalignment implicit in the formalism, and that it should be only a matter of sufficient care and cleverness to achieve a given level of precision. Does the requirement of a finite total mass in the system fix the problem? Remarkably, the answer is NO!

Let us consider an infinite sequence of masses $(m_0, m_1, m_2, \cdots)$ aligned in order along a line, as shown. Suppose further that the mass m_0 is given a velocity v_0 toward the first mass m_1, which is at rest, along with all the other masses. After the collision the mass m_0 leaves with velocity V_0, and the mass m_1 goes forward with velocity v_1. The mass m_1 strikes the mass m_2, leaving with velocity V_1, the mass m_2 going forward with velocity v_2. This collisional round continues until all subsequent balls have been undergone two collisions. The requirements of energy and momentum conservation in this round of collisions are

$$\begin{aligned} m_k\, v_k &= m_k\, V_k + m_{k+1}\, v_{k+1} \\ \tfrac{1}{2} m_k\, v_k^2 &= \tfrac{1}{2} m_k\, V_k^2 + \tfrac{1}{2} m_{k+1}\, v_{k+1}^2 \end{aligned}$$

where $k = 0, 1, \cdots$. Equivalently, we have

$$\begin{aligned} v_k &= \frac{m_{k+1} + m_k}{2m_k}\, v_{k+1} \\ V_k &= v_{k+1} - v_k = \frac{m_k - m_{k+1}}{m_k + m_{k+1}}\, v_k \end{aligned}$$

Let us write these relations as

$$\begin{aligned} v_{n+1} &= \frac{2}{1+\mu_n}\, v_n \\ V_n &= \frac{1-\mu_n}{1+\mu_n}\, v_n \\ \text{where} \qquad \mu_k &= \frac{m_{k+1}}{m_k} \end{aligned}$$

Rather complete discussions of these collisional sequences are given in References [5] -[7]. Here we shall draw special attention to the case in which all particles

recoil with a common speed; i. e., $V_n = V$ for all n. When this occurs, the particles move in lock step after the first round of collisions, and there are no subsequent collisions. In terms of the parameters

$$\alpha_n = \frac{1+\mu_n}{1-\mu_n}$$

the requirement of a constant recoil speed V may be written as

$$\begin{aligned} v_{n+1} &= \frac{2}{1+\mu_n}\, v_n \\ \alpha_{n+1}\, V &= \frac{2}{1+\mu_n}\, \frac{1+\mu_n}{1-\mu_n}\, V \\ \alpha_{n+1} &= \frac{2}{1-\mu_n} = 1 + \frac{1+\mu_n}{1-\mu_n} \\ \alpha_{n+1} &= \alpha_n + 1 \end{aligned}$$

There is a one-parameter family of solutions to this latter recursive formula for α_n:

$$\alpha_n = \lambda + n$$

where the parameter $\lambda > 1$ is otherwise arbitrary. The intermediate velocities are

$$v_n = \alpha_n\, V = (\lambda + n)\, V$$

and the parameter λ may determined from the original velocity of the incident ball:

$$\lambda\, V = v_0$$

The corresponding mass ratios are

$$\mu_n = \frac{\alpha_n - 1}{\alpha_n + 1} = \frac{\lambda + n - 1}{\lambda + n + 1}$$

We may determine the masses themselves:

$$\frac{m_n}{m_0} = \prod_{k=0}^{n-1} \mu_k = \frac{\lambda\,(\lambda - 1)}{(\lambda + n)\,(\lambda + n - 1)}$$

Therefore, the total mass of all the balls is

$$M = \sum_{n=0}^{\infty} m_n = \lambda\, m_0$$

The initial momentum and kinetic energy are given by

$$\begin{aligned} P_i &= m_0\, v_0 \\ 2\,T_i &= m_0\, v_0^2 \end{aligned}$$

and the final values are

$$\begin{aligned} P_f &= M\,V = m_0\, v_0 \\ 2\,T_f &= M\,u^2 = \frac{m\, v_0^2}{\lambda} \end{aligned}$$

For this case, as well as for a wide variety of other cases, momentum is conserved whereas energy has been lost. The intermediate velocities v_n increase with n, and in the limit the intermediate kinetic energy approaches a non-zero limit:

$$2\,T_n^{\,\rm lost} = m_n\, V_n^2 = m_0\, u^2 \frac{\lambda\,(\lambda-1)}{(\lambda+n)\,(\lambda+n-1)}\,(\lambda+n)^2 \to m_0\, v_0^2 \left(1-\frac{1}{\lambda}\right)$$

This amount of energy disappears into the continuum in the process.

The case of a common recoil speed can be compared with the completely inelastic collision of a particle of mass m_0 and initial speed v_0 with a solid body of mass $M-m$ that is initially at rest. The two masses coalesce during the collisions, and afterward the entire system of mass M moves with speed V. For the collision the momentum is conserved:

$$m_0\, v_0 = M\, u$$

whereas this amount of energy is lost – presumably converted into "heat".

$$2\,T^{\,\rm lost} = m_0\, v_0^2 - M\, u^2 = m_0\, v_0^2 \left(1-\frac{1}{\lambda}\right)$$

This energy is converted into "heat" in this inelastic process.

The original collision sequence could be interpreted as a microscopic rendition of this corresponding inelastic collision, without the requirement of "binding forces" to keep the final mass intact. However, it is a mystery as to how an elastic collisional sequence could possibly mimic an inelastic process.

Note also that the intermediate speeds $(\lambda+n)\,v$ become arbitrarily large at large n, and this fact might suggest that our result might be a spurious nonrelativistic limit of an more appropriate relativistic process. We will discuss the relativistic version in the next section.

8.3 Relativistic Collisions

The relativistic equations reflecting energy and momentum conservation for the nth collision (with $c = 1$) are

$$\begin{aligned} \gamma(v_n)\, m_n &= \gamma(V_n)\, m_n + \gamma(v_{n+1})\, m_{n+1} \\ \gamma(v_n)\, m_n\, v_n &= \gamma(V_n)\, m_n\, V_n + \gamma(v_{n+1})\, m_{n+1}\, v_{n+1} \end{aligned}$$

where $\gamma(w) = 1/\sqrt{1-w^2}$ is the Lorentz factor. By adding and subtracting these equations we may obtain

$$\begin{aligned} \epsilon(V_n) - \epsilon(v_n) &= \mu_n\,(1 - \epsilon(v_{n+1})) \\ \epsilon^{-1}(V_n) - \epsilon^{-1}(v_n) &= \mu_n\,(1 - \epsilon^{-1}(v_{n+1})) \end{aligned}$$

where $\epsilon(w) = (1-w)\,\gamma(w) = \sqrt{(1-w)/(1+w)}$. Equivalently, we have

$$\epsilon(V_n) = \frac{\epsilon(v_{n+1})}{\epsilon(v_n)} = \frac{\mu_n + \epsilon(v_n)}{1 + \mu_n\,\epsilon(v_n)}$$

For the relativistic version of the constant recoil process, the final speeds V_n have a common value V, so that $\epsilon(V_n) = \epsilon(V) = \eta$ for all n. We may then solve recursively to obtain

$$\epsilon(v_n) = \eta^n\,\epsilon(v_0)$$

The corresponding mass ratios we obtain

$$\mu_n = \frac{\eta - \epsilon(v_n)}{1 - \eta\,\epsilon(v_n)} = \frac{\eta - \eta^n\,\epsilon(v_0)}{1 - \eta^{n+1}\,u(v_0)}$$

The masses themselves are

$$m_n = m_0\,\eta^n\,\frac{(1-\epsilon(v_0))\;(\eta - \epsilon(v_0))}{(1-\eta^n\,\epsilon(v_0))\;(\eta - \eta^n\,\epsilon(v_0))}$$

and the total mass is

$$M = \frac{1-\epsilon(v_0)}{1-\eta}$$

The quantities $\epsilon(v_n)$ asymptotically approach zero, so that the corresponding speeds v_n approach the speed of light. The limiting energy of the nth ball after the first collision is

$$E_n = m_n\,\gamma(v_n) = \frac{m_n}{2}\left(\epsilon(v_n) + \frac{1}{\epsilon(v_n)}\right) \to \frac{m_0}{2\,\epsilon(v_0)\,\eta}\,(\eta - \epsilon(v_0))\;(1 - \epsilon(v_0)) = E'$$

This quantity is also equal to the asymptotic value of the momentum at large n, so that both energy and momentum are lost in this collisional sequence.

It is instructive to compare this process with one in which a mass m_0 with initial speed v_0 collides with a solid object of mass $M-m$, initially at rest. After the collision sequence the mass M that recoils with speed V, and also a massless particle of energy E' – such as a photon – is produced. The requirements of energy and momentum conservation are

$$\begin{aligned} m_0\,\gamma(v_0) + (M - m_0) &= M\,\gamma(V) + E' \\ m_0\,\gamma(v_0)\,v_0 &= M\,\gamma(V)\,V + E' \end{aligned}$$

Equivalently,

$$\begin{aligned} m_0\,\epsilon(v_0) + (M - m_0) &= M\,\epsilon(V) \\ m_0\,\epsilon^{-1}(v_0) + (M - m_0) &= M\,\epsilon^{-1}(V) + 2E' \end{aligned}$$

The first equation, written in terms of $\eta = \epsilon(V)$, is

$$M = m_0\,\frac{1-\epsilon(v_0)}{1-\eta}$$

whereas the second relation gives

$$E' = \frac{m_0}{2\,\epsilon(v_0)\,\eta}\,(\eta - \epsilon(v_0))\,(1-\epsilon(v_0))$$

Thus, this process has the same kinematical form as the relativistic collisional sequence.

In summary, the elastic relativistic collisional sequences with a constant velocity of recoil are microscopic renditions of single collisions in which a massless particle is produced.

8.4 Classical Statistical Mechanics

Suppose that an infinite set of masses $\{m_n\}$, with finite total mass M, as in the previous two sections, is confined by walls to a one-dimensional region of finite length. Suppose, as before, that initially the heaviest particle, with mass m_0, has speed v_0, the rest of the particles being at rest. Let the particles subsequently collide elastically with one another, and subsequently at the walls. This seems like a simple extension of what we have done above, but it is not.

To see this, given the case of constant recoil, as in Sect. 1.2 (nonrelativistic), or Sect. 1.3 (relativistic), let us consider the state of the system after all the balls have collided with one another, but before any of them have reached a wall. They are all moving with the same speed towards the left, and will continue to do so until they hit the leftmost wall. Or will they? Which ball will strike the wall first? No ball can do so, for if ball number n were to hit the wall, its neighbor to its left should have struck the wall first, and this applies to any ball at all. The inconsistency is identical to the one discussed by Alper and Bridger [8], in which an additional ball approaches the point of accumulation of an infinite set of balls that are all initially at rest. Indeed, by looking at the constant-recoil scenario from a co-moving frame of reference, we see a wall moving to the right, approaching the point of accumulation of the positions of an infinite set of stationary balls: exactly the Alper-Bridger case. In Ref. [9] we have advocated a radical way around the impasse, namely that of embracing Aristotle's "potential infinity", in contrast to "actual infinity". The distinction is that between the limit of a finite system, as the size grows without bound, and an infinite system ab initio . A merely potentially infinite set of balls can collide with a wall, and moreover energy and momentum are both conserved throughout all collisional processes. We shall accordingly consider the statistical mechanics in one dimension of a finite, but large number of balls, with a view to inferring what happens in the limit of an infinite system.

Let us assume that the non-relativistic collisional system evolves into a condition of statistical equilibrium. In equilibrium, we suppose that it would be equally likely for a particle to lie anywhere within the region of phase space corresponding to a given total energy. Let us consider a system of N non-interacting non-relativistic particles with masses

$$m_n = m_0/2^n \qquad \text{where} \qquad n = 0, \ldots, N-1$$

which is confined along a line of length L. The amount of phase space of this system for energy E is

$$\Omega(E) = \int dx_0 \ldots \int dx_{N-1} \int dp_0 \ldots \int dp_{N-1}\, \delta\left(E - E_0 - \ldots - E_{N-1}\right)$$

where the single particle energies are

$$E_n = \frac{p_n^2}{2\, m_n}$$

For the case of impenetrable particles along the line, the conditions

$$0 \le x_0 \le x_1 \le \ldots \le x_{N-1} \le L$$

must be maintained. Thus the volume integral is

$$\int_0^L dx_0 \int_{x_0}^L dx_1 \ \ldots \int_{x_N-2}^L dx_{N-1} = \frac{L^N}{N!}$$

When the system achieves thermal equilibrium at an (absolute) temperature T, the quantity $\Omega(E)$ must be multiplied by the Boltzmann factor $\exp(-\beta E)$ to determine the occupation probabilities. The partition function is

$$\begin{aligned} Z(\beta, N, m_n) &= \int_0^\infty dE\, \Omega(E) e^{-\beta\, E} = \frac{L^N}{N!} \prod_{n=0}^{N-1} f(\beta, m_n) \\ \text{where} \qquad f(\beta, m_n) &= \int_{-\infty}^\infty dp_n\, \exp(-\beta\, p_n^2/(2\, m_n)) = \sqrt{\frac{2\,\pi\, m_n}{\beta}} \end{aligned}$$

Thus we obtain

$$Z(\beta, N, m_n) = \frac{L^N}{N!} \left(\frac{2\pi}{\beta}\right)^{N/2} \left(\prod_{n=0}^{N-1} m_n\right)^{1/2}$$

The average single particle energies are

$$\overline{E}_n = -\frac{\partial}{\partial \beta} \log f(\beta, m_n) = \frac{1}{2\,\beta} = \frac{k_B\, T}{2}$$

and the total energy is

$$E = -\frac{\partial}{\partial \beta} \log Z = \frac{N}{2\,\beta} = \frac{N\, k_B\, T}{2}$$

The energy is partitioned equally among all the particles – The Equipartition Theorem. In particular, the heaviest particle, with mass m_0, would have an average speed of

$$\overline{u}_0 = \sqrt{\frac{2\, E}{N m_0}}$$

whereas the lightest particle, of mass $m_0/2^{N-1}$, would have an average speed of

$$\overline{u}_{N-1} = \sqrt{\frac{2^N\, E}{N m_0}}$$

For the case of "potential infinity" (i. e., the limit $N \to \infty$ with total energy fixed) each particle would have a negligible amount of energy, with the heavier ones moving slowly and the lighter ones moving rapidly. This system in equilibrium would produce an average force

$$F = \frac{1}{\beta} \frac{\partial}{\partial L} \log Z = \frac{N}{\beta L} = \frac{N k_B T}{L}$$

on the immovable walls at each end of the containment region. This relation is the one-dimensional analogue of the Ideal Gas Law. The force of containment is finite, but the only possible equilibrium temperature is absolute zero.

Whether or not statistical equilibrium ever arises, we still expect the lighter particles to have more substantial speeds, with the heavier ones moving more slowly over the course of time, as a result of the approach to equilibrium. In a preliminary numerical study of one-dimensional collisions a finite number of distinguishable, impenetrable balls, we obtain results that are consistent with this viewpoint. Namely, when the initial kinetic energy of the balls is fixed, and the number of balls is increased, the energy is seen to be shared with all the balls, when averaged over a sufficiently large time interval.

The relativistic case is rather similar in character. The amount of Lorentz invariant relativistic phase space for total energy E is

$$\Omega(E) = \int dx_0 \, \ldots \int dx_{N-1} \int \frac{dp_0}{E_0} \, \ldots \int \frac{dp_{N-1}}{E_{N-1}} \, \delta \left(E - E_0 - \ldots - E_{N-1} \right)$$

where the single particle energies are

$$E_n = \sqrt{p_n^2 + m_n^2}$$

The corresponding partition function is

$$Z(\beta) = \frac{L^N}{N!} \prod_{n=0}^{N-1} z(\beta, m_n)$$

where

$$\begin{aligned} z(\beta, m_n) &= \int_{-\infty}^{\infty} \frac{dp_n}{\sqrt{p_n^2 + m_n^2}} \exp\left(-\beta \sqrt{p^2 + m_n^2} \right) \\ &= 2 \int_0^{\infty} dt \, \exp(-\beta m_n \cosh t) = 2 K_0(\beta m_n) \end{aligned}$$

where K_0 is the modified Hankel (Kelvin) function of order 0. For asymptotically limiting cases we obtain

$$z(\beta, m_n) = \begin{cases} 2\,\log 2/(\beta\, m_n) & \beta\, m_n \ll 1 \\ \sqrt{2\,\pi/(\beta\, m_n)}\,\exp(-\beta\, m_n) & \beta m_n \gg 1 \end{cases}$$

The average single particle energies are

$$\overline{E}_n = -\frac{\partial}{\partial\beta}\,\log z(\beta, m_n) = \begin{cases} 1/\beta & \beta\, m_n \ll 1 \\ m_n + 1/(2\,\beta) & \beta\, m_n \gg 1 \end{cases}$$

The kinetic energies of the particles approach $k_B T/2$ when $k_B T \ll m_n$, whereas when $k_B T \gg m_n$ they approach $k_B T$. For intermediate masses they lie between these limits. The full partition function is

$$Z(\beta) = \frac{(2L)^N}{N!}\,\prod_{n=0}^{N-1} K_0(\beta\, m_n)$$

and the total energy is

$$E = -\frac{\partial}{\partial\beta}\,\log Z = \sum_{n=0}^{N-1} E_n$$

Thermal equilibrium can occur only at absolute zero temperature, as was true for the non-relativistic case.

The force at the ends of the region of containment are

$$F = \frac{1}{\beta}\,\frac{\partial}{\partial L}\,\log Z = \frac{N}{\beta\, L} = \frac{N\, k_B\, T}{L}$$

This one-dimensional version of the perfect gas law[4] remains valid for a finite number of particles with an arbitrary distribution of masses. It remains valid even for an infinite number of particles.

We find that, at thermal equilibrium, the mechanical energy of the one dimensional closed systems remains constant over the course of time. However, whatever the dimension, when the number of particles N is increased with a fixed amount of mechanical energy, the equilibrium temperature T decreases. In the

[4]The three-dimensional version, $P = N\, k_B\, T/V$ is also correct for relativistic systems.

limit $N \to \infty$ that temperature must approach absolute zero. Consequently, the energy is still present inside the one-dimensional box, but it cannot be recovered by thermal contact with a body at finite absolute temperature.

In Sections 1.2 and 1.3, we encountered violation of the first law of thermodynamics (that is, the law of conservation of energy) with an actual infinity of balls. In the present section we considered a potential infinity of balls (i.e., the limit of the energy of N balls, as $N \to \infty$), but they are confined to a finite interval. Although the first law of thermodynamics is observed, the second law of thermodynamics ensures that the energy is irrecoverable. In the next section we shall see that quantum mechanical considerations do not alleviate this situation.

8.5 Quantum Statistical Mechanics

Over a century ago, it was pointed out by Rayleigh and Jeans that the Planck quantum hypothesis for the description of electromagnetic radiation within a cavity was deficient. In effect, the thermal energy inside would have to be shared among the infinite set of modes of electromagnetic radiation in the cavity – leading to an impossibility. It was subsequently realized that the higher frequency modes of cavity radiation had simply been "frozen out" in the Planck formula, as a result of quantum effects. Furthermore, a classical description was inadequate, and an explanation involving the quantum theory was required.

Surely, a quantum mechanical description is also required to analyze the interaction of these systems, which contain an infinite number of particles. In the quantum mechanical formulation, we might like to drop the notion of impenetrability of the masses. Furthermore, we would be able to analyze individual collisions of localized mass states for a very short time interval, and would achieve no insight as to the long-term behavior of the system. Therefore, we will consider the situation in which the masses are confined to a one-dimensional region, and also are in a state of thermal equilibrium.

For N (distinguishable) particles of individual mass m_n, each moving freely in the region $[0, L]$, the (non-relativistic) Hamiltonian is

$$H = \sum_{n=0}^{N-1} H_n = \sum_{n=0}^{N-1} \frac{p_n^2}{2\, m_n}$$

The single-particle eigenstates for the nth particle are

$$\begin{aligned} \psi_n^{\{k\}}(x) &= \sqrt{\frac{2}{L}} \sin \frac{k\,\pi\,x_n}{L} \\ H_n\,\psi_n^{\{k\}}(x) &= E_n^{\{k\}}\,\psi_n^{\{k\}}(x) \\ E_n^{\{k\}} &= \frac{k^2\,h^2}{8\,m_n\,L^2} \end{aligned}$$

where h is Planck's constant and $k = 1, 2, \cdots$. We use these eigenstates as an orthonormal basis in calculating the quantum mechanical partition function:

$$Z^{QM}(\beta, N, m_n) = \text{Trace } \exp(-\beta\,H) \;=\; \sum_{n=0}^{N-1} f(\beta, m_n)$$

where the single particle traces are

$$f(\beta, m_n) = \text{Trace } \exp(-\beta\,H_n) \;=\; \sum_{k=1}^{\infty} \exp(-\beta\,E_n^{\{k\}})$$

For the case $\beta\,E_n^{\{1\}} \ll 1$, many lower-lying energy eigenstates are likely to be populated. Roughly speaking, the eigenstates $E_n^{\{k\}}$ are populated with significant probability when $k = 1, \ldots, K$, where

$$\beta\,E_n^{\{K\}} \approx 1$$

Consequently

$$f(\beta, m_n) \approx K \approx \sqrt{\frac{8\,m_n\,L^2}{\beta\,h^2}}$$

For the other extreme case, $\beta\,E_n^{\{1\}} \gg 1$, only the state $k = 1$ is occupied with significant probability, and we obtain

$$f(\beta, m_m) \approx \exp\left(-\beta E_n^{\{1\}}\right) = \exp\left(\frac{-\beta\,h^2}{8\,m_n\,L^2}\right)$$

The average energy for the nth particle is

$$< E_n >= -\frac{\partial}{\partial\beta}\;\log f(\beta, m_n) = \begin{cases} 1/(2\,\beta) = k_B\,T/2 & \beta\,E_n^{\{1\}} \ll 1 \\ E_n^{\{1\}} = h^2/(8\,m\,L^2) & \beta\,E_n^{\{1\}} \gg 1 \end{cases}$$

and the average force induced at the walls by this particle is

$$< F_n >= \frac{1}{\beta L}\,\frac{\partial}{\partial L}\,\log f(\beta, m_n) = \begin{cases} 1/(\beta) & \beta\, E_n^{\{1\}} \ll 1 \\ \\ h^2/(4\, m_n\, L^3) & \beta\, E_n^{\{1\}} \gg 1 \end{cases}$$

In both cases note that $< F_n >= 2 < E_n > /L$. The average energy and force on the walls for all N particles are additive:

$$\begin{aligned} < E > &= \sum_{n=0}^{N-1} < E_n > \\ < F > &= \sum_{n=0}^{N-1} < F_n > \end{aligned}$$

So long as the condition

$$\frac{h^2}{8\, m_n L^2} \ll k_B\, T$$

is met, the total energy and force on the walls are

$$< E >= \frac{N\, k_B T}{2} \qquad \text{and} \qquad < F >= \frac{N\, k_B T}{L}$$

in accordance with the results of classical statistical mechanics.

However, when this condition is not met, only a few low-lying states are significantly populated, and the equilibrium state of the system differs significantly from that of its classical counterpart. In this circumstance, the energy and force on the walls become significantly larger than they are in the classical system. In the extreme case

$$\frac{h^2}{8\, m_n L^2} \gg k_B\, T$$

the mass m_n lies predominantly in its ground state, $k = 1$.

The problem of particles in a box stands in strong contrast to that of cavity radiation, as considered by Planck. There are an infinite number of modes of electromagnetic radiation inside the cavity, with corresponding frequencies ν_n. When the walls of the cavity are held at temperature T, each mode for which

$$h\, \nu_n \ll k_B\, T$$

contains a large number of quanta k_n of electromagnetic radiation – photons. On the other hand, for

$$h\,\nu_n \gg k_B\,T$$

the modes contain very few photons, if any, on average. In fact the very high frequency modes are completely unpopulated – they do not participate in the dynamics. The fact that the occupation number k_n can be zero for modes of cavity radiation of high frequency ν_n is responsible for the reduction of their contribution to the energy and force on the walls. This fact is responsible for the difference in quantum effects for the two systems.

In summary, the difficulty in the statistical treatment of the system of Zeno balls persists in the quantum theory. Indeed, quantum effects serve to increase the agitation of light particles that are confined.

References

[1] Pérez Laraudogoitia, J.: A Beautiful Supertask. Mind **105**, 81-83 (1996).

[2] Earman, J. and Norton, J.D.: Comments on Laraudogoitia's 'Classical Particle Dynamics, Indeterminism and a Supertask'. Brit. Jour. Phil. Sci. **49**, 123-33 (1998).

[3] Pérez Laraudogoitia, J.: Earman and Norton on Supertasks that Generate Indeterminism. Brit. Jour. Phil. Sci. **50**, 137-41 (1999).

[4] Beck, A.: Hase und Schildkröte (Deutsch), Selecta Mathematica V, Band 201, Heidelberger Taschenbücher, Springer-Verlag, 1-230 (1979).
The Hare and the Tortoise, and other paradoxes, by Anatole Beck, to be published.

[5] Atkinson, D.: Losing Energy in Classical, Relativistic and Quantum Mechanics. Stud. Hist. Phil. Mod. Phys. **38**, 170-180 (2007).

[6] Atkinson, D.: A Relativistic Zeno Process. Synthese **160**, 5-12 (2008).

[7] Atkinson, D., and Johnson, P. W.: Nonconservation of Energy and Loss of Determinism I. Infinitely Many Colliding Balls, Found. Physics **39**, 937-957 (2009).

[8] Alper, J.S., Bridger, M.: Newtonian Supertasks: A Critical Analysis. Synthese **114**, 335-369 (1998).

[9] Atkinson, D. and Johnson, P. W.: Nonconservation of Energy and Loss of Determinism II. Colliding with an Open Set, Found. Physics **40**, 179 (2010).

9

The Tao Solution for the Stefan Problem: An Alternate Derivation

S. Nair[1]
Illinois Institute of Technology[2]

Abstract

The problem of a freezing front moving into a liquid in a one dimensional semi-infinite domain with arbitrary initial and boundary conditions was solved by Tao in 1978 using the heat polynomials and the repeated integrals of the complementary error function. As the field equations in the frozen and liquid domains are linear, solutions may be constructed using the Laplace transform. The problem is essentially nonlinear due to the interface conditions. Here, an alternate derivation using this approach is provided, which demonstrates certain symmetry between the temperature distributions in the frozen and liquid domains. A Laplace inversion integral operator is used in this derivation to obtain the repeated integrals of the complementary error function and their derivatives of arbitrary orders in their respective power series, which play a crucial role in the Tao solution.

The author is indebted to Professor Tom Erber for inspiring discussions on various subjects of mutual interest.

[1]Professor of Mechanical Engineering and Professor of Applied Mathematics
[2]Department of Mechanical, Materials and Aerospace Engineering

9.1 Introduction

Moving or free boundary problems arise naturally in many physical phenomena. A classical example is the freezing of water (or melting of ice) where the interface separating the two phases is a moving boundary. The presence of such interfaces renders otherwise linear heat conduction problems nonlinear. The freezing problem was first considered by Neumann in his unpublished lecture notes (*circa* 1860). He found exact solutions for a frozen front moving into a semi-infinite domain due to an applied sub-zero, constant temperature at the origin, using a similarity variable. On the basis of a special case of the Neumann solution published by Stefan previously, this class of problems are known as *Stefan problems* in the literature (Carslaw and Jaeger [1]). Analytical solutions for the semi-infinite domain problem with arbitrary initial and boundary conditions were unavailable for the Stefan problem for over a hundred years until the series solutions due to Tao [2–7].

Meanwhile, a host of approximate methods have been proposed to solve the one dimensional Stefan problem. The approximate methods may be classified as analytic, numerical or mixed. The solutions using the perturbation method, due to Weinbaum and Jiji [8] and Charach and Zoglin [9] fall under the first group. As in all perturbation solutions the need for a small parameter limits the applicability of these solutions. Finite difference and finite element methods form the bulk of numerical methods. Detailed appraisals of these methods have been published by Rubinstein [10], Ockendon and Hodgkins [11], Wilson, Solomon, and Boggs [12] and Fukusako and Seki [13]. Under mixed methods falls the use of a sequence of global approximating functions for the temperature or enthalpy distribution. The heat balance integral method due to Goodman [14] is an example of this. A mixed method due to Dursunkaya and Nair [15], using spectral methods on a transformed problem gives a simple system of ordinary differential equations as an approximation for certain ranges of parameters. Another class of mixed methods stem from the integral equation formulation of moving boundary problems by Kolodner [16], Boley [17], Nair [18] and Tao [19]. The formulations by Kolodner and Tao employ the Green's functions of moving heat sources at the interface while Boley and Nair use overlapping extended frozen and liquid domains with variable boundary temperatures to satisfy the interface conditions. An extensive bibliography on the moving boundary problem can be found in a book by Crank [20].

In what follows, we provide an alternate derivation of the Tao [2] solution using the Laplace transform formulism. The present derivation reveals certain symmetry in the temperature distributions in the frozen and liquid regions.

Moreover, the approach presented here appears to be more convenient than that employed in the original solution. A number of algebraic complexities can be avoided by using the present approach. The repeated integrals of the complementary error function and their derivatives of arbitrary order come out in series form. We also avoid the formula for the multinomial coefficients (see Abramowitz and Stegun [21]) through the use of a recurrence relation. The present formulation may be of use for numerical implementation and extensions to other problems.

9.2 Formulation

We consider a liquid with freezing point T_0, occupying the region $0 < x < \infty$ with the initial temperature $\bar{W}(x) > T_0$. At time $t = 0$, the boundary $x = 0$ is subjected to a temperature $U(t) < T_0$. Denoting the liquid properties by symbols with a bar over them and the solid properties by plain symbols, we have

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}, \qquad 0 < x < s(t), \tag{9.1}$$

$$\bar{\alpha} \frac{\partial^2 \bar{\theta}}{\partial x^2} = \frac{\partial \bar{\theta}}{\partial t}, \qquad s(t) < x < \infty, \tag{9.2}$$

where α and $\bar{\alpha}$ represent the diffusivities and $s(t)$ is the solid-liquid interface location.

The boundary and initial conditions are

$$T(0,t) = U(t), \qquad \bar{\theta}(x,0) = \bar{W}(x). \tag{9.3}$$

For this freezing problem the interface conditions are

$$T(s,t) = \bar{T}(s,t) = T_0, \qquad \left[k \frac{\partial T}{\partial x} - \bar{k} \frac{\partial \bar{T}}{\partial x} \right]_{x=s} = \rho L \frac{ds}{dt}, \tag{9.4}$$

where k and $\bar{k}$ are the conductivities, ρ is the density and L is the latent heat.

The temperatures may be scaled by introducing

$$\theta = T/T_0 \qquad \bar{\theta} = \bar{T}/T_0,$$
$$u(t) = U(t)/T_0, \qquad \bar{w}(x) = \bar{W}(x)/T_0 \tag{9.5}$$

The resulting equations are

$$\frac{\partial^2 \theta}{\partial x^2} = \frac{\partial \theta}{\partial t}, \qquad \frac{\partial^2 \bar{\theta}}{\partial x^2} = \frac{\partial \bar{\theta}}{\partial t}, \tag{9.6}$$

$$\theta(0,t) = u(t), \qquad \bar{\theta}(x,0) = \bar{w}(x), \tag{9.7}$$

$$\theta(s,t) = \bar{\theta}(s,t) = 1, \qquad \left[S\,\frac{\partial\theta}{\partial x} - \bar{S}\frac{\partial\bar{\theta}}{\partial x}\right]_{x=s} = \frac{ds}{dt}, \tag{9.8}$$

where

$$S = \frac{k\,T_0}{\rho\,L}, \qquad \bar{S} = \frac{\bar{k}\,T_0}{\rho\,L}. \tag{9.9}$$

It is useful to consider certain properties of the Laplace transform before we attempt a solution of the above equations. We define

$$f(p) = \mathcal{L}[f(t), t \to p] = \int_0^\infty f(t)\,e^{-p\,t}\,dt, \tag{9.10}$$

with its inverse

$$\begin{aligned} f(t) &= \frac{1}{2\pi i}\int_C f(p)e^{pt}\,dp = \frac{1}{2\pi i}\int_C (p/t)f(p/t)e^{p}\,dp/p \\ &= \mathcal{I}\left[(p/t)f(p/t)\right], \end{aligned} \tag{9.11}$$

where the integral operator $\mathcal{I}$, which is independent of time t and its operand are of use in extracting the time dependence explicitly from the Laplace transforms, in particular, when the transforms are in the form of power series in p or $\sqrt{p}$. The inverse integral is performed over the Bromwich contour C.

The Tao solution [2] involves a representation of the solution in terms of the heat polynomials and the repeated integrals of the complementary error function. We note that the complementary error function has the series representation (Abramowitz and Stegun, [21]),

$$\operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(1-k/2)}\left(\frac{x}{\sqrt{t}}\right)^k, \tag{9.12}$$

when $x/\sqrt{t}$ is finite. As this series is absolutely convergent, by term-by-term integration, the Laplace transform is obtained as,

$$\mathcal{L}\left[\operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right)\right] = \frac{1}{p}\,e^{-\sqrt{p}x} = \frac{1}{p}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}(x\,\sqrt{p})^k. \tag{9.13}$$

When, in our calculations, evaluation of the inverse of $\sqrt{p}$- terms are needed, evaluating through integration in the complex p-domain,

$$\mathcal{I}\left[\left(\frac{p}{t}\right)^{(k/2-1)}\right] = \frac{1}{t^{k/2}\,\Gamma(1-k/2)}. \tag{9.14}$$

Applying this to the series in (9.13), we recover the series representation (9.12), as long as $x/\sqrt{t}$ is finite. We caution that in general Laplace transforms $f(p)$ which are unbounded as $Re(p) \to \infty$ do not have an inverse. If $(\sqrt{p})^k$ where $k > 0$ are interpreted as belonging to the derivative of the error function (9.12) at $x = 0$, the term-by-term inversion gives the correct result and serves as a useful operational approach.

In the Laplace transform domain the differential equations (9.6) become

$$\frac{\partial^2 \theta}{\partial x^2} = p\,\theta - w(x), \qquad \frac{\partial^2 \bar{\theta}}{\partial x^2} = p\,\bar{\theta} - \bar{w}(x), \tag{9.15}$$

where $w(x)$ is an unknown (fictitious) initial condition for the frozen domain. The solutions of these can be readily written as

$$\theta = a(p)e^{-qx} + \frac{1}{p}(1 - D^2/q^2)^{-1} w(x), \qquad q^2 = p/\alpha, \tag{9.16}$$

$$\bar{\theta} = \bar{a}(p)\, e^{-\bar{q}x} + \frac{1}{p}(1 - D^2/\bar{q}^2)^{-1} \bar{w}(x), \qquad \bar{q}^2 = p/\bar{\alpha}, \tag{9.17}$$

where a and $\bar{a}$ are arbitrary functions of p, and D is the differential operator $\partial/\partial x$.

Using the initial condition on θ,

$$a(p) = u(p) - \frac{1}{p}\,(1 - D^2/q^2)^{-1}\, w(0), \tag{9.18}$$

where $D^n w(0)$ denotes $D^n w(x)$ evaluated at $x = 0$.

Following Tao [2], we assume the boundary condition at $x = 0$ to be analytic in $\sqrt{t}$ and the initial condition analytic in x. In series representations

$$\begin{aligned} u(t) &= \sum_{n=0} u_n (\alpha t)^{n/2}/\Gamma(1 + n/2), \\ \{u(p), a(p), \bar{a}(p)\} &= \frac{1}{p}\sum_{n=0} \{u_n, a_n, \bar{a}_n\}/q^n, \\ \{w(x), \bar{w}(x)\} &= \sum_{n=0} \{w_n, \bar{w}_n\} x^n/n!, \end{aligned} \tag{9.19}$$

$$(1 - D^2/q^2)^{-1}\{w(x), w(0)\} = \sum_{m=0,2} \left\{ \sum_{n=0} w_{n+m} x^n/n!, w_m \right\} /q^m. \tag{9.20}$$

From equations (9.18) and (9.19) we have

$$a_n = u_n - w_n\, e_n \tag{9.21}$$

where e_n is 1 when n is 0 or even and it is 0 when n is odd.

Using the inversion operator with explicit time, and $\tau = \sqrt{t}$, the solutions become

$$\theta = \mathcal{I}\left[\sum_{n=0} a_n \tau^n e^{-qx/\tau}/q^n + \sum_{n=1}\sum_{m=0,2} w_{n+m} x^n q^{-m} \tau^m/n!\right] \tag{9.22}$$

and its spatial derivative

$$D\theta = \mathcal{I}\left[-\sum_{n=0} a_n \tau^{n-1} e^{-qx/\tau}/q^{n-1} + \sum_{n=1}\sum_{m=0,2} w_{n+m} x^{n-1} \tau^{m-1} q^{-m}/(n-1)!\right], \tag{9.23}$$

and the corresponding functions for the liquid domain are obtained by attaching the bar to q, a_n and w_n.

If x is replaced by $s(t)$, the interface conditions (9.8) can be expressed in terms of the above expressions. Thus,

$$\mathcal{I}\sum_{n=0}\left[a_n\, \tau^n\, e^{-qs/\tau}/q^n + \sum_{m=0,2} w_{n+m} s^n \tau^m q^{-m}/n!\right] = 1, \tag{9.24}$$

$$\mathcal{I}\sum_{n=0}\left[\bar{a}_n \tau^n e^{-\bar{q}s/\tau}/\bar{q}^n + \sum_{m=0,2} w_{n+m} s^n \tau^m \bar{q}^{-m}/n!\right] = 1, \tag{9.25}$$

$$S\mathcal{I}\left[-\sum_{n=0} a_n \tau^{n-1} e^{-qs/\tau}/q^{n-1} + \sum_{n=1}\sum_{m=0,2} w_{n+m} s^{n-1} \tau^{m-1} q^{-m}/(n-1)!\right]$$
$$-\bar{S}\mathcal{I}\left[-\sum_{n=0} \bar{a}_n \tau^{n-1} e^{-\bar{q}s/\tau}/\bar{q}^{n-1} + \sum_{n=1}\sum_{m=0,2} w_{n+m} s^{n-1} \tau^{m-1} \bar{q}^{-m}/(n-1)!\right]$$
$$= \frac{ds}{dt}. \tag{9.26}$$

Assuming

$$s/\tau = s_0 + s_1\tau + s_2\tau^2 + \cdots = \sum_{n=0} s_n \tau^n, \tag{9.27}$$

$$\tau \frac{ds}{dt} = \sum_{n=0}(n+1)s_n\tau^n, \qquad (s/\tau)^k = \sum_{n=0}\tau^n Z_n^{(k)}, \tag{9.28}$$

where $Z_n^{(k)}$ are related to the multinomial coefficients (see Tao, 1978). We may use the recurrence relations

$$Z_n^{(1)} = s_n, \qquad Z_n^{(m+1)} = \sum_{k=0}^{n} Z_n^{(m)} s_{n-k}, \tag{9.29}$$

to construct the full matrix of $Z_n^{(k)}$. We also note that $Z_n^{(k)} = 0$ whenever $n < 0$.

To isolate the various powers of τ we need the elementary result,

$$\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\tau^{n+k}A_{nk} = \sum_{n=0}^{\infty}\tau^n\sum_{k=0}^{n}A_{k(n-k)}. \tag{9.30}$$

Matching the coefficients of equal powers of τ and using the inversion relation (9.14), and the derivative of s from (9.28), equations (9.24) - (9.26), can be written as

$$\sum_{n=0}^{m}[A_{mn}\,a_n + B_{mn}\,w_n] = \delta_{m0}, \tag{9.31}$$

$$\sum_{n=0}^{m}[\bar{A}_{mn}\,\bar{a}_n + \bar{B}_{mn}\,\bar{w}_n] = \delta_{m0}, \tag{9.32}$$

$$S\sum_{n=0}^{m}[C_{mn}a_n + D_{mn}w_n] - \bar{S}\sum_{n=0}^{m}\left[\bar{C}_{mn}\bar{a}_n + \bar{D}_{mn}\bar{w}_n\right] = (m+1)s_m/2, \tag{9.33}$$

where δ_{m0} is the Kronecker delta, and with n_e representing the largest even integer less than or equal to n,

$$\begin{aligned}
A_{mn} &= \sum_{k=0,1}^{\infty}\frac{(-1)^k}{k!}\frac{(\sqrt{\alpha})^{n-k}Z_{m-n}^{(k)}}{\Gamma(1-k/2-n/2)},\\
B_{mn} &= \sum_{k=0,2}^{n_e}\frac{1}{(n-k)!}\frac{(\sqrt{\alpha})^{k}Z_{m-n}^{(n-k)}}{\Gamma(1+k/2)},
\end{aligned} \tag{9.34}$$

$$\begin{aligned}
C_{mn} &= \sum_{k=1,2}^{\infty}\frac{(-1)^k}{(k-1)!}\frac{(\sqrt{\alpha})^{n-k}Z_{m-n}^{(k-1)}}{\Gamma(1-k/2-n/2)},\\
D_{mn} &= \sum_{k=1,3}^{n_e}\frac{1}{(n-k-1)!}\frac{(\sqrt{\alpha})^{k}Z_{m-n}^{(n-k)}}{\Gamma(1+k/2)},
\end{aligned} \tag{9.35}$$

and $\bar{A}_{mn}, \bar{B}_{mn}, \bar{C}_{mn}$,and $\bar{D}_{mn}$ are obtained from the above by replacing α by $\bar{\alpha}$.

We note that the system of equations ((9.31)-(9.33)) are, as expected, triangular and it can be solved sequentially. In equation (9.31), when a_n are expressed in terms of the boundary and initial condition coefficients u_n and w_n using (9.21), equation (9.31) can be solved for w_n in terms of u_n and s_n. Similarly equation (9.32) can be solved for $\bar{a}_n$ in terms of $\bar{w}_n$ and s_n. This illustrates the symmetry of the solutions in the two phases. Finally, equation (9.33) gives s_n.

From the series representation for A_{nn}, it can be seen that

$$A_{nn} = i^n \operatorname{erfc}\left(\sqrt{\alpha} s_0/2\right), \qquad \bar{A}_{nn} = i^n \operatorname{erfc}\left(\sqrt{\bar{\alpha}} s_0/2\right), \tag{9.36}$$

as seen in Tao [2].

9.3 Conclusion

A brief alternate derivation of the Tao solution for the Stefan problem is given using the Laplace transform and its inversion formula. This new derivation clearly illustrates the symmetry between the solutions in the frozen and liquid domains in regard to their structure. The initial conditions for the frozen domain solution and the boundary conditions for the liquid domain solution form a sequence of unknown coefficients which can be solved recursively from the triangular system of equations. Instead of the Taylor series employed by Tao using repeated differentiation of the repeated integrals of the complementary error functions, the explicit-time representation of the inverse operator used here allows isolation of the various powers of $\sqrt{t}$ directly. The present approach has the advantage of reducing the algebraic manipulations by retaining the above mentioned symmetry, avoiding Faa di Bruno differentiation rules and, to some extent, replacing the multinomial coefficients through a recurrence relation. In view of the detailed solution given by Tao [2], we do not discuss the existence and convergence questions here.

References

[1] Carslaw, H. S. and Jaeger, J. C., (1959), *Conduction of Heat in Solids*, Oxford University Press, England.

[2] Tao, L. N.,(1978), The Stefan problem with arbitrary initial and boundary conditions, *Quart. Appl. Math.* 36:223–233.

[3] Tao, L. N.,(1979), On free boundary problems with arbitrary initial and flux conditions, *Z. Angew. Math. Phys.* 30:416–426.

[4] Tao, L. N.,(1979), Free boundary problems with radiation boundary conditions, *Quart. Appl. Math.* 37:1-10.

[5] Tao, L. N.,(1979), On solidification problems including the density jump at the moving boundary, *Quart. J. Mech. Appl. Math.* 32:175–185.

[6] Tao, L. N.,(1980), The analyticity of solutions of the Stefan problem, *Arch. Rational Mech. and Anal.*72:285–301.

[7] Tao, L. N.,(1981), The exact solutions of some Stefan problems with prescribed heat flux, *J. Appl. Mech.* 48:732–736.

[8] Weinbaum, S. and Jiji, L. M., (1977), Singular perturbation theory for melting or freezing in finite domains initially not at the fusion temperature, *J. Appl. Mech.* 44:25–30.

[9] Charach, Ch. and Zoglin, P., (1985), Solidification in a finite initially overheated slab, *International Journal of Heat and Mass Transfer*, 28:2261–2268.

[10] Rubinstein, L. I., (1971),*The Stefan Problem*, American Mathematical Society, Providence, RI.

[11] Ockendon, J. R. and Hodgkins, W. R., (1975),*Moving Boundary Problems in Heat Flow and Diffusion*, Clarendon Press, Oxford.

[12] Wilson, D. G., Solomon, A. D., and Boggs, P. T., (1978) *Moving Boundary Problems*, Academic Press, New York.

[13] Fukusako, S. and Seki, N., (1987), Fundamental aspects of analytical and numerical methods on freezing and melting heat transfer problems, *Annual Review of Numerical Fluid Mechanics and Heat Transfer*(T. C. Chawla, editor), 1:351–402.

[14] Goodman, T. R., (1958), The heat balance integral and its applications to problems involving a change of phase, *Journal of Heat Transfer*,80:335–341.

[15] Dursunkaya, Z. and Nair, S., (1990), A moving boundary problem in a finite domain, *J. Appl. Mech.*, 57:50–56.

[16] Kolodner, I. I., (1956), Free boundary problem for the heat equation with applications to problems of change of phase, *Comm. Pure and Appl. Math.* 9:1–31.

[17] Boley, B. A., (1961), A Method of Heat Conduction Analysis of Melting and Solidification Problems, *J. Math. Phys.*, 40:300–313.

[18] Nair, S., (1994), Numerical solution of moving boundary problems using integral equations, *Transport Phenomena in Solidification, ASME AMD*, 182:109–118.

[19] Tao, L. N., (1986), A method for solving moving boundary problems, *SIAM J. Appl. Math.*46:254–264.

[20] Crank, J., (1984),*Moving Boundary Problems*, Oxford University Press, England.

[21] Abramowitz, M. and Stegun, I. A., (1965), *Handbook of Mathematical Functions*, Dover Publications, New York.

10

A Brief History of IIT Physics

Harold N Spector[1]
Illinois Institute of Technology

The physics department at IIT can trace its origin to the early years of Armour Institute of Technology, which is one of the predecessors of IIT. When I joined the department in 1966 as an associate professor, Robert Malhiot was the acting chair, having succeeded Paul Copland, who had been the chair for several years up to 1964. The following people were faculty members at that time:

Ray Burnstein	exper. high energy
Forrest Cleveland	atomic spectroscopy
Nguyen Dzoan	exper. plasma
Tom Erber	megagauss fields and complex systems
Fred Ernst	theor. high energy
Leonard Grossweiner	exper. condensed matter: color centers
Arthur Harris	exper. condensed matter
Isadore Hauser	theor. high energy
Caroline Herzenberg	Mössbauer effect (lunar rocks)
Jordan Markham (IITRI chair)	theor. condensed matter: color centers
Robert Malhiot	theor. high energy
Esther Segal	exper. nuclear
Robert Warnock	theor. high energy
Harold Weinstock	exper. condensed matter
Earl Zwicker	exper. condensed matter: color centers

[1]Professor of Physics, Emeritus: E-mail: spector@iit.edu

Robert Estin, who had been heavily involved in physics education, left the department shortly before I arrived. Howard Rubin, an experimental particle physicist, joined the department at the same time as I did. Caroline (Littlejohn) Herzenberg left the physics department the year after I arrived, going to IITRI to investigate lunar rocks using the Mössbauer effect. Arthur Harris and Nguyen Dzoan also left in the late 1960s. Art went to Zenith Corporation and Dzoan took a position in the electrical engineering department at Notre Dame University.

The department hired three new assistant professors, Cheuk Chau, who collaborated with Harold Weinstock in the Low Temperature Laboratory, Robert Quigley, a theoretical condensed matter physicist who would collaborate with Jordan Markham, and Porter Johnson, a theoretical particle physicist, who collaborated with Robert Warnock. Fred Ernst dropped research in particle physics to return to his first love, the general theory of relativity. He organized a relativity group at IIT together with Robert Malhiot and Isadore Hauser. Fred gained renown in discovering and developing the Ernst equation. Richard Isaacson joined the relativity group as a visiting assistant professor. Robert Malhiot became the chair in 1968, a position he held until 1970. He was promoted to the rank of professor.

Robert Quigley left IIT after several years to join the physics department at Western Washington University, where he switched from condensed matter physics to astronomy. He became a professor there and served for several years as chair of the department. Fritz Herlach, who had been at Frascati National Laboratory in Italy, joined the department as associate professor and collaborated with Tom Erber on generating megagauss magnetic fields. After conducting a successful experiment at SLAC exploring the effects of such fields with Erber, he was promoted to professor. Herlach left after a few years to take an academic position at the University of Leuven, in Belgium.

In 1970, Leonard Grossweiner became chair of the department, a position he held until 1981. He switched his research interests from condensed matter physics to biophysics, an area in which he continued to work for the rest of his life. William Brennan, a former student of Len's, was hired as an assistant professor and also served as assistant chair, staying for a few years until he left the department for a job in industry. Porter Johnson then took on the position of assistant chair. Forrest Cleveland retired during this period because he had reached the age of 65 for mandatory retirement. He then went to the University of Kentucky. Jeff Davis, a condensed matter experimentalist working on optical properties of solids, was hired as an assistant professor. Together with Fritz Herlach and me, he explored the effect of megagauss magnetic fields on the

optical properties of semiconductors.

Jeff left IIT for San Diego State University where he became a professor and a fellow of the Optical Society of America. In the early 1970s, Porter was promoted to associate professor and Cheuk left IIT to join the California State University in Chico. Wu-Ki Tung, a particle theorist, joined IIT as an associate professor. Richard Isaacson left IIT to take a position with the National Science Foundation in Washington and became one of the driving forces in pushing for LIGO. Esther Segal left the department at about the same time. Peter Silverman, a condensed matter experimentalist, came to IIT as an assistant professor for potential collaboration with Harold Weinstock in low temperature physics. He played an important role in setting up the Copland Laboratory, but soon left IIT for a job in industry.

Also, during this period of time, Gerald Cohn and Joseph Baugher were hired as assistant professors to work in the area of biophysics, and Chumin Fu was hired as an assistant professor in the area of experimental particle physics. It was in the early 1970s that the department reached its maximum size of over 20 faculty members. From 1977-79, Nicholas Karnezos, who got his degree from the University of Athens, was hired as a visiting assistant professor. While he was in the department, he collaborated in the low temperature laboratory with Harold Weinstock, investigating the effects of radiation damage on the properties of superconductors. When he left the department he joined AT&T Teletype, which later became Lucent Technologies.

Several people were promoted to the rank of professor during the 1970s. Harold Weinstock and Ray Burnstein were promoted in the early 1970s while I was promoted in 1976. Bob Warnock left the department to to move to San Francisco, where his wife took a faculty position at a prominent medical school there. Tim Rynne, who had received his Ph.D. from IIT working with Tom Erber, became a visiting assistant professor in the department where he continued to work with Tom on megagauss physics and with me on free carrier absorption in semiconductors in strong magnetic fields. Chumin Fu, Gerald Cohn, and Joe Baugher all left IIT after a few years. Joe went to AT&T Teletype where he remained until around 2000. Gerald Cohn went to Abbott Laboratories where he remained for several years. In this period, Jordan Markham retired from the department, moving to North Carolina and then to Arizona.

Tom Hsiang, a condensed matter experimentalist, and Leonard Lis, an experimental biophysicist, were hired to replace them. Jim Hanlon, an experimental particle physicist was hired to replace Chumin Fu. In 1980 Wu-Ki Tung was also promoted to professor. Tim Rynne left IIT and founded his own company, and was replaced by John Collins, an elementary particle theorist who came

at the rank of associate professor. John gained some renown for his work in renormalization groups in particle physics.

In 1981, Len Grossweiner stepped down from the chairmanship of the department after having served in this position for eleven years. Len continued to do research in biophysics, changing his direction so as to promote the use of photodynamic therapy for the treatment of various cancers, especially that of the esophagus. He established a collaboration with people at Ravenswood Hospital which he maintained even after he retired from IIT in 1996. Wu-Ki Tung then became chair of physics, a position he continued in for three years until 1984.

During Wu-ki's term as chair, Fred Ernst, Porter Johnson and Isidore Hauser were all promoted to the rank of professor. Tom Hsiang left IIT for a position in an electrical engineering department at the University of Rochester. John Zasadzinski, a condensed matter experimentalist working on tunneling in superconductors, who had received his Ph.D. at Iowa State, and Dmitri Niarchos, another condensed matter experimentalist who received his Ph.D. in Greece, were hired as tenure track assistant professors. In 1983, Carlo Segre, another condensed matter experimentalist who received his Ph.D.from UCSD, was hired as assistant professor. In 1983, Leonard Lis left IIT for Kent State University and Jim Longworth, an experimental biophysicist who had received his Ph.D. from University of Sheffield in England was brought in as a visiting associate professor. Jim had been at Oak Ridge and had been president of the Photobiology Society.

In 1984, Porter Johnson became chair of the department. When he assumed the chair, the physics department (as well as several other academic departments) was faced with severe budgetary constraints, forcing the elimination of several positions. Several faculty members left IIT during this period. Jim Hanlon took a staff position at Fermilab, and Dmitri Niarchos returned to Greece. Isadore Hauser retired after having spent a sabbatical leave in Mexico. Harold Weinstock left to become a program director at AFOSR. Fred Ernst left the department for a position at Clarkson University. A year later, Robert Malhiot retired from IIT.

In 1987 Tim Morrison, who received a Ph.D. from the University of Illinois at Urbana, left Argonne National Laboratory to become an associate professor. Together with Carlo Segre, Tim took initiatives that led the beginning of the Synchrotron Center at IIT. As a result of his successful efforts, Grant Bunker, who obtained a Ph.D. from the University of Washington, was hired as an associate professor. Grant succeeded in starting the BIOCAT beamline at the Advanced Photon Source. Tim received tenure around 1990 and Grant in the

mid 1990s. Two other new faculty members were hired in the 1990's. Liam Coffey, a condensed matter theorist who received his Ph.D. from the University of Chicago, was hired as an assistant professor in 1990. Dan Kaplan, an experimental particle physicist who had been at Northern Illinois University, was hired as an associate professor in 1994.

In the early 1990s the department lost two of its three theoretical particle physicists: Wu-Ki Tung, who went to Michigan State University, and John Collins who went to Pennsylvania State University. Also, Earl Zwicker retired in 1991, although he remained active in some of the educational programs, including the SMILE program, until 2006. In 1993, the biology and chemistry departments were merged, with Tim Morrison and Carlo Segre becoming the chair and associate chair, respectively, of the merged department. Also, at this time, Nobel Laureate Leon Lederman, former director of Fermilab who had been at the University of Chicago, joined IIT and became a member of the department.

In 1995, the physics department was merged with the chemistry and biology departments to form an umbrella Department of Biological, Chemical and Physical Sciences, ending the era of physics as an independent academic department. Physics has remained as a division within the merged department. At the time of the merger, Dean Chapman from Brookhaven National Laboratory was hired as an associate professor to be the director of the synchrotron center.

After the merger in 1995, Liam Coffey was promoted to the rank of associate professor. Chris White in experimental particle physics and Linda Spentazouris in experimental accelerator physics were hired as assistant professors in the physics division of BCPS, and both subsequently received tenure. Dan Kaplan and Grant Bunker had also been granted tenure. In addition, Dan Kaplan, Tim Morrison, and Grant Bunker, and Carlo Segre were promoted to the rank of professor.

In 1999, Tom Erber was given the rank of Distinguished Professor at IIT. In 2000, Larry Scott, a theoretical biophysicist, came to IIT at the rank of professor and as the chair of the BCPS department, and remained chair for a five year term. John Zasadzinski subsequently became the next chair of the BCPS department. Ray Burnstein and I retired in 2001. In 2008 both Tom Erber and Porter Johnson retired. All of us have remained professionally active.

11

Electromagnetic Whispering Gallery Modes

Robert L. Warnock[1]
SLAC National Acccelerator Laboratory
Stanford University

Abstract

Coherent Synchrotron Radiation (CSR) from electron storage rings is now well established, in both continuous and bursting modes, following initial observations of the bursting mode reported in 2000-2002. The phenomenon was studied theoretically in 1988, on the basis of two models of the storage ring vacuum chamber: a smooth torus with rectangular cross section, and a cylindrical pillbox. Both models predict coherent radiation into "whispering gallery modes" in which fields of very short wavelength (much shorter than the wave guide cutoff of the chamber) are concentrated near the outer wall. The Fourier spectrum of the field at multiples of the particle revolution frequency shows a series of sharp peaks, each peak resolved into contributions from closely spaced modes having phase velocity close to the particle velocity. Bergstrom at the Canadian Light Source made the important observation that CSR spectra measured at the NSLS VUV synchrotron light source in 2001 show peaks that fit this theory, evaluated with the actual parameters of the machine. Finely

[1]warnock@slac.stanford.edu

resolved spectra from the CLS show similar peaks, which are tentatively understood as whispering gallery modes. To confirm this interpretation is not so easy as at NSLS because the outer vacuum chamber wall at CLS has large deviations from circular form. There is urgent need of a theory including wall perturbations that would clarify the physical picture and allow comparisons with experiment. The paper is dedicated to Thomas Erber on his 80th birthday, and the last section contains some recollections of life with Tom and others at Illinois Institute of Technology in the early 1960's.

11.1 Coherent Synchrotron Radiation (CSR)

Imagine N electrons moving in a circle of radius R with angular velocity ω_0. Suppose that their angular positions θ_i relative to a reference particle are independent identically distributed random variables with probability density $\lambda(\theta)$. Then the expected value of power radiated at angular frequency $\omega = n\omega_0$ is

$$P(n) = (\omega_0 e)^2 \mathrm{Re} Z(n, n\omega_0) \left(\left(\frac{1}{2\pi} \right)^2 N + N(N-1)|\lambda_n|^2 \right) . \qquad (11.1)$$

where λ_n is the Fourier transform at mode number n of $\lambda(\theta)$) and $Z(n, \omega)$ is the radiation impedance at wave number n/R and frequency ω. The first term is understood as the incoherent power due to radiation from N electrons acting individually. It represents the bulk of power radiated in a storage ring, and has a broad spectrum ranging up to a cutoff proportional to γ^3. The second term, the coherent power, comes from electrons radiating collectively at wavelengths for which the bunch form has appreciable Fourier components. Its N^2 dependence makes it interesting for a high intensity radiation source, but its presence depends on the bunch form having Fourier components where the impedance is appreciable.

The radiation impedance depends on the electromagnetic environment. For free space the physics was largely understood as early as 1912, through investigations of G. A. Schott [1]. The impedance can be derived from Schott's approach, which is recounted in text books such as Jackson and Landau and Lifshitz. Beginning with an unpublished report of Schwinger [2] dated 1945, attention was given to the effect of nearby conductors. Betatron accelerators were being developed, and it was necessary to understand the effect of the metallic vacuum chamber on the electron's motion. Schwinger and later Nodvick and Saxon [3] modeled the vacuum chamber by two infinite conducting parallel plates, with

the beam circulating in the midplane. The impedance for this model displays an exponential cutoff, being negligible for wavelengths greater than a "shielding cutoff" given by

$$\lambda_0 \approx 2h(h/R)^{1/2} \ , \tag{11.2}$$

where h is the plate separation. If h and R are typical for storage rings (one or a few centimeters for h, and a few meters for R), this is much less than the waveguide cutoff for the structure, which is of order h. By (11.1) and (11.2) one cannot expect to see coherent radiation unless λ_n is appreciable for

$$n > \frac{2\pi R}{\lambda_0} \approx \pi(R/h)^{3/2}. \tag{11.3}$$

Conditions for producing CSR were achieved by Nakazato *et al.*[4] at Tohoku University in 1989. They used a short bunch from a linac, which probably had a ragged form with short wavelength components, and a single bend with small R in a relatively roomy vacuum chamber.

In storage rings as they were usually operated before ca. 2000, the bunch would normally have a smooth, Gaussian-like form, and would be so long as to have no significant Fourier components with n satisfying (11.3). Around 2000 strong evidence of CSR began appearing in the storage rings of several synchrotron light sources [5, 6, 7, 8, 9, 10]. Without unusual values of h, R, and σ (the nominal bunch length), the necessary conditions were obtained through an instability of the bunch form, giving it a "microstructure" with Fourier components at mode numbers satisfying (11.3). At high peak beam current, this instability can be induced by the coherent radiation of the bunch acting on itself, a coherent "radiation reaction". The CSR induced through an instability comes in bursts, with a time between bursts being a substantial fraction of the longitudinal damping time τ of the ring (the damping being due to incoherent radiation in quanta and the attendant diffusion in phase space).

A theoretical model of bursting was given in [11], based on the parallel plate model of radiation shielding and Vlasov-Fokker-Planck dynamics. The mechanism of the model is roughly as follows: (a) with high peak current the coherent radiation reaction causes microbunching, which gives high Fourier components satisfying (11.3), and a burst follows; (b) rapid mixing in phase space due to the nonlinear and time dependent self-force causes a lengthening and smoothing of the bunch, until (11.3) is no longer satisfied and the burst is terminated; (c) the damping from incoherent radiation shortens the bunch gradually over a time comparable to τ, which increases the peak current, again causing microbunching, satisfaction of (11.3), and a consequent burst.

11.2 Effects of a Closed Vacuum Chamber

When I first arrived at SLAC in 1987, one of the widely experienced accelerator physicists, Phil Morton, suggested that I look at fields induced by a beam circulating in a torodial chamber with rectangular cross section. (In fact he put it a bit more strongly than a suggestion; it would be in my "best interests" to do the problem). This would be essentially different from the parallel plate model in not allowing radiation of energy to infinity. Phil had been thinking in terms of an eigenfunction expansion, the eigenfunctions of the structure being so-called cross products of Bessel functions ([12], §9.1.32) A knowledge of zeros of the cross product is required to satisfy the boundary conditions. By an elementary method for differential equations I was able to avoid the eigenfunction expansion, getting an expression stated directly in terms of Bessel functions without reference to their zeros [13]. In parallel work unknown to us at the time, K.-Y. Ng treated the problem with eigenfunctions [14]. He and I later joined forces to analyze the sub-resonant frequency region by my method, which I think was essential for that case [15].

The analysis showed that the physics is almost the same for the torus and the cylindrical pillbox, provided that in the toroidal case the beam is not too close to the inner radius. In either case the wave function is concentrated near the outer wall, so presence or absence of the inner wall has little effect. Here I review and later extend the theory for the pillbox. The analysis is in cylindrical coordinates (r, θ, z) with the polar axis along the axis of the circular cylinder. The radius of the pillbox is b and its planar ends are at $z = \pm g,\ h = 2g$. The walls of the chamber are taken to be perfectly conducting, but Ref.[13] shows how to incorporate wall resistance. All components of electric and magnetic fields can be expressed in terms of the axial components $E_z,\ H_z$. The first step is to make a Laplace transform of the fields in the time t, and Fourier developments in θ and z, choosing the trigonometric functions in z so that the boundary conditions on the planar surfaces (tangential E and normal H equal to zero) are satisfied term-by-term in the z-series. Thus the axial fields are represented as

$$E_z(r,\theta,z,t) = \int_{\mathrm{Im}\omega=v} d\omega e^{-i\omega t} \sum_{n=-\infty}^{\infty} e^{in\theta} \sum_{p=0}^{\infty} \cos\alpha_p(z+g) E_{znp}(r,\omega)\ , \quad (11.4)$$

$$H_z(r,\theta,z,t) = \int_{\mathrm{Im}\omega=v} d\omega e^{-i\omega t} \sum_{n=-\infty}^{\infty} e^{in\theta} \sum_{p=1}^{\infty} \sin\alpha_p(z+g) H_{znp}(r,\omega)\ , \quad (11.5)$$

where $v > 0$ and $\alpha_p = \pi p/h$ is the vertical mode number. The initial values from the Laplace transform are put equal to zero. The other field components and the charge and current densities have similar developments with the trigonometric functions of z chosen by the following scheme:

$$(E_r, H_\theta, H_r, E_\theta, \rho, J_r, J_\theta, J_z) \leftrightarrow (\sin, \cos, \cos, \sin, \sin, \sin, \sin, \cos) \ . \tag{11.6}$$

The Laplace-Fourier coefficients of E_z, H_z satisfy Bessel equations with sources. With SI units,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_{znp}}{\partial r}\right) + \left(\gamma_p^2 - \frac{n^2}{r^2}\right)E_{znp} = Z_o\alpha_p c\rho_{np} \tag{11.7}$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial H_{znp}}{\partial r}\right) + \left(\gamma_p^2 - \frac{n^2}{r^2}\right)H_{znp} = -\frac{1}{r}\frac{\partial}{\partial r}(rJ_{\theta np}) \ , \tag{11.8}$$

$$\gamma_p^2 = \left(\frac{\omega}{c}\right)^2 - \alpha_p^2 \ , \quad \alpha_p = \frac{\pi p}{h}, \quad Z_o = \left(\frac{\mu_o}{\epsilon_o}\right)^{1/2} = 120\pi\ \Omega \ . \tag{11.9}$$

Since in the model studied the particles of the beam have only azimuthal motion, $J_r = J_z = 0$. The general solutions of (11.7) and (11.8) that are regular at $r = 0$ have the form

$$\begin{aligned} E_{znp}(r,\omega) &= A_{np}J_n(\gamma_p r) + e_{znp}(r,\omega) \ , \\ H_{znp}(r,\omega) &= C_{np}J_n(\gamma_p r) + h_{znp}(r,\omega) \ , \end{aligned} \tag{11.10}$$

where e_{znp} and h_{znp} are particular solutions regular at $r = 0$. From these solutions one can express all fields, using formulas obtained by algebra from Maxwell's equations after the Laplace-Fourier transform, namely,

$$E_{rnp} = -\frac{1}{\gamma_p^2}\left[\alpha_p\frac{\partial E_{znp}}{\partial r} + Z_o\frac{\omega}{c}\frac{n}{r}H_{znp}\right] \ , \tag{11.11}$$

$$H_{\theta np} = \frac{i}{\gamma_p^2}\left[\frac{1}{Z_o}\frac{\omega}{c}\frac{\partial E_{znp}}{\partial r} + \alpha_p\frac{n}{r}H_{znp}\right] \ , \tag{11.12}$$

$$H_{rnp} = \frac{1}{\gamma_p^2}\left[\frac{1}{Z_o}\frac{\omega}{c}\frac{n}{r}E_{znp} + \alpha_p\left(\frac{\partial H_{znp}}{\partial r} + J_{\theta np}\right)\right] \ , \tag{11.13}$$

$$E_{\theta np} = -\frac{i}{\gamma_p^2}\left[\alpha_p\frac{n}{r}E_{znp} + Z_o\frac{\omega}{c}\left(\frac{\partial H_{znp}}{\partial r} + J_{\theta np}\right)\right] . \tag{11.14}$$

The boundary conditions at $r = b$ are $H_r = 0,\ E_z = E_\theta = 0$. By (11.13) and (11.14) these conditions are satisfied if

$$E_{znp}(b,\omega) = 0 \ , \quad \frac{\partial H_{znp}}{\partial r}(b,\omega) = 0 \ , \tag{11.15}$$

since the current is zero at the boundary. By applying (11.15), (11.10), (11.13), and (11.14), it is seen that the boundary conditions immediately determine the unknown coefficients as

$$E_{znp}(r,\omega) = -\frac{J_n(\gamma_p r)}{J_n(\gamma_p b)} e_{znp}(b,\omega) + e_{znp}(r,\omega) \ , \tag{11.16}$$

$$H_{znp}(r,\omega) = -\frac{J_n(\gamma_p r)}{\gamma_p J_n'(\gamma_p b)} h'_{znp}(b,\omega) + h_{znp}(r,\omega) \ , \tag{11.17}$$

where primes indicate derivatives.

To specify the source terms in (11.7) and (11.8) let us take a simple model of the charge density of a bunch of total charge q, namely

$$\begin{aligned} \rho(r,\theta,z,t) &= q\lambda(\theta-\omega_o t)H(z)W(r)\,, \\ \int_0^{2\pi} \lambda(\theta)d\theta &= \int_{-g}^{g} H(z)dz = \int_0^b W(r)rdr = 1 \ , \end{aligned} \tag{11.18}$$

where $\lambda(\theta+2\pi)=\lambda(\theta)$ and H and W are some functions concentrated near $z=0$ and $r=R$, respectively. Thus the form of the longitudinal density $\lambda(\theta)$ is time independent, but a generalization to allow a deforming density $\lambda(\theta,t)$ is possible. The parallel plate model with deforming bunch is discussed in [16]; the pillbox model could be treated similarly. The current density which with (11.18) satisfies the continuity equation is

$$(J_r,\ J_\theta,\ J_z) = (0,\ \beta c\rho r/R,\ 0) \ , \tag{11.19}$$

where the revolution frequency is $\beta c/R=\omega_0$. The Laplace-Fourier transform of ρ for $\mathrm{Im}\omega>0$ is

$$\begin{aligned} \rho_{np}(r,\omega) &= \frac{1}{2\pi}\int_0^\infty dt e^{i\omega t}\frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-in\theta}\frac{1}{g}\int_{-g}^{g} dz \sin\alpha_p(z+g)\rho(r,\theta,z,t) \\ &= \frac{iq\lambda_n H_p W(r)}{2\pi(\omega-n\omega_0)} \ , \end{aligned} \tag{11.20}$$

with

$$\lambda_n = \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta}\lambda(\theta)d\theta \ , \quad H_p = \frac{1}{g}\int_{-g}^{g}\sin\alpha_p(z+g)H(z)dz \ . \tag{11.21}$$

By using the method of variation of parameters and the value of the Wronskian of Bessel functions ([12], §9.1.15) one finds the required particular solutions

of (11.7) and (11.8). Choosing for simplicity the radial distribution

$$W(r) = \frac{\delta(r-R)}{r} \ , \tag{11.22}$$

and applying (11.20,11.19), the solutions can be stated explicitly as

$$e_{znp}(r,\omega) = -\frac{\pi}{2} Z_o \alpha_p c \Phi p_n(\gamma_p r, \gamma_p R)\Theta(r-R) \ , \tag{11.23}$$

$$h_{znp}(r,\omega) = -\frac{\pi}{2} \beta c \gamma_p \Phi q_n(\gamma_p r, \gamma_p R)\Theta(r-R) \ , \tag{11.24}$$

$$\Phi = \frac{iqH_p\lambda_n}{2\pi(\omega - n\omega_0)} \ , \tag{11.25}$$

where $\Theta(x) = 1$ for $x \geq 0$ and 0 otherwise. Here and in the sequel the notation for cross products of Bessel functions follows Ref.[12]:

$$p_n(x,y) = J_n(x)Y_n(y) - Y_n(x)J_n(y) \ , \tag{11.26}$$

$$q_n(x,y) = J_n(x)Y_n'(y) - Y_n(x)J_n'(y) \ , \tag{11.27}$$

$$r_n(x,y) = J_n'(x)Y_n(y) - Y_n'(x)J_n(y) \ , \tag{11.28}$$

$$s_n(x,y) = J_n'(x)Y_n'(y) - Y_n'(x)J_n'(y) \ . \tag{11.29}$$

From the above results and (11.14) one can assemble $E_{\theta np}$, which is the main quantity of interest since it gives the component of the electric field in the direction of motion of the beam, hence the work done by the field on the beam, equal in magnitude to the energy radiated. In practice it is adequate to work with the average of E_θ with respect to the transverse distribution,

$$\begin{aligned}
\mathcal{E}_\theta(\theta,t) &= \int_0^b r dr \int_{-g}^{g} dz W(r) H(z) E_\theta(r,\theta,z,t) = \\
&= \int d\omega e^{-i\omega t} \sum_{n,p} e^{in\theta} E_{\theta np}(R,\omega) \int dz \sin\alpha_p(z+g) H(z) \\
&= \int d\omega e^{-i\omega t} \sum_{n,p} e^{in\theta} E_{\theta np}(R,\omega) g H_p \ .
\end{aligned} \tag{11.30}$$

Accordingly, the impedance $Z(n,\omega)$ is defined by

$$\hat{\mathcal{E}}_\theta(n,\omega) = g\sum_p H_p E_{\theta np}(R,\omega) = -2\pi R \ Z(n,\omega)\hat{I}(n,\omega) \ , \tag{11.31}$$

$$\hat{I}(n,\omega) = \frac{iq\omega_0\lambda_n}{2\pi(\omega - n\omega_0)} \ . \tag{11.32}$$

Here $\hat{I}(n,\omega)$ is the Laplace-Fourier transform of the current

$$I(\theta,t) = \int dr \int dz J_\theta(r,\theta,z,t)\,,$$

so that (11.31) is "Ohm's Law" with voltage per turn of $-\hat{\mathcal{E}}_\theta(n,\omega)/2\pi R$.

Now the impedance can be assembled from (11.14,11.16,11.17,11.23,11.24,11.31). The result is [17]

$$\begin{aligned} Z(n,\omega) &= i\pi^2 Z_0 g R \sum_{p=0}^{\infty} H_p^2 \Bigg[\frac{\omega R}{c} \frac{J'_{|n|}(\gamma_p R)}{J'_{|n|}(\gamma_p b)} s_{|n|}(\gamma_p b, \gamma_p R) \\ &\quad + \frac{n}{\beta}\left(\frac{\alpha_p}{\gamma_p}\right)^2 \frac{J_{|n|}(\gamma_p R)}{J_{|n|}(\gamma_p b)} p_{|n|}(\gamma_p b, \gamma_p R) \Bigg]\,. \end{aligned} \tag{11.33}$$

The properties of Bessel functions under a reflection $n \to -n$ have been used to write the result in a form valid for both signs of n. The threshold for CSR in this model comes at the lowest frequency for which a Bessel function in the denominator of (11.33) has a zero, assuming $\omega = n\omega_0$. This in turn is about where $\gamma_p b = n$, since the Bessel function begins oscillating about zero when its argument is close to n. From this the threshold for the p-th vertical mode is as stated in Eq.(90) of [13],

$$n > n_{0p} = \frac{\pi p b}{\sqrt{2}h}\left(\frac{R}{b-R}\right)^{1/2}\,. \tag{11.34}$$

Now from the inversion theorems for Laplace and Fourier transforms the so-called induced voltage can be retrieved [16],

$$V(\theta,t) = -2\pi R\mathcal{E}(\theta,t) = \sum_n e^{in\theta} \int_{\mathrm{Im}\ \omega = v} d\omega e^{-i\omega t} Z(n,\omega)\hat{I}(n,\omega)\,, \tag{11.35}$$

together with the radiated power as the negative of the work done on the beam per unit time ([16], Eq.(20)),

$$\mathcal{P}(t) = q\omega_0 \sum_n e^{in\omega_0 t}\lambda_n^* \int_{\mathrm{Im}\ \omega = v} d\omega e^{-i\omega t} Z(n,\omega)\hat{I}(n,\omega)\,. \tag{11.36}$$

The integral over ω can be evaluated by deforming the contour to an infinite semi-circle in the lower half-plane. This can be done because $Z(n,\omega)$ is analytic

in ω except for poles on the real axis, and tends to a constant at infinity in complex directions [16]. Poles come from zeros of the factors $J_n(\gamma_p b),\ J_n'(\gamma_p b)$, and γ_p^2 in the denominators of (11.33). With the usual notation $j_{ns},\ j_{ns}'$ for zeros of $J_n,\ J_n'$, the positive pole positions are defined as follows:

$$\begin{aligned}
\omega_{nps}^{TE} &= \frac{c}{b}\left[j_{ns}'^2 + (\alpha_p b)^2\right]^{1/2}, \quad s = 1, 2, \cdots \\
\omega_{nps}^{TM} &= \frac{c}{b}\left[j_{ns}^2 + (\alpha_p b)^2\right]^{1/2}, \quad s = 1, 2, \cdots \\
\omega_p^{WG} &= \alpha_p c
\end{aligned} \tag{11.37}$$

Let us label the positive pole positions as ω_j, where j is a multiindex such as $j = (TE, n, p, s)$. Each pole at ω_j has a counterpart at $-\omega_j$, and with an appropriate definition [16] of $\gamma_p(\omega)$ at complex ω one has the reflection property $Z(-n, -\omega) = Z^*(n, \omega)$. The notations TE, TM, WG mean "transverse electric", "transverse magnetic", and "waveguide". The nomenclature is not the usual one, since here the transverse direction is transverse to the axis of the cylinder, not transverse to the beam as in discussions of wave guide modes in a beam tube. On the other hand, the notation WG is to recognize that $\alpha_p c$ is the waveguide cutoff frequency for a guide formed by infinite parallel plates with separation h.

To evaluate the integral write $\mathsf{R}(n, \omega_j)$ for the residue of the pole at ω_j, and recall that $\hat{I}$ has a pole at $n\omega_0$, for now supposed distinct from all the other poles. Deformation of the contour gives

$$\begin{aligned}
V(\theta, t) \;=\; & \int_{\mathrm{Im}\ \omega = v} d\omega e^{-i\omega t} Z(n, \omega)\hat{I}(n, \omega) = q\omega_0 \sum_n \lambda_n Z(n, n\omega_0) e^{in(\theta - \omega_0 t)} \\
& + q\omega_0 \sum_n e^{in\theta}\lambda_n \sum_j \left[\frac{e^{-i\omega_j t}\mathsf{R}(n, \omega_j)}{\omega_j - n\omega_0} - \frac{e^{i\omega_j t}\mathsf{R}(n, -\omega_j)}{\omega_j + n\omega_0}\right] .
\end{aligned} \tag{11.38}$$

Notice that if $n\omega_0$ should hit exactly one of the $\pm\omega_j$, which will happen only "by accident" for a general choice of ω_0, then the second term of (11.38) has a polar singularity but it is canceled by a corresponding pole from the first term since

$$Z(n, n\omega_0) \sim \frac{R(n, \pm\omega_j)}{n\omega_0 - \omega_j} . \tag{11.39}$$

Thus the electric field is finite at the TM and TE resonances, which seems entirely reasonable. On the other hand, one reaches the false conclusion that the field is infinite when $n\omega_0 = \omega_j$ if the Fourier transform with respect to

time is used in a formal and unjustified way, in place of the well justified Laplace transform. The true Fourier transform does not exist, simply because fields do not decay at $t = \pm\infty$. The formal Fourier transform of $I(\theta, t)$ is $\tilde{I}(n, \omega) = q\omega_0\lambda_n\delta(\omega - n\omega_0)$ in place of (11.32). Since this is not analytic in ω, one cannot account appropriately for the poles of $Z(n, \omega)$ which constitute its essence. Rather, the formal Fourier inversion is an integral along the real ω-axis giving only the first term V_1 of (11.38) which blows up if $n\omega_0 \to \omega_j$:

$$\begin{aligned} V_1(\theta, t) &= q\omega_0 \sum_n e^{in\theta}\lambda_n \int_{\mathrm{Im}\ \omega=0} d\omega e^{-i\omega t} Z(n, \omega)\delta(\omega - n\omega_0) \\ &= q\omega_0 \sum_n e^{in(\theta-\omega_0 t)} Z(n, n\omega_0)\ . \end{aligned} \tag{11.40}$$

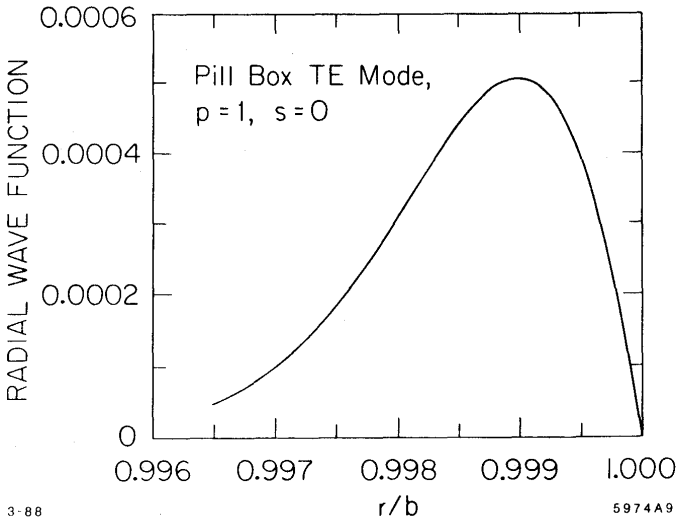


Figure 11.1: Typical TE radial wave function, $p = 1,\ s = 1$ ($s = 0$ in previous notation)

When the chamber is given finite conductivity, as it was in Ref.[13], the resonance pole positions acquire a negative imaginary part, so that the second term in (11.38) vanishes at large t; in practice "large" is a matter of a few revolution times. There is also a new contribution from a branch cut $[0, -i\infty)$,

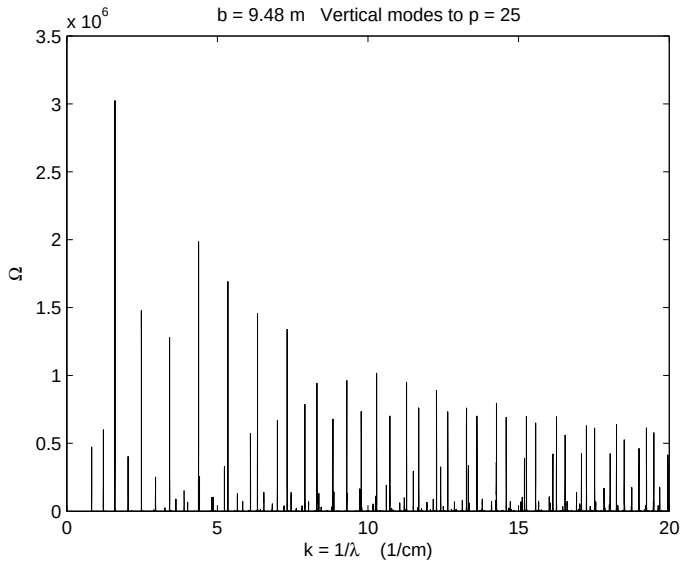


Figure 11.2: Re$Z(n, n\omega_0)$ for parameters of NSLS light source, vs. wave number $1/\lambda$ in units of cm^{-1}. All vertical modes up to $p = 25$ are included.

but it vanishes as well at large t. Thus with resistive walls V_1 is effectively the full result for V, and similarly for the power $P(t)$, which is time-dependent at large t.

11.3 Possible Observation of Whispering Gallery Modes

Let us now plot Re$Z(n, n\omega_0)$ using the resistive wall model, with a choice of parameters for the NSLS VUV synchrotron light source at Brookhaven National Laboratory as stated in [19]: $R = 1.91$m, $w = 8$ cm, $h = 4$ cm, where R is the bending radius and w is the horizontal width of the chamber. Since I do not know exactly the horizontal location of the beam relative to the chamber, I put the beam at $r = R$ and adjust the outer wall radius b to fit the data. This gives $b = 1.948$ cm, meaning that the beam is 2mm off center in the chamber. The toroidal and pillbox models give nearly identical resonant frequencies, agreeing

to about three digits.

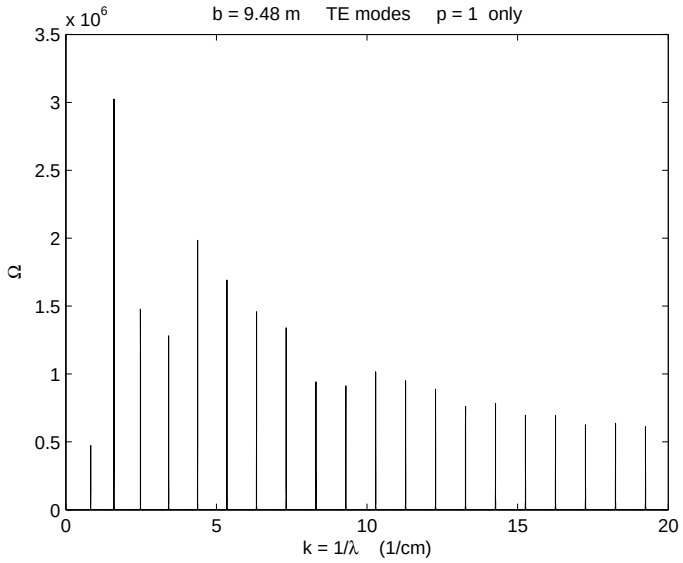


Figure 11.3: Contribution to $\mathrm{Re}Z(n, n\omega_0)$ from TE modes with $p = 1$ only.

The plot in Fig.11.2 includes all vertical modes up to $p = 25$. More and more of these can be excited as n increases. The plot of Fig.11.3 shows just the TE modes with $p = 1$; these tend to dominate at the lower frequencies. Note that in these plots many values of n and many closely spaced resonance poles contribute to each peak. Every n is plotted, with graphical interpolation as though n were a continuous variable. For a discussion of the spread of n within a peak and close-up graphs of peaks, see Section 7 of [13]. The spread is given by

$$\frac{\Delta n}{n} = \frac{R}{dQ} , \quad d = b - R , \tag{11.41}$$

where Q is the quality factor determined by the wall resistance, $10^4 - 10^5$ for an aluminium chamber in the examples of [13].

Remarkably, an observation of the far IR spectrum at NSLS VUV by Carr *et al.* showed a series of peaks in correspondence with Fig.(11.2) [19]. Their data are shown in Fig.11.4. The spectrometer was a Michelson interferometer

with resolution down to $0.01\,\mathrm{cm}^{-1}$. The spectrum is obtained as the cosine Fourier transform of the interferogram produced by scanning the path length differential between the two halves of a split beam [21]. The radiation goes to the instrument through a diamond window, which does not allow transmission for wavelengths longer than 5mm. At longer wavelengths a large glass window was used with microwave RF measurement techniques complementing interferometry. The measured long wavelength spectrum is shown in Fig.11.5, with open circles for the RF data and the black circles for interferometry. The spectrum of Fig.11.4 appears in both incoherent and coherent radiation, the latter in bursting mode. The spectrum of Fig.11.5 was taken with coherent radiation only, since there was not enough incoherent intensity.

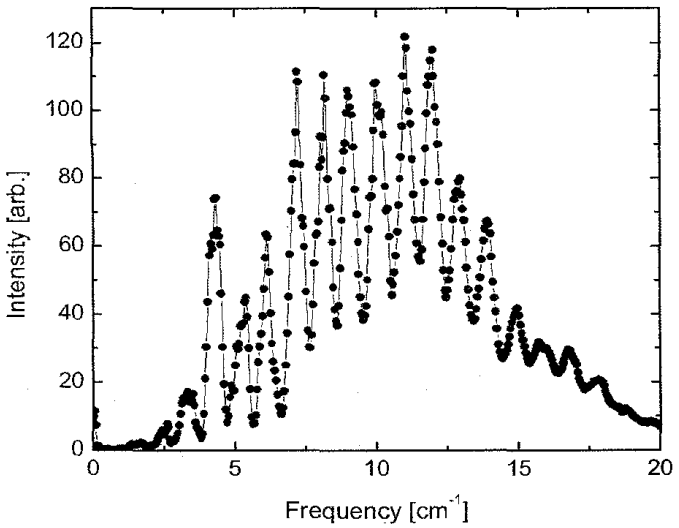


Figure 11.4: Far IR spectrum measured at NSLS

To get an idea of how the theoretical spectrum appears with smoothing over small structures, I convolved the graph of Fig.11.2 with a Gaussian kernel, choosing the sigma of the Gaussian to be $0.19\,\mathrm{cm}^{-1}$ (230 units of n). This was to imitate experimental resolution and give peaks with widths something like those of Fig.11.4. As shown in Fig.11.6, this raises the floor and leaves mostly large peaks close to the TE modes. The low frequency experiment had better resolution, so for its range I convolve with a narrower Gaussian with a sigma of

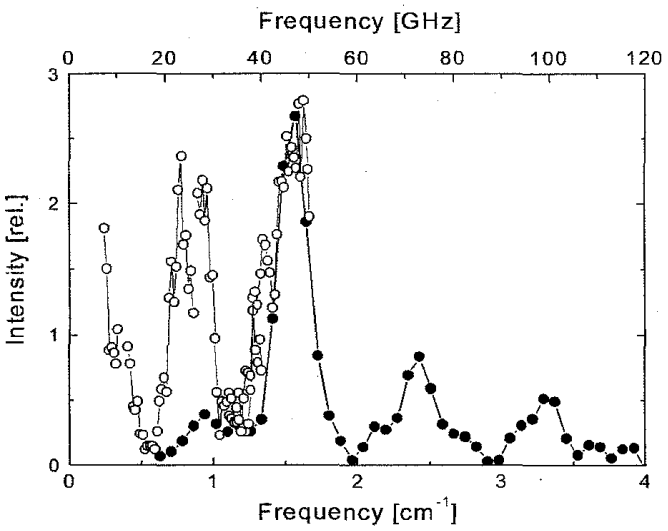


Figure 11.5: Low frequency spectrum measured at NSLS. The black dots are IR data from an interferometer, the open dots from RF measurements.

$0.037\,\mathrm{cm}^{-1}$ to get the plot of Fig.11.7.

Now I will compare the positions of peaks in Figs. 11.6 and 11.7 with the data. The power spectrum depends on the unknown bunch spectrum $|\lambda_n|^2$, and its relation to experiment involves experimental frequency resolution and detector efficiency. The main emphasis should then be on positions of peaks, not their height, although height still does correspond to intensity in some vague and rough sense. Table 1.1 gives experimental values read from the graphs (since I have not had access to data tables); entries with a star ($*$) correspond to small shoulders that could be questioned as authentic signals. Theoretical values are from the plot files for Figs. 11.7, 11.6. In Fig.11.8 I display results of Table 1.1 graphically, omitting the doubtful starred entries.

It is remarkable that the model fits so well with the actual machine parameters. The first ten theoretical lines shown in Fig. 11.8 agree very well with experiment. The experimental line at $0.93\mathrm{cm}^{-1}$ is unexplained. It is striking that the theory accounts for the unequal spacing of peaks at the lowest wave numbers. As is seen in Table 1.1, even the starred data from small shoulders seem to have counterparts in theory. At higher wave numbers the experimental

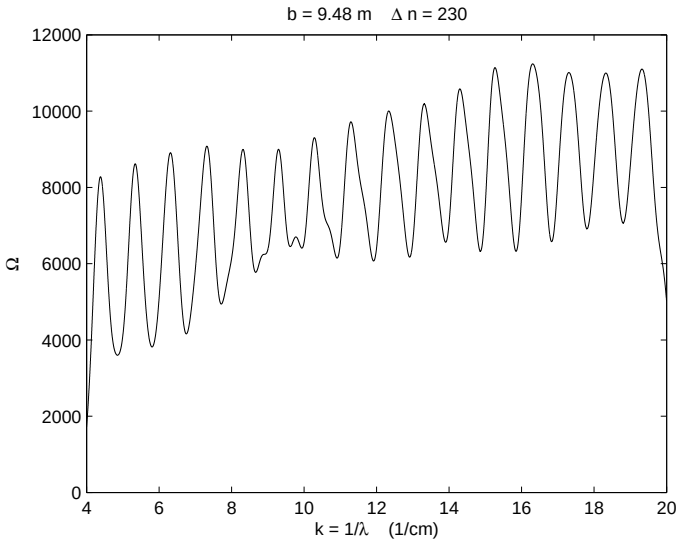


Figure 11.6: High frequency part of spectrum of Fig.11.1, convolved with a Gaussian to average small structures.

peaks are in one-to-one correspondence with theory, but are consistently lower in position. The average experimental spacing for $k \geq 9\text{cm}^{-1}$ is 1cm^{-1}, in agreement with the constant theoretical value. Remember that the radial wave function of the s-th TE mode has s nodes, so the resemblance to theory at high wave numbers is evidence for up to 19 nodes of the field in the radial direction!

Microwave signals at wave numbers less than $0.5\,\text{cm}^{-1}$ are below the CSR threshold of $0.827\,\text{cm}^{-1}$. Here a resonance-like peak is not displayed, and the signal could be due to an ordinary broad band impedance from some element of the vacuum chamber, not related to curvature of orbits, provided that the bunch had strong Fourier components around 4 cm wavelength. In fact, evidence was found for such components in streak camera data showing large density modulation around 5 cm wavelength [19]. This could be associated with the sawtooth oscillations in bunch length found in simulations of the same storage ring [11]. The signal at $0.93\,\text{cm}^{-1}$ is slightly above the CSR threshold but cannot be explained as a curvature effect. Nothing forbids machine impedance effects above the threshold. It would be interesting to do simulations including

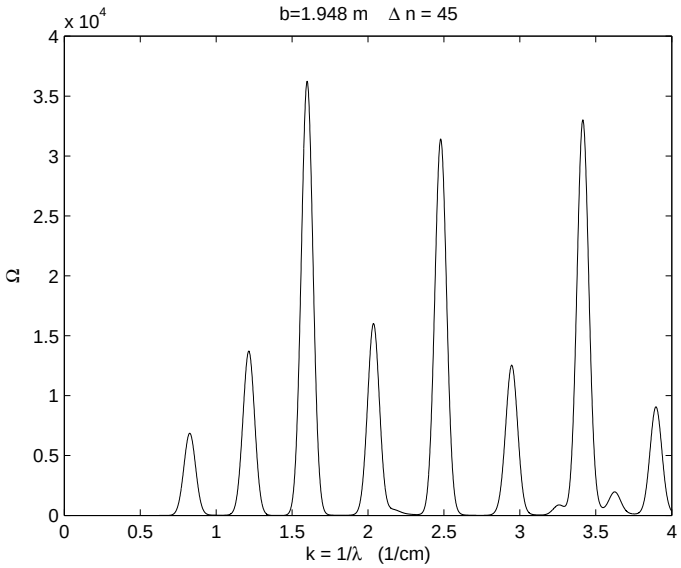


Figure 11.7: Low frequency part of spectrum of Fig.11.1, convolved with a Gaussian.

both radiation and machine impedances to see if their effects mix near the CSR threshold.

This correspondence of NSLS VUV data with theory was first noticed by Jack Bergstrom in 2007 [20]. Rather than using a plot like Fig.11.2, he made the useful observation that a plot of $\mathcal{R}/Q$ as a continuous function of frequency $\omega = n\omega_0$ conveys similar if not identical information, since it has maxima at the resonances; the formulas used are Eqs. (112),(113) of [13].

The authors of Ref.[19] attempted to understand their spectra in terms of interference between direct light and light reflected with a π phase shift from the chamber wall. As far as I know this has not led to a predictive theory, and is criticized by Bergstrom for requiring implausibly complete destructive interference, and also for being unable to account for the non-uniform spacing of the lowest modes that comes out of our model automatically [20].

Bergstrom's motivation to understand the peaks in the spectrum came from measurements at his own laboratory, the Canadian Light Source (CLS), using a spectrometer with resolution down to $0.001\,\mathrm{cm}^{-1}$. This instrument, a Bruker

Table 11.1: Theoretical spectrum compared to data of Figs.11.4,11.5

Exp.	**Thy.**	**Exp.**	**Thy.**
0.80	0.827	6.10	6.31
0.93	—	7.25	7.32
1.32	1.21	8.25	8.32
1.57	1.60	9.00	9.29
2.10*	2.04	10.0	10.28
2.40	2.48	11.1	11.29
2.76*	2.94	12.0	12.33
3.10*	3.26	12.8	13.31
3.30	3.41	13.8	14.3
3.66*	3.62	15.0	15.3
3.88*	3.90	15.8	16.3
4.20	4.38	16.7	17.3
5.25	5.34	18.0	18.3

IFS 125 HR, gives finely resolved spectra like that of Fig.11.9. The storage ring provides CSR in a bursting mode at 2.9 GeV, or in a continuous mode at 1.5 GeV by means of low momentum compaction. Salient features of the interferograms are remarkably stable under changes of the machine set-up. Changes in energy, synchrotron parameters, and number of bunches leave prominent structures unchanged [20]. This suggests that spectral patterns are indeed determined by the vacuum chamber, as the whispering gallery theory would predict.

Although beautiful spectra are produced at CLS, it is not as easy to compare with experiment as at NSLS VUV. Vacuum chambers in dipole bends at NSLS have a smooth circular cylindrical wall following the beam, just as in the simple theory, whereas the chamber at the IR port of CLS is far different, having the triangular form with a secondary triangular excursion around a vacuum pumping port as shown in Fig. 11.10. The average distance d from the beam to the wall is something like 21 cm, as compared to 3.2 cm in the pipes going into and out of the dipole chamber. The dipole chamber forms a sort of cavity, and the question arises of its effect on whispering gallery modes. Are there modes which are resonances of the entire vacuum chamber of the ring but which have localized field patterns that depend sensitively on the local wall profile near the IR port? The spectrometer would see the spectrum of these local fields, since

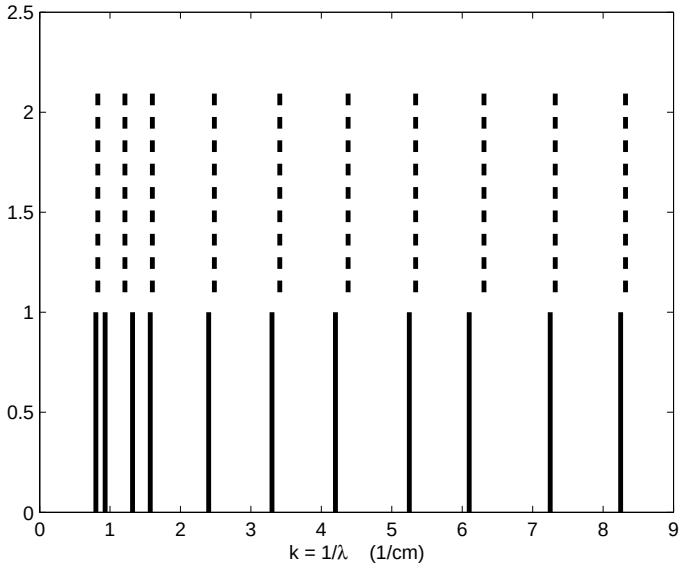


Figure 11.8: Comparison of experimental and theoretical spectra. Solid lines are from experiment, dashed lines from theory.

the IR is extracted in that neighborhood.

To get some hint about this question, Bergstrom devised a procedure of averaging the impedance of the pillbox model as a function of varying wall location $b(\theta)$. He averages over just the large triangular chamber with the IR port, not over the whole ring, and finds spectra agreeing with experiment in some limited domain. At present there is no theoretical justification for this local averaging, but it is certainly interesting that it seems to work.

Having doubts about the effect of the wall on the spectra, since it is so far from the beam, Bergstrom and colleagues proposed an experiment to modify the wall. They will drop a metallic tube into the chamber through the circular pumping port that already exists. The calculation of averaged pillbox impedance indicates that a big shift in the spectrum should occur. The experiment was approved and is scheduled for April, 2010.

If the spectrum shifts as expected, that would provide a big boost to the whispering gallery theory in a qualitative way. It will still be important to

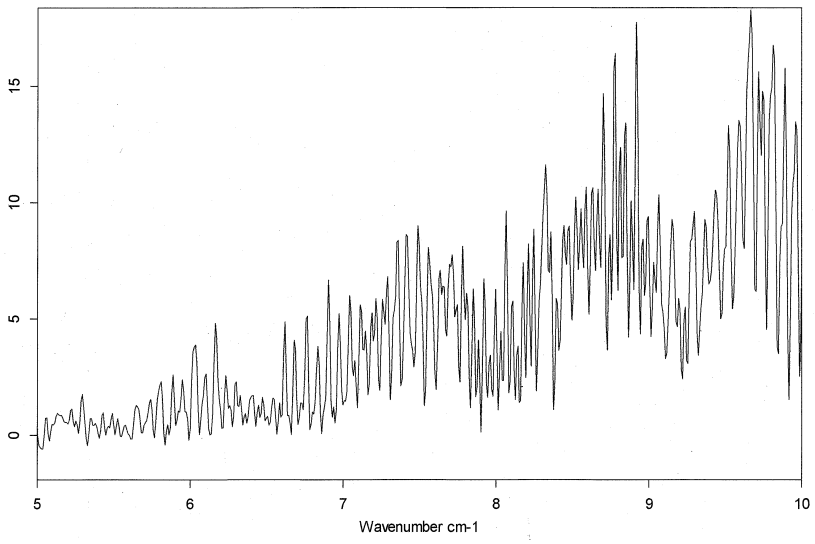


Figure 11.9: Spectrum of CSR with high resolution spectrometer at CLS

provide a quantitative theory for CLS and other storage rings. The challenge for theory is both to fit complex spectra as in Fig. 11.9 and to give a physical picture of perturbed whispering gallery modes. A further motivation for improved understanding comes from the striking observation at CLS of coherence of radiation from successive bunches in a train, for up to 8 bunches, this with CSR in continuous mode [23]. This leads to superintense radiation proportional to $(NN_b)^2$, where N_b is the number of bunches. The coherence requires the bunches to have nearly identical charge distributions, presumably describable as Haïssinski equilibria. A model of the wake field from the radiation impedance in the perturbed vacuum chamber would be a big help toward understanding these distributions.

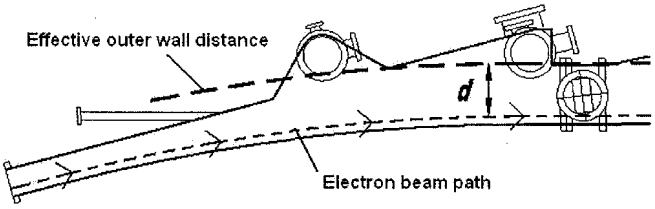


Figure 3: Top view of dipole chamber showing electron beam radius with arrows and distance d to the outer wall. M1 extraction mirror located in rectangular opening in chamber at right, outside of the magnet windings (not shown).

Figure 11.10: Dipole vacuum chamber at IR port of CLS.

11.4 Reminiscence of IIT at Mid-century

I arrived at IIT in 1962, after nine years of a soft life in graduate school and post-doctoral research at big universities. The transition was slightly shocking. You could still smell the Chicago Stock Yards, even though they had been closed for several years, and the neighborhood around the campus had a bombed-out look after "urban renewal". Trains rumbled past on both sides of the campus, so loud as to bring a lecture to a dead stop. The teaching load was cruel, and graduate courses could be taught only with heavy use of black coffee since they were all at night. The elegant buildings of Mies van der Rohe, which had attracted me to IIT in the first place, were ice cold in the winter and like an

oven in the summer. Along with the Mies buildings there were some shabby old structures still in use, and a late-Victorian red brick survivor called Main Building. The latter had a stained glass window bearing the message "From heaven falls light for youth if youth will walk thereby". Unfortunately it turned out that many among our youth were indeed expecting enlightenment to fall on their heads, without much work on their part.

All that aside, I immediately felt at home because of the vitality of Chicago and a warm welcome from Tom Erber and other good colleagues including Izz Hauser, Bob Malhiot, Jordan Markham, Len Grossweiner, and Caroline Herzenberg. Tom's office in Siegel Hall was next to mine and was a center of discussions, with students and various colleagues, often mathematicians and engineers, coming and going. Although his research and mine went in different directions, it was easy to find topics of conversation since Tom was interested in almost anything. In teaching we both admired the European style of exposition in its great years: Sommerfeld, Born, Pauli, Whittaker and Watson, *et al.*. We both thought that mathematics and good physics ought to join hands, not be at odds as so often happens. Tom did very wide reading, often in odd corners of the literature. At my job interview before joining the department he gave me a reference to a paper by Aronszajn on Herglotz functions, something that came up in a paper I had just written. I was awfully impressed with such erudition. As in his reading, Tom's research ranged over all sorts of topics, with wit and élan. Above all he loved electrodynamics, and I hope that he will find the above story amusing. He indirectly contributed to this work by putting me in touch with Gerald Baumgartner, who helped me understand the tricky asymptotics of Bessel functions. Baumgartner's thesis with Tom on the asymptotics of Whittaker functions was motivated by a hard problem in quantum electrodynamics having some relation to synchrotron radiation.

I must also mention the pleasure of working with George Bart, then an IIT graduate student who benefited from Tom's famous course in electrodynamics and did his thesis with me. His later work with Fenster and me introduced some of the methods used in [13] and our paper [18] launched me on my late career in accelerator physics.

Acknowledgments

I am tremendously indebted to Prof. Jack Bergstrom for his appreciation of Ref.[13], which gave part of the motivation for the fascinating experiments at the Canadian Light Source. This research was partially supported by the U.S.

Department of Energy under contract No. DE-AC03-76SF00515.

References

[1] G. A. Schott, "Electromagnetic Radiation", (Cambridge University Press, 1912).

[2] J. Schwinger, "On Radiation by Electrons in a Betatron", unpublished report dated 1945. A transcription made by M. Furman is available as Lawrence Berkeley National Laboratory Report No. LBL-39088 (1996).

[3] J. S. Nodvick and D. S. Saxon, Phys. Rev. **96**, 180 (1954).

[4] T. Nakazato *et al.*, Phys. Rev. Lett. **63**, 1245 (1989).

[5] Å. Anderson, M. Johnson, and B. Nelander, Opt. Eng. **39**, 3099 (2000).

[6] U. Arp *et al.*, Phys. Rev. ST Accel. Beams **4**, 054401 (2001).

[7] G. Carr, S. Kramer, J. Murphy, R. Lobo, and D. Tanner, Nucl. Instr. Meth. Phys. Res. A **463**, 387 (2001).

[8] B. Podobedov *et al.*, Proc. 2001 IEEE Part. Accel. Conf., p.1921; S. Kramer *et al.*, Proc. 2002 Euro. Part. Accel. Conf.

[9] M. Abo-Bakr *et al.*, Phys. Rev. Lett. **88**, 254801 (2002).

[10] J. Byrd *et al.*, Phys. Rev. Lett. **89**, 224801 (2002)

[11] M. Venturini and R. Warnock, Phys. Rev. Lett. **89**, 224802 (2002).

[12] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions", (Dover, New York, 1965).

[13] R. L. Warnock and P. Morton, Part. Accel. **25**, 113 (1990). This work was completed two years before its publication, and was reported in SLAC-PUB-4562 (1988).

[14] K.-Y. Ng, Part Accel. **25**, 153 (1990).

[15] K.-Y. Ng and R. L. Warnock, Phys. Rev. D **40**, 231 (1989).

[16] R. Warnock, R. Ruth, M. Venturini, and J. A. Ellison, Phys. Rev. ST Accel. Beams **8**, 014402 (2005).

[17] For correspondence of (11.33) to Eq.(50) in Ref.[13] put $\Lambda_p = (gH_p)^2$. The impedance was defined slightly differently in that paper, with a square step distribution in place of a general $W(z)$ in the averaging step of Eq.(11.30) above. Note also that in the preprint of [13], SLAC-PUB-4562, the factor Λ_p was defined differently by a factor of 2, but the formulas for the impedance agree in preprint and published paper.

[18] R. L. Warnock, G. R. Bart, and S. Fenster, Part. Accel. **12**, 179 (1982).

[19] G. L. Carr, S. L. Kramer, N. Jiswari, L. Mihaly, and D. Talbayev, Proc. 2001 Part. Accel. Conf., Chicago.

[20] J. Bergstrom, Canadian Light Source, private communication, 2007-2010.

[21] J. W. Goodman, "Statistical Optics", (Wiley, New York, 2000), §5.1.

[22] T. May, J. C. Bergstrom, L. O. Dallin, and D. R. T. Appadoo, IEEE Proc. 33rd Intl. Conf. on Infrared Millimeter and Terahertz Waves, Los Angeles, 2008.

[23] B. E. Billinghurst, T. May, J. Bergstrom, M. DeJong, and L. Dallin, Workshop on Infrared and Microwave Spectroscopy, Banff, 2009, AIP Conf. Proceedings **1214**, pp. 10-12.

12

Hysteresis in Iron, Nickel and STMs

Harold Weinstock[1]
Air Force Office of Scientific Research[2]

I don't think that in my more than 50 years engaged in basic research in one form or another that I have ever met a scientist with a more global knowledge of physics than Tom Erber. Certainly one long-term highlight of my 2 decades as an IIT faculty member in Physics was my interaction with Tom, and especially the interaction that occurred during the 1982-83 academic year when I was on a sabbatical at the Naval Research Lab (NRL).

When I first arrived at NRL, the plan had been to learn how to use a Superconducting Quantum Interference Device (SQUID) in the form of a second-order magnetic gradiometer and then use this system to detect magnetic signatures of prototype cryogenic refrigerators. The only problem was that not one of these refrigerators was delivered that year or ever. I did hit upon other potential applications that included the first measurements using a SQUID for non-destructive evaluation and taking the SQUID gradiometer to NIH to perform neuromagnetic measurements on patients with epilepsy. However, the major activity was devoted to a suggestion Tom made. Namely, Tom wanted to test a hypothesis concerning magnetic hysteresis. Specifically, he wanted to know if the magnetic gradiometer could observe the change in ambient magnetic field when a particular magnetic domain reoriented itself. This behavior in which a domain reorients itself due to a change in external magnetic field or due to some other

[1]Email: harold.weinstock@afosr.af.mil
[2]Arlington, VA 22203-1768

ambient change (such as an increase in temperature) is known as a Barkhausen jump.

In order to measure the change in the magnetic field near the surface of a magnetic field, a number of conditions needed to be satisfied. First, one needed a high purity iron specimen with randomized magnetic domains. That part was easy thanks to a five-nines pure iron cylinder the size of a pill supplied by an NRL colleague. The harder part was to be able to randomize the magnetic domains in a region with almost no magnetic field and then immediately afterward to slowly apply a magnetic field and observe the change in response of the SQUID magnetic gradiometer placed just above the iron "pill." When I explained this problem to another NRL scientist, he recalled that there was a very special facility at the NASA Goddard Space Flight Center (GSFC), about 20 miles north of NRL. This was exactly what was needed. The NASA facility was in a relatively small building in a forest, technically still on the grounds of GSFC, but reachable only via a private road and a couple of miles from the main GSFC campus. The building was constructed entirely of non-magnetic material, and cars had to be parked at least some minimum distance from it. Inside this rather special building was a 3-axis Braunbeck coil configuration, with each set of coils being 12.7-meters in diameter, in other words, 3 mutually perpendicular Helmholtz coils. Using a set of 3 orthogonal flux-gate magnetometers outside these coils and a feedback system of current into the Braunbeck coils, it was possible to create a spherical region of about 2 meters in diameter in the center that was only about 10^{-6} of the earth's magnetic field. This facility had been built and used initially to check on the magnetic moment of the lunar excursion module used by astronauts on the moon, and it was used also to simulate tumbling of rockets in the earth's magnetic field. By the time I contacted NASA in 1983, this special facility had little use, so those in charge of it were happy to have visitors who would pay $250 a day to use the facility and thus help employ the technician assigned to it.

This technician spent most of his time in a nearby building at the controls of the magnetic coil system. Thus, I was basically alone to carry out a set of procedures that Tom had suggested. Once the iron specimen was properly degaussed in the zero-field environment inside the Braunbeck coil region, I would slowly ramp up the current through one set of coils in an orthogonal direction to the axis of the cylindrical iron specimen. This was done in small increments, after which I cycled the field back to zero. This was then repeated many times, but each time increasing the maximum field by a little more than in the previous cycle. By monitoring the SQUID output signal as a function of time on a strip-chart recorder - an instrument perhaps unknown to young scientists today -

it was possible to track the relatively slow change in ambient field. However, beyond some threshold field, it was possible to observe a sharp shift in field value due to a Barkhausen jump, that is, a reordering of one or more magnetic domains in the polarizing field. While at this special NASA facility, I often called Tom while the experiment was in progress, and we had many fruitful discussions. He made major contributions to the experimental aspects of this research.

The details of what transpired are found in a 19-page Physical Review paper [1] that Tom and I coauthored with my NRL colleague, Marty Nisenoff, whose major contribution was to educate me on the operation of the SQUID system. Briefly, a major finding of this research (as Tom had anticipated) was that there is a magnetic field threshold for producing Barkhausen jumps and correspondingly, a threshold for hysteretic behavior. By cycling the external field back to zero, one can anneal away the jumps that occurred at all fields less than or equal to the highest field used in this sequence of "training" exercises, thus extending the range of magnetic field that one could operate in without experiencing hysteretic behavior. However, this magnetic training worked only up to a point, or rather only up to a second threshold. Beyond this second threshold, no amount of cycling back and forth to zero applied field could eliminate the Barkhausen jumps.

Tom had anticipated the behavior just discussed because he had been studying hysteretic behavior in a variety of physical systems, and I'm sure there are other contributions to this volume that deal with this phenomenon, especially in the area of structural mechanics. Nevertheless, I believe it is Tom's encyclopedic knowledge of basic physics and his multilingual skills that resulted in this seminal contribution to the study of hysteresis in magnetic systems. Tom was able to show that this magnetic behavior was identical in form to that of acoustic emission from metallic structures subjected to cyclic stress. Tom even knew about the first report of this acoustic emission by a certain J. Kaiser in a rather obscure (to me at least) German journal, Arch. Eisenhüttenwesen, in 1953. Tom always refers to this phenomenon as the Kaiser effect. Although there are metallurgists or materials scientists who use this nomenclature, if you check Wikipedia^{R} for the Kaiser effect, it will refer to a cosmological occurrence observed by a person named N. Kaiser.

When I returned to IIT in August 1983 from my sabbatical year, I was armed with a vast amount of stripchart paper detailing not only the hysteretic behavior of the aforementioned iron "pill," but I also had similar types of data associated with magnetic field cycles on iron whiskers, slender single crystals that had been grown and studied in the laboratory of Professor R. V. Coleman

at the University of Virginia. These iron whiskers had been found to have typically only about 6 magnetic domains per whisker. Trying to mount them in order to obtain cyclic data as was done for the cylindrical, multigrain iron pill, was a considerable challenge, and again Tom had been most helpful. I pointed out to Tom that there were no Barkhausen jumps to be seen, but there did appear to be a threshold field at which hysteresis occurred, albeit quite continuously as opposed to isolated Barkhausen jumps. Later on, in analyzing the data I brought with me, Tom analyzed these 2 different sets of data using an extension of ideas put forward in 1942 by Louis Néel, the French scientist who was awarded the Nobel Prize in Physics in 1970 for his work on magnetic materials. The paper Tom cited was written in French. Tom liked to point out that even under Nazi occupation, Néel was able to do good scientific research. With the data I showed Tom, he was able to create energy "landscapes" that could explain both the Barkhausen jumps in the multicrystalline specimen and in the iron whiskers. Tom analyzed just about every piece of data I had taken and worked for untold hours to present an overall analysis that was the major part of the Physical Review paper mentioned earlier.

I was fortunate to receive an ASEE Summer Fellowship to return to NRL in 1984. Again at Tom's urging, I obtained a nickel specimen that I subjected to the same type of regimen as was done on the iron specimens, with basically the same general results. This work was the basis for our second publication [2], this time in the Journal of Applied Physics.

I took a leave of absence for the 1984-85 academic year to be a visiting program manager at the Air Force Office of Scientific Research (AFOSR), but I returned regularly during the year to take care of some duties associated with a grant from the Office of Naval Research. Naturally, I spent time with Tom discussing his favorite subject, hysteresis. At that time there was much excitement about the invention of a new tool, the scanning tunneling microscope (STM) by 2 IBM-Zürich scientists, Gerd Binnig and Heinrich Rohrer, an invention that would earn them the 1986 Nobel Prize in Physics. The key element in this invention was the ability to achieve atomic resolution of a surface by changing the voltage on a piezoelectric material known colloquially as PZT.[3]

Tom was concerned that the resolution of this now commonplace instrument was compromised by hysteresis in the cycling of the piezoelectric actuator. In those early years in the life of STMs and various related scanning instruments, you needed to build one yourself. In mid 1986 I had taken a permanent position at AFOSR, and one of the projects I managed was work on superconducting

[3]OK, it's lead, zirconate, titanate. I'll spare the reader the detailed chemical description.

materials in the laboratory of Allen Goldman at the University of Minnesota. Under other funding an STM had been constructed in his lab, and when I told Allen what we would like to do, he volunteered to let Tom and me direct a post doc and 2 grad students in setting up a series of cycles to check for hysteresis. Of note is the fact that one of the students, Ed Nowak, had been an IIT undergraduate majoring in Physics. (Ed is now a Professor of Physics at the University of Delaware.) The experiment was designed to control the vertical position of the STM tip through several cycles over the surface of a graphite specimen. Once again, as Tom had predicted, there were regions of reversible and irreversible responses. These results were published in the Journal of Applied Physics in 1990 [3].

Over the years since, Tom and I have spoken often, always discussing, at least in part, his universal interest in hysteresis. Additionally, he seems to be in touch with a broad swath of what is going on in the world of physics. It is sobering to realize that despite reaching the age of 80, he maintains his keen insight into physical phenomena and discoveries, and although he only recently attained emeritus status, he still maintains his quest for scientific discovery. I believe Tom has been a member of the IIT faculty for over 50 years. He has been an exciting teacher, and students whom I steered to his courses invariably thanked me for this advice. Tom, by any account, has been a stellar member of the IIT faculty during his long tenure. May he continue to flourish and to achieve success in the golden years ahead.

Thanks Tom for your wisdom and friendship these many years.

References

[1] H. Weinstock, T. Erber and M. Nisenoff, Threshold of Barkhausen emission and onset of hysteresis in iron. Phys. Rev. B31, 1535-1553 (1985).

[2] H. Weinstock and T. Erber, Training of Barkhausen emission in nickel and iron. J. Appl. Phys. 63, 3952-3954 (1988).

[3] T. Erber, K. A. McGreer, E. R. Nowak, J-C. Wan and H. Weinstock, Onset of hysteresis measured by scanning tunneling microscopy. J. Appl. Phys. 68, 1370-1372 (1990).

13

GEM and the $K^*(892)$

David White[1][2]
School of Science and Mathematics
Roosevelt University

Abstract

The Gluon Emission Model (GEM), which accurately describes the widths of all known vector mesons, is a model for vector meson production in which vector mesons arise by virtue of quark spin-flip with accompanying gluon emission. Since the spin of the $K^*(892)$ meson is unity, application of GEM towards the $K^*(892)$, long thought to be an excited state of the kaon, is effected for purposes of ascertaining the likelihood that it may actually be itself a vector meson. The current Meson Table published by the Particle Data Group has the $K^*(892)$ listed as a two-component system - the uncharged $K^*(892)^0$ with mass of $896\,MeV$ and width of $50.8\,MeV$ and the charged $K^*(892)^{\pm}$ with mass of $892\,MeV$ and width of $50.7\,MeV$. Surprisingly, GEM indicates that the $K^*(892)$ acts as a standard vector meson stemming from a three-quark base formed by the up (u), down (d), and strange(s) quarks. Specifically, based upon width calculations through the use of GEM associated with various possibilities as to the $K^*(892)$'s construction, the $K^*(892)$ is best repre-

[1]Email:dwhite@roosevelt.edu

[2]"GEM and the $K^*(892)$", republished here by consent of David King, Chair of the Intellectbase International Consortium, who offers his congratulations on the contributions of Thomas Erber over a long and productive career.

sented as a linear combination, χ, of quark/anti-quark states given by $\chi = c\,(uu^* + dd^* + ss^*)$, where $c = 1/\sqrt{3}$ and the * indicates the relevant anti-quark in the expression for χ. For such construction GEM indicates a theoretical width for each mode of the $K^*(892)$ (charged and neutral) as virtually a match to published experimental determinations of same.

Keywords

gluon emission model; $K^*(892)$; vector mesons; isospin = 1/2; quark spin-flip

13.1 Introduction

Conventional wisdom has established the $K^*(892)$ meson as an excited ($J = 1$) state of the K meson. As such, the neutral mode of the $K^*(892)$, the $K^*(892)^0$, is thought to involve a bound state of the down (d) quark and the strange (s) anti-quark (s^*). Its anti-particle, herein designated by $[K^*(892)^0]^*$, involves, therefore, d^* and s, again, by conventional wisdom (PDG (2004), p. 20). Meanwhile, the charged mode of the $K^*(892)$, the $K^*(892)^{\pm}$, would be characterized by a bound state of the up (u) quark and the s^* or vice versa, i.e., u^* and s (PDG (2004, p.20). According to PDG (2004), p.28, the $K^*(892)$ decays nearly 100% into various πK products, and, presumably through the use of coincidence analysis, experiments indicate that decays involving no net charge associated with the decay products take place at essentially the same rate as decays involving one unit of charge associated with same, as the decay width, Γ_0, of the $K^*(892)^0$ is inferred to be essentially the same as the decay width, $\Gamma_{\pm}$, of the $K^*(892)^{\pm}$ [$50.7 \pm 0.6 MeV$ vs. $50.8 \pm 0.9\, MeV$, respectively]. Also according to PDG (2004), p. 28, the neutral decay mode is associated with a meson mass of $896\, MeV$, the mass, M_0, of the $K^*(892)^0$, while the charged decay mode stems from a meson mass of $892\, MeV$, the mass, $M_{\pm}$, of the $K^*(892)^{\pm}$.

Above is recounted, as noted, the conventional wisdom straight out of PDG (2004) and as it has been expressed in the literature for decades concerning the $K^*(892)$. Below, we wish to apply some "unconventional wisdom" to the situation, the results of which will show that a reinterpretation as to the structure of the $K^*(892)$ appears to be in order. Specifically, assuming the context of high-energy colliding beams as providing for meson production, we apply the Gluon

Emission Model (GEM) to the situation involving the $K^*(892)$.[3] Since $J = 1$ for the $K^*(892)$, the stated application of GEM is valid, as we may assume that the $K^*(892)$ arises via quark spin-flip with accompanying gluon emission.[4]

The particular value realized through the application of GEM is that it yields a formula for the width of any $J = 1$ meson in its ground state, a formula which is highly sensitive to the assumed quark structure of the given $J = 1$ meson under consideration. Specifically, GEMdetermined widths are each proportional to $\sum_i q_i^4$, where i represents a generic quark type (u, d, or s herein) comprising the decay products common to the meson and qi represents the basic unit of charge of the given quark type in units of the proton charge, i.e., $q_u = 2/3$, $q_d = 1/3$, and $q_s = 1/3$.

Hence, the conventionally assumed quark structure of the $K^*(892)^0$ and the $K^*(892)^{\pm}$ can be used for GEM width formulas, the theoretical results then compared with experimental determinations. We will find significant discrepancies with "conventional wisdom", to be sure. However, GEM will allow for a new picture as to the structure of the $K^*(892)$ one that is as simple as it is reasonable.

13.2 Application of GEM

From White (2008), Eq. 4, the GEM formula governing the width, $\Gamma_v(GEM)$, of any conventional vector meson, v, of mass, m_v, in its ground state is given by:

$$\Gamma_v(GEM) \approx \left(\frac{m_\rho}{m_v}\right)^3 \left(\sum_i q_i^4\right) \left[\ln \frac{m_v}{50\,MeV}\right]^{-1} \times 1960\,MeV\,, \qquad (13.1)$$

where $m_v = 776\,MeV$ is the mass of the ρ meson (PDG (2004), p. 4). As noted in White (2008), however, the $K^*(892)$ is not a "conventional" vector meson, because it has an isospin value of 1/2, and so unlike the "conventional" vector mesons such as the ρ, the ϕ, the J, and the υ, half of the $K^*(892)$'s energetically allowed decay routes are forbidden, i.e., decays into a pion and a kaon are allowed, but decays into pion pairs are not.[5] For the time being, considering the $K^*(892)$ as a composite structure of charged and neutral modes of average mass, $894\,MeV$, we postulate that its width is given by Eq.(13.1)

[3] For literature as to the roots of the Gluon Emission Model and its early application to hadron production phenomena see Close (1979) and White (1985).

[4] See White (2008) and Dalitz (1977).

[5] For a discussion on isospin, see Ohanian (1987), pp. 439-441.

above with the sum over q_i^4 as $q_u^4 + q_d^4 + q_s^4 = 18/81$, except that we must multiply the right hand side by 18/35, because, as a composite structure, u, d, and s quark types are common both to the meson and its collection of allowed decay products, whereas the forbidden route only involves u and d quarks in the decay products, so that the relevant sum over q_i^4 would be 17/81. Hence, the allowed route is favored over the forbidden one by the factor 18 to 17.[6] Denoting the width of the composite structure as determined by GEM as $\Gamma_C(GEM)$, we find from Eq.(13.1)

$$\Gamma_C(GEM) \approx \frac{18}{35}\left(\frac{776}{894}\right)^3 \frac{18}{81}\left[\ln\frac{894}{50}\right]^{-1} \times 1960\,MeV \approx 50.80\,MeV\ , \quad (13.2)$$

a figure certainly well representative of the experimentally determined width of either the $K^*(892)^0$ or the $K^*(892)^{\pm}$ noted in the Introduction.

Now, building upon the composite structure idea, if we consider the $K^*(892)^0$ as a linear combination of u, d and s quark / anti-quark pairs much in keeping with the description of the ρ as a linear combination of uu^* and dd^* objects, or of the ϕ as a linear combination of ss^* (forming its kaon branch), uu^* and dd^* (together forming its non-kaon branch) objects whose decay products bear no net charge, we can invoke GEM to obtain the theoretical width of the $K^*(892)^0$ under the above assumption. Specifically, denoting its GEM-determined width as $\Gamma_0(GEM)$, we find:

$$\Gamma_0(GEM) \approx \frac{18}{35}\left(\frac{776}{896}\right)^3 \frac{18}{81}\left[\ln\frac{896}{50}\right]^{-1} \times\ 1960\,MeV \approx 50.42\,MeV\ , \quad (13.3)$$

a figure well in accord with the experimental finding of $50.7 \pm 0.6\,MeV$ noted in the Introduction.

Proceeding similarly with the $K^*(892)^{\pm}$, whose decay products would bear, of course, one unit of charge and denoting its GEM-determined width as $\Gamma_{\pm}(GEM)$, we find:

$$\Gamma_{\pm}(GEM) \approx \frac{18}{35}\left(\frac{776}{892}\right)^3 \frac{18}{81}\left[\ln\frac{892}{50}\right]^{-1} \times\ 1960\,MeV \approx 51.18\,MeV\ , \quad (13.4)$$

a figure also well in accord with the experimental finding of $50.8 \pm 0.9\,MeV$ noted above in the Introduction.

Obviously, to consider the $K^*(892)^0$ as a construction containing only the (conventionally assumed) d and s quark types would not lead to an accurate

[6] See White (2008), Section IV.

width determination of it via GEM by any stretch of the imagination. The term involving the sum over quark charges would be only 2/81, thus producing the corresponding "weight factor" (scaling the allowed decay route) of $2/(2+1) = 2/3$, since only the d quark type would be common to both the forbidden decay products (pions) and the meson. The resulting width of the $K^*(892)^0$ as determined by GEM would thus be only $7.26\,MeV$. Neither would considering the $K^*(892)^{\pm}$ as a construction containing only the (conventionally assumed) u and s quark types give via GEM a determination of its width as accurate as seen in Eq.(13.4). Here, the term involving the sum over quark charges would be (17/81), thus producing a "weight factor" of $17/(17+16) = 17/33$, since, here, only the u quark type would be common to both the forbidden decay products (again, pions) and the meson. The resulting width of the $K^*(892)^{\pm}$ as determined via GEM would be $48.42\,MeV$, a figure lying outside the range of experimental uncertainty. The most reasonable and, in our view, correct assumption as to the actual structure of the $K^*(892)$, as indicated by GEM, therefore, would be that it comprises a composite structure of uu^*, dd^*, and ss^* in equal measure.

13.3 Discussion of Results

The structure indicated by GEM for the $K^*(892)$ charged mode or uncharged mode – i.e., a linear combination of uu^*, dd^*, and ss^* in equal measure, we will represent by

$$\chi = \frac{1}{\sqrt{3}}\,(uu^* + dd^* + ss^*) \tag{13.5}$$

As such, the $K^*(892)$ would not be characterized as a "strange meson" in the usual sense, as said term usually denotes a structure of the form, $\chi_s = sx^*$ (or s^*x), where "x" denotes a given type of quark other than "s". Indeed, χ is much more akin to the theoretical structures of the light, unflavored vector mesons, such as the ρ and the ϕ. If one assumes the decay of χ takes place via s (or s^*) joining either u^* (or u) or d^* (or d) on a random basis, all possible $K\,\pi$ decay products would show up with equal probability, such in keeping with the widths of the charged mode and neutral mode being essentially equal. Under our assumptions it is thus seen that all possible $K\,\pi$ products are realized in all possible permutations with equal likelihood consistent with experimental findings. Hence, it appears eminently evident that the $K^*(892)$ is not "strange" at all. Rather, it clearly should be regarded as just another one of the garden variety light, unflavored vector mesons joining ρ, ω, and ϕ as the third most

massive in the group of four such objects.

13.4 Concluding Remarks

In a sense, GEM has provided for the discovery (perhaps "uncovery" would be a better word) of a "missing link" between the ρ/ω and the ϕ. The ρ and ω, for instance, are theoretically constructed from linear combinations of uu^* and dd^* structures, and, in keeping with that, give rise to various π decay products (no Ks). The ϕ, on the other hand, is thought to be a linear combination of uu^*, dd^*, and ss^* objects, similar to our suggested construction of the $K^*(892)$ (see Eq.(13.5)), but the ϕ is massive enough to decay into two kaons, which it does predominately. So, the predominant decay mode of the ϕ is characterized by s joining u^* and the associated s^* joining u, for example, so that two Ks can be emitted. Where is the vector meson less massive than the ϕ but more massive than the ρ/ω which can give rise to the intermediate type of decay, i.e., a π plus a K? It's quite a mystery once one begins to think about it! However, much as Arthur Conan Doyle's famous purloined letter, we believe the missing meson can be found in plain view for anyone who really looks for it: It's right there on p. 28 of *PDG* (2004) for all to see.

Epilogue: Personal Comments by Dave White

Thanks to Tom Erber I received my very first job opportunity in 1972 as a part-time instructor at IIT while still finishing up my PhD thesis on the quantum mechanical description of synchrotron radiation by electrons in uniform magnetic fields. But it was more than that; it was a place to be ... a place to learn ... to learn how to write for journals ... to learn how to collaborate ... to learn to share ideas ... to learn about things other than my prime interest ... things such as the quantum shifting of many-magnet systems, the generation of quasi-random numbers via simple polynomials, and the few-loop theorem in random number generation, which still fascinates me to this day. The place to be was a desk dedicated to me in Tom's magnet lab, a desk I frequented most days a week, including lots of Saturdays (following Tom's example), for eight years, carrying me to the fall of 1980, at which time I obtained my first tenure-track appointment at Roosevelt University. I'll never forget the sight of Tom (1973) almost running down the hall towards the magnet lab, he calling "White! White!", my first draft of my first single-authored physics paper, subsequently published by Physical Review D, waving wildly in his right hand. "'This' is

for speaking," he said. "Huh?" of course was my reply. "You have too many 'this-es' in here; 'this' is a word you use when you can point to what is meant by 'this'. You don't use 'this' in anything meant for the written communication of scientific ideas. And by the way ... you have a 'soley' in here, too."

Today, believe me, I know that one "this" is one "this" too many in formal scientific writing, something that I have passed on to all my physics students at Roosevelt since. To this day, Tom is still the best pure scientist I have ever known. He always insisted that things be done for the right reason ("we don't do it for the fame"). I learned to "do it" because I thought it was on a path towards "the truth" and because it was fun to embark on such a path. Along those lines I never met anyone so all business in the pursuit of fun stuff to pursue. I lost the sense of fun in the pusuit of physics for one score years (1983 - 2003), during which time I wrote science fiction for the pure, all business fun of it, and such endeavor I see now as being necessary to get me from some measure of emotional instability towards a better outlook on life. Now, I am "doing it" with gusto ... having tons of fun with it ... but, maybe with ... oh ... 15.6% of the "business" exhibited by my mentor over the entirety of his teaching career (sorry, Tom). The last paper I published before my "hiatus from physics" was an extension of the work I had been doing regarding hadron production via virtual synchrotron radiation. I'll never forget, either, how a major feature of said extension, viz., F. Close's Gluon Emission Model, found its way into that 1985 IJTP publication. I was at my desk in the magnet lab one day circa 1982[7] when Tom walked in with a load of books and papers in arm, a load which, in the manner of the Japanese tea ceremony (more business + fun all wrapped together), he "placed/dumped" on my desk. "I thought you might want to take a look at this; I think you'll find it interesting," he said. "Close has a real interesting extension to QED; you might be able to work it into your presentation. And by the way, you could renormalize the absorption cross-section formula to a width, you know; you might then obtain a standard as a check to your synchrotron results." Boy, did I have fun with that unsolicited pile of information! There was Close's text containing the Gluon Emission Model ... also a small book by Feynman on quarks and gluons, in which various QED processes were "analogized". I went straight to Merzbacher to find the absorption cross-section formula associated with quantum states at rest which decay to the ground state and saw a way to "analogize" the QED situation to a width formula for vector mesons, as well. And ... in 2004 that's exactly what I picked up on again. To my amazement, the meson data in 2004 had become not only voluminous, but also each vector

[7] I would still venture to my desk in the magnet lab on weekends until the spring of 1984.

meson data set in the PDG's meson table had become incredibly more "tight" (precise) than was evidenced in 1983. Coupled with the personal computer with its internally-borne $EXCEL^{TM}$ spreadsheet programs, getting back into physics again had all of a sudden become interesting and fun again. In a non-trivial way, then, Tom's guiding influence had always been present within me, even during my hiatus. I just needed some time for the experimentalists to get with it and start measuring things better and for the folks at $INTEL^{TM}$ to come up with a computer I could carry around with me. So ... thanks, Tom! The paper is an application of Close's Gluon Emission Model ... it was a blast writing it (and no "this-es" will you find therein).

References

[1] PDG "Mesons" (2004), accessed online November 7, 2008 http://pdg.lbl.gov/2004/tables/mxxx.pdf .

[2] White, D. (2008) "The Gluon Emission Model for Hadron Production Revisited", Journal of Interdisciplinary Mathematics, 11 (4), pp. 543-551.

[3] Close, F. (1979) An Introduction to Quarks and Partons, Academic Press.

[4] White, D. (1985) "Calculation of the Strong Coupling Constant, α_s, from Considerations of Virtual Synchrotron Radiation Resulting in Hadron Pair Emission", International Journal of Theoretical Physics, 24 (2), pp. 201-216.

[5] Dalitz, R. H. (1977) "Glossary for New Particles and New Quantum Numbers', Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, 355 (1683), p. 601.

[6] Ohanian, H. C. (1987) Modern Physics, Second Edition, Prentice Hall, pp. 439-441.

14

The Heisenberg-Euler Lagrangian: an Example of an Effective Field Theory

Walter Dittrich[1]
Institute for Theoretical Physics
University of Tuebingen (Germany)

14.1 Introduction

I consider it an honor to contribute an article to the "Festschrift" on the occasion of Professor Tom Erber's eightieth birthday. Although our world lines only crossed a few times in the past, I vividly remember our discussions at UCLA in the nineteen eighties. What impressed me most was Tom's universal, encyclopaedic knowledge, which went far beyond our beloved physics. During his active time he accomplished a series of brilliant works which have left their marks on a large class of students and colleagues alike, especially those working in classical and quantum electrodynamics. It is this field in which we shared a common interest and about which I will reminisce in the following. In particular, I will outline the basic idea which lies behind any effective field theory, taking as an example the Heisenberg-Euler non-linear effective Lagrangian.

The physics of effective actions has grown and diversified tremendously in recent decades. So I thought it would be a good idea to sketch by way of example the techniques of constructing an effective field theory starting from QED. But before I start my contribution: I wish you, Tom, all the best for your future life, scientifically and otherwise. Ad multos annos!

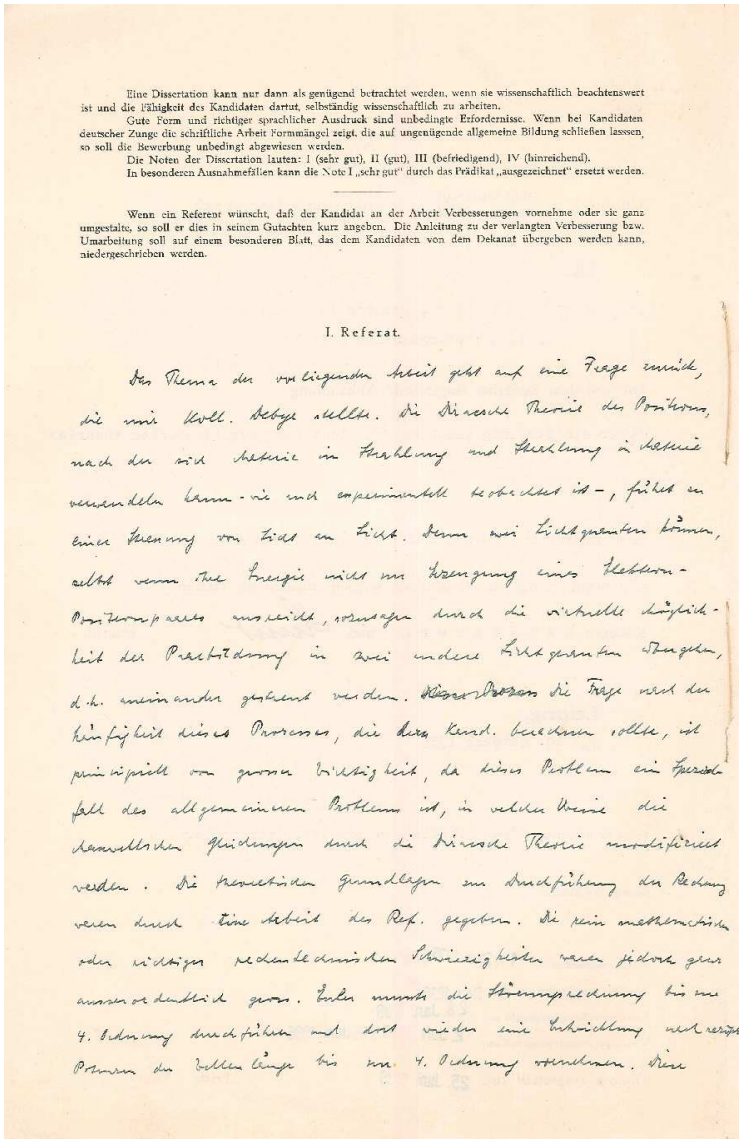
[1]eMail: qed.dittrich@uni-tuebingen.de

14.2 Hans Euler's Ph.D. Thesis: Scattering of Light by Light

Hans Euler's celebrated Ph.D. thesis [1] had its origin in a question that Peter Debye posed to Werner Heisenberg when they were both professors at the university of Leipzig in the mid-thirties. Debye wanted to know whether it would not be possible to calculate the amplitude for the scattering of light by light in the framework of the newly discovered quantum theory of electrons and photons by P.A.M. Dirac. Since light-light scattering implies violation of the superposition principle, this would thus require a modification of Maxwell's equations and hence the replacement of the linear theory of classical electrodynamics by a nonlinear one showing the nonlinear corrections due to Dirac's quantum theory of electrons. This was achieved by Hans Euler.

In the thirties Heisenberg's institute in Leipzig hosted scholars from all around the world. In addition, he had some excellent young, promising students like F. Bloch and E. Teller. However the person most relevant to our subject of photon-photon scattering was his higly talented student Hans Euler. Not much is known about him because he was killed shortly after Nazi Germany started the war against the Soviet Union; more precisely: he was already missing on the east front on June 21, 1941, although the German attack officially started one day later. Attempts after the war to find out more about the fate of Hans Euler with the aid of his physics colleagues and Russian officials were in vain.

Hans Euler's Ph.D. thesis is a masterpiece in handling the calculation of low-energy photon-photon scattering. (For Heisenberg's evaluation of Euler's thesis, see the copy of the original, handwritten document, the German transcription and its English translation below.) His final result, the effective Lagrangian of QED, is presented in a form that is precisely the same as that which was later rederived more elegantly by V. Weisskopf and J. Schwinger.



Eine Dissertation kann nur dann als genügend betrachtet werden, wenn sie wissenschaftlich beachtenswert ist und die Fähigkeit des Kandidaten dartut, selbständig wissenschaftlich zu arbeiten.
Gute Form und richtiger sprachlicher Ausdruck sind unbedingte Erfordernisse. Wenn bei Kandidaten deutscher Zunge die schriftliche Arbeit Formmängel zeigt, die auf ungenügende allgemeine Bildung schließen lassen, so soll die Bewerbung unbedingt abgewiesen werden.
Die Noten der Dissertation lauten: I (sehr gut), II (gut), III (befriedigend), IV (hinreichend).
In besonderen Ausnahmefällen kann die Note I „sehr gut" durch das Prädikat „ausgezeichnet" ersetzt werden.

Wenn ein Referent wünscht, daß der Kandidat an der Arbeit Verbesserungen vornehme oder sie ganz umgestalte, so soll er dies in seinem Gutachten kurz angeben. Die Anleitung zu der verlangten Verbesserung bzw. Umarbeitung soll auf einem besonderen Blatt, das dem Kandidaten von dem Dekanat übergeben werden kann, niedergeschrieben werden.

I. Referat.

Das Thema der vorliegenden Arbeit geht auf eine Frage zurück, die mir Voll. stellte. Die diracsche Theorie des Positrons, nach der sich Materie in Strahlung und Strahlung in Materie verwandeln kann - wie auch experimentell festgestellt ist -, führt zu einer Streuung von Licht an Licht. Denn zwei Lichtquanten können, selbst wenn ihre Energie nicht zur Erzeugung eines Elektron-Positronpaares ausreicht, vorübergehend durch die virtuelle Möglichkeit der Paarbildung in zwei andere Lichtquanten übergehen, d.h. aneinander gestreut werden. Diese Frage nach der Häufigkeit dieses Prozesses, die Herr Kemd. berechnen sollte, ist prinzipiell von großer Wichtigkeit, da dieses Problem ein Spezialfall des allgemeineren Problems ist, in welcher Weise die klassischen Gleichungen durch die diracsche Theorie modifiziert werden. Die theoretischen Grundlagen zur Durchführung der Rechnung waren durch eine Arbeit des Ref. gegeben. Die rein mathematischen oder richtiger rechentechnischen Schwierigkeiten waren jedoch ganz außerordentlich groß. Euler mußte die Störungsrechnung bis zur 4. Ordnung durchführen und dort wieder eine Entwicklung nach reziproken Potenzen der Energie bis zur 4. Ordnung vornehmen. Diese

Figure 14.1: Copy of Heisenberg's recommendation of Euler's PhD thesis (*kindness of the Universitätsarchiv Leipzig*)

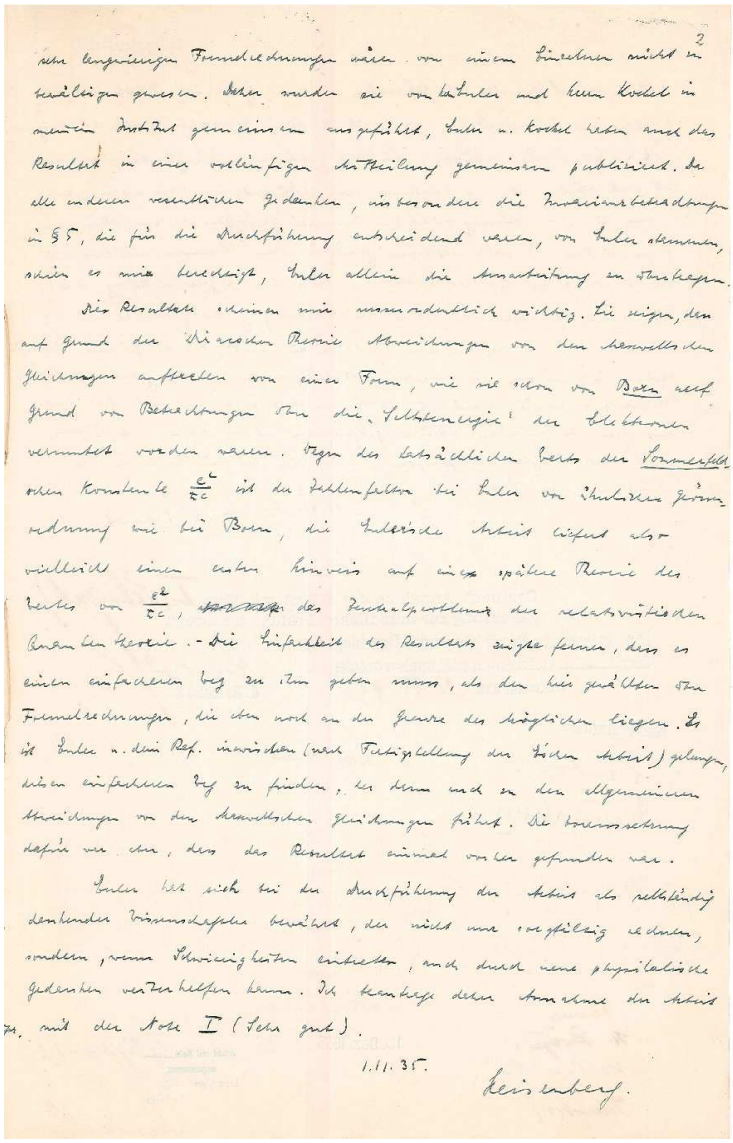


Figure 14.2: Copy of Heisenberg's recommendation of Euler's PhD thesis (*kindness of the Universitätsarchiv Leipzig*)

English Translation of Heisenberg's Evaluation of Euler's PhD Thesis

The topic of the present thesis goes back to the question that my colleague Debye once asked me. Dirac's theory of the positron, according to which matter can be transformed into radiation and radiation into matter - as can be observed experimentally - leads to a scattering of light by light. Two light quanta can, even if their energy is not sufficient to create an electron-positron pair, be transformed into two other light quanta through the - so to speak - virtual possibility of pair building, i.e., can be scattered by each other. The question of how often these processes occur, which was to be calculated by the candidate, is principally of great importance, since the problem is a special case of the more general problem of how Maxwell's equations are modified by Dirac's theory. The theoretical foundations for doing the calculations were laid by the work of the referee [*Heisenberg*]. The purely mathematical difficulties in performing the calculations were, however, formidable. Euler had to use perturbation calculation to 4^{th} order and then make an expansion to 4^{th} order in the reciprocal wavelength. Such a lengthy process of solving of equations could not be accomplished by a single person. This was done at my institute by Mr. Euler together with Mr. Kockel. Euler and Kockel also conjointly published the result in a preliminary publication. Since all the other significant ideas are attributable to Euler, in particular the invariance discussions in §5 which were decisive for the execution, it seemed to me only justified to assign Euler alone to work it out. The results seem to me to be extremely important. They show, on the basis of Dirac's theory, that deviations from Maxwell's equations appear in a form similar to that already conjectured by Born on the basis of considerations of the "self energy" of the electron. Due to the actual value of Sommerfeld's constant $e^2/\hbar c$, Euler's numerical factor is of the same size as Born's. Thus, Euler's thesis possibly gives a first indication of a later theory of the value of $e^2/\hbar c$, the main problem in relativistic quantum theory. – The simplicity of the results further showed that there must be a more straightforward way than that chosen here, solving formulas that lie at the limit of what is possible. Meanwhile Euler and the referee (after finishing Euler's work) were able to find a simple path that finally also leads to the general deviations from Maxwell's equations. The prerequisite for this was, however, that the result had been found once before. In accomplishing this work, Euler proved himself an independently thinking scientist, able not only to calculate meticulously but also, when confronted with difficulties, to surmount them with new physics ideas. I therefore advocate acceptance of this work with the grade 1 (very good).

1.11.35
Heisenberg

14.3 Photon-Photon Scattering, General Remarks

The quantum theory of photon-photon scattering in its full generality is a rather complicated theory. It is, however, not my intention to consider photon-photon scattering experiments over every possible energy regime. Having in mind the creation of an effective purely electromagnetic classical field theory, I will limit myself to photons in the incoming state with momentum transfer much less than the electron mass. Most importantly, since we do not observe any electrons, we integrate out - in the path-integral sense - the electron fields (heavy degrees of freedom) which influence the dynamics of the "light" fields, the photon fields, or the classical electromagnetic fields in our case. Clearly, the effective action should give a complete description of the dynamics of the photon field without reference to the unobserved electron field. All of the interesting low-energy effects of the fundamental spinor QED will be reproduced, while the high-energy behaviour becomes distorted. This is the price of making low-energy photon physics calculable.

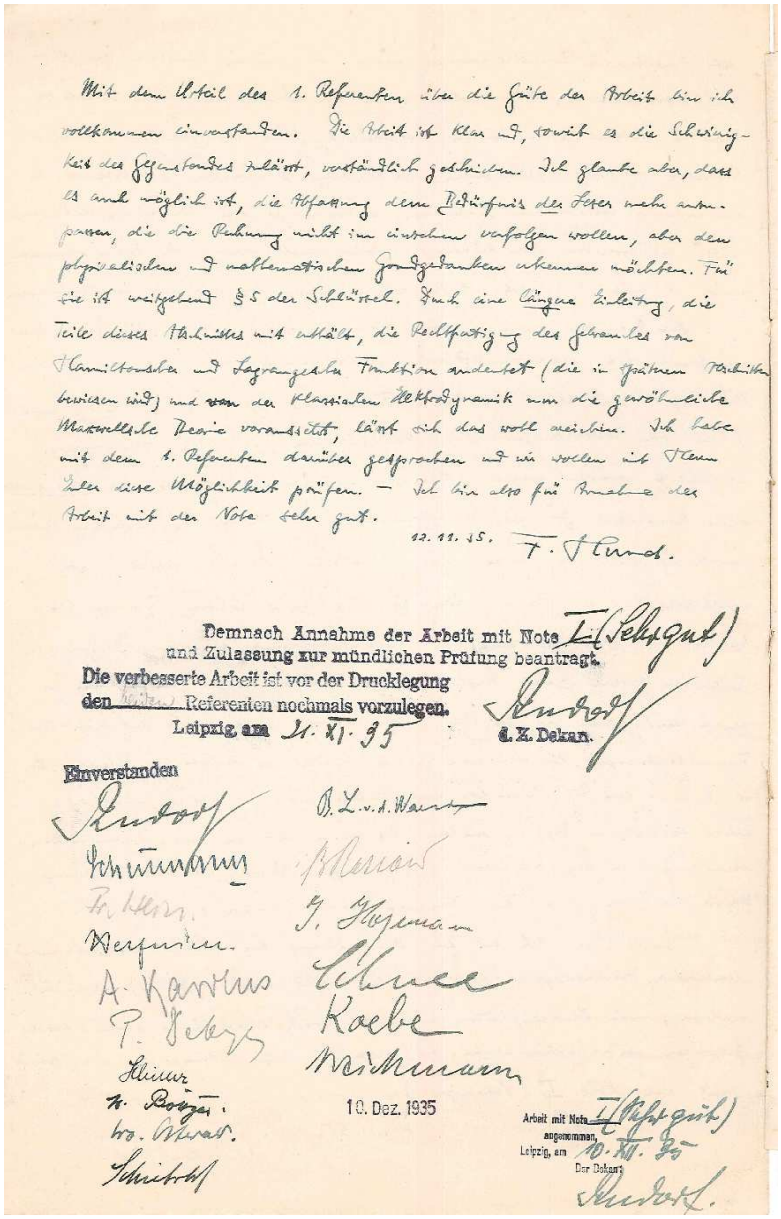
The dynamics of self-interacting low-energy photons is determined by a Lagrangian which is assumed to be hermitean, invariant under Lorentz, gauge, charge conjugation and parity transformations. The general Lagrangian therefore reads

$$\mathcal{L}\left(\mathcal{F},\mathcal{G}\right) = -\mathcal{F} + a\,\mathcal{F}^2 + b\,\mathcal{G}^2 + \dots \tag{14.1}$$

where we have introduced the gauge and Lorentz invariants of the electromagnetic field:

$$\begin{aligned} \mathcal{F} &= \frac{1}{4}\,F_{\mu\nu}\,F^{\mu\nu} = \frac{1}{2}\left(\vec{B}^2 - \vec{E}^2\right), \\ \mathcal{G}^2 &= \left(\frac{1}{4}\,F_{\mu\nu}\,\tilde{F}^{\mu\nu}\right)^2 = \left(\vec{E}\cdot\vec{B}\right)^2, \qquad \tilde{F}^{\mu\nu} = \frac{1}{2}\,\epsilon^{\mu\nu\lambda\sigma}\,F_{\lambda\sigma}\,. \end{aligned}$$

Euler, Kockel and Heisenberg [2] and, independently, Weisskopf [3] were the first to discuss an effective Lagrangian of the type (14.1) and determine the coefficients a and b. If we had no knowledge of standard QED, we could take (14.1), representing a phenomenological theory of interacting photons, calculate various Green's functions and scattering amplitudes and then adjust the unknown constants a and b to the measured photon-photon scattering cross setions at



Mit dem Urteil des 1. Referenten über die Güte der Arbeit bin ich vollkommen einverstanden. Die Arbeit ist klar u. soweit es die Schwierigkeit des Gegenstandes erlaubt, verständlich geschrieben. Ich glaube aber, dass es auch möglich ist, die Fassung dem Bedürfnis des Lesers mehr anzupassen, die die Rechnung nicht im einzelnen verfolgen wollen, aber den physikalischen u. mathematischen Grundgedanken erkennen möchten. Für sie ist wichtig §5 der Schlüssel. Auch eine längere Einleitung, die Teile dieses Abschnittes mit enthält, die Reduktion des Gebrauches von Hamiltonscher und Lagrangescher Funktion ändert (die in jüngeren Arbeiten bewiesen ist) und von der klassischen Elektrodynamik nur die gewöhnliche Maxwellsche Theorie voraussetzt, lässt sich das wohl erreichen. Ich habe mit dem 1. Referenten darüber gesprochen und wir wollen ihn Herrn Euler diese Möglichkeit prüfen. — Ich bin also für Annahme der Arbeit mit der Note sehr gut.

12. 11. 35. F. Hund.

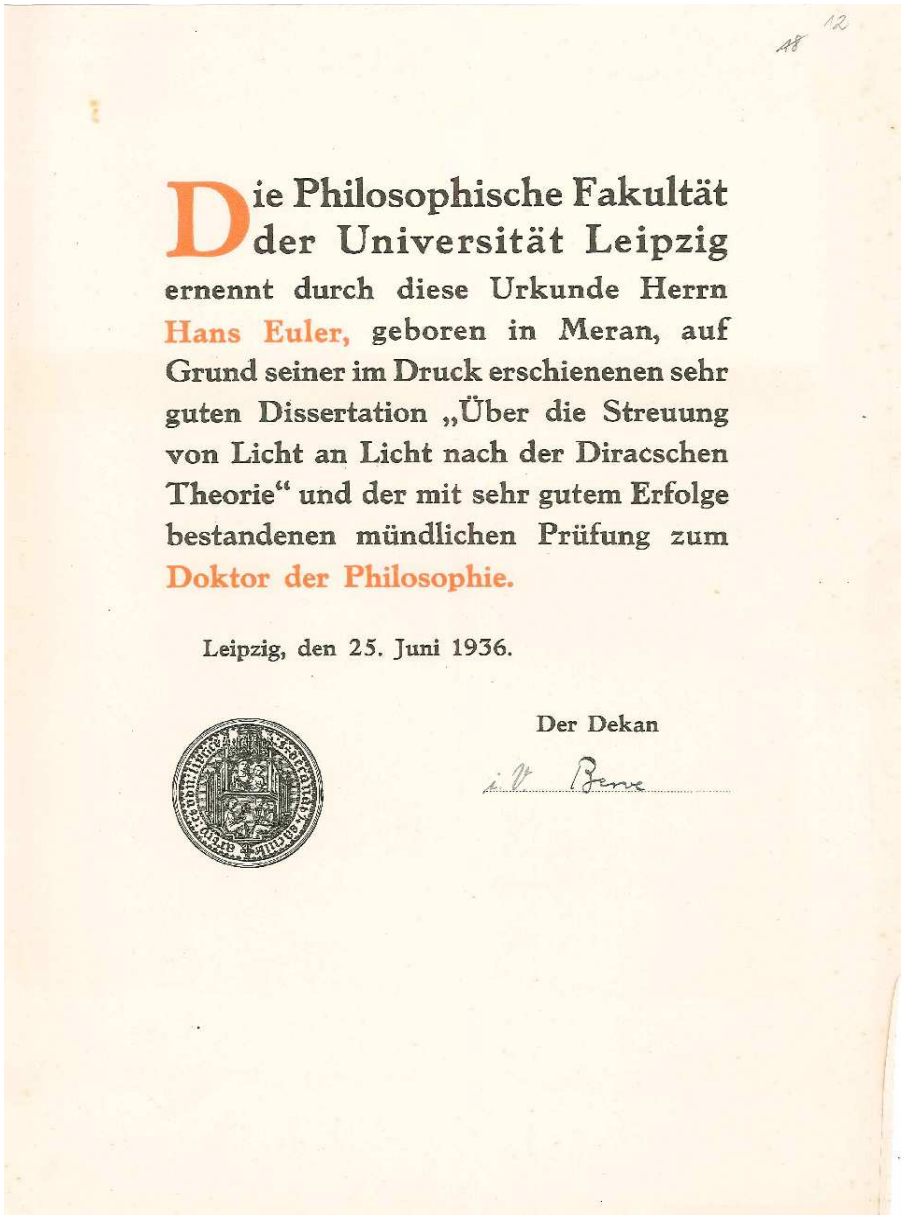
Demnach Annahme der Arbeit mit Note I (sehr gut)
und Zulassung zur mündlichen Prüfung beantragt.
Die verbesserte Arbeit ist vor der Drucklegung
den Referenten nochmals vorzulegen.
Leipzig, am 11. XI. 35
d. Z. Dekan.

Einverstanden

10. Dez. 1935

Arbeit mit Note I (sehr gut)
angenommen,
Leipzig, am 10. XII. 35
Der Dekan

Figure 14.3: Friedrich Hund's affirmation of Heisenberg's recommendation of Euler's PhD thesis and signatures of colleagues (*kindness of the Universitätsarchiv Leipzig*)



Die Philosophische Fakultät
der Universität Leipzig
ernennt durch diese Urkunde Herrn
Hans Euler, geboren in Meran, auf
Grund seiner im Druck erschienenen sehr
guten Dissertation „Über die Streuung
von Licht an Licht nach der Diracschen
Theorie“ und der mit sehr gutem Erfolge
bestandenen mündlichen Prüfung zum
Doktor der Philosophie.

Leipzig, den 25. Juni 1936.

Der Dekan

Figure 14.4: Doctoral certificate awarded to Hans Euler by the Universität Leipzig (*kindness of the Universitätsarchiv Leipzig*)

low energy. Fortunately, we possess a fairly good knowledge of QED in quite a range of energy, which enables us to compute a and b from first principles. To achieve this we merely have to require that the Lagrangian of fundamental QED and the effective Lagrangian (14.1) yield the same results at low energy. Ensuring that the predictions following from the effective theory (14.1) agree with the fundamental theory to any order of accuracy is called *matching*. What we are matching are transition amplitudes or Green's functions in both theories. These techniques are common to all those perturbatively renormalizable field theories and their respective effective low-energy theories in which the heavy degrees of freedom are integrated out.

14.4 One-Loop Effective Lagrangian in Spinor-QED

The simplest Feynman graph that leads to photon-photon scattering in QED is given in Fig. 14.5:

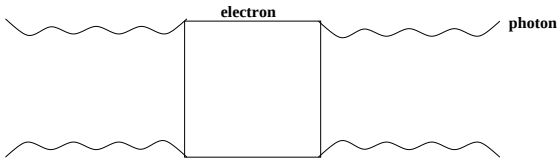


Figure 14.5: Box graph for $\gamma - \gamma$ scattering to 4^{th} order

Assuming low frequency $\gamma - \gamma$ scattering, i.e., treating the external photons as external prescribed fields, we can, more generally, study the electron loop to all orders in the external field, which is sketched in Fig. 14.6:

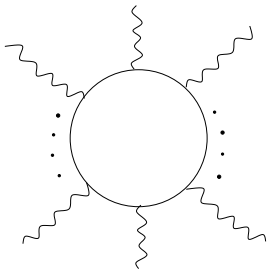


Figure 14.6: Electron loop in external field to all orders

Hence the central object we shall be interested in is the vacuum-to-vacuum persistence amplitude

$$\langle 0_+|0_-\rangle^A = \exp\left\{i\,W^{(1)}\,[A]\right\}\,.$$

Integrating out the unobserved electron degrees of freedom leads to the one-loop contribution $W^{(1)}$ in the effective action:

$$\exp\left\{i\,W^{(1)}\,[A]\right\} = \int d\psi\,d\bar{\psi}\,\exp\left\{-i\int d^4x\,\bar{\psi}\left[\gamma\cdot\left(\frac{1}{i}\,\partial - eA\right) + m\right]\psi\right\}\,. \tag{14.2}$$

The Gauß-type integral in (14.2) with anticommuting classical fields ψ, $\bar{\psi}$ can be evaluated with the result

$$\exp\left\{i\,W^{(1)}\,[A]\right\} = \det\left[\gamma\left(\frac{1}{i}\,\partial - eA\right) + m\right] = \det\left(G\,[A]^{-1}\right)\,.$$

This yields

$$W^{(1)}\,[A] = i\,\ell n\,\det\,G\,[A] = i\,Tr\,\ell n\,G\,[A]\,.$$

Because action functionals are defined only up to a constant, we may exploit this freedom to write

$$W^{(1)}\,[A] = i\,Tr\,\ell n\,(G\,[A]\,/G\,[0])\,. \tag{14.3}$$

Here, $G\,[0] = G$ is the electron propagator in the field-free case, connected with $G\,[A]$ by

$$G\,[A] = G\,(1 - e\gamma\cdot A\,G)^{-1}\,.$$

Tr indicates the trace in both spinor and configuration space.

To explicitly calculate the effective Lagrangian for the effect which an arbitrary number of external photon lines can have on a single electron loop, I have at least half a dozen strategies at my disposal. Some of them can be looked up in ref. [4].

Limiting ourselves to a constant magnetic field B in z-direction, we obtain

$$\mathcal{L}^{(1)}\,(B) = \frac{1}{8\pi^2}\int_0^\infty \frac{ds}{s^3}\,e^{-im^2s}\left[(eBs)\cot\,(eBs) + \frac{1}{3}\,(eBs)^2 - 1\right]\,. \tag{14.4}$$

The integral in (14.4) can be explicitly computed, e.g., by dimensional or ζ-function regularization with the result

$$\begin{aligned}\mathcal{L}^{(1)}(B) &= -\frac{1}{32\pi^2}\left\{\left(2m^4-4m^2(eB)+\frac{4}{3}(eB)^2\right)\left[1+\ell n\left(\frac{m^2}{2eB}\right)\right]\right.\\ &\left.+4m^2(eB)-3m^4-(4eB)^2\zeta'\left(-1,\frac{m^2}{2eB}\right)\right\}. \qquad (14.5)\end{aligned}$$

The result (14.4), which was obtained for a pure magnetic field, can be generalized for arbitrary constant external electromagnetic fields. One only has to substitute for B the gauge-invariant Lorentz scalars $\mathcal{F}$ and $\mathcal{G}^2$. The resulting expression was already achieved by Euler and Heisenberg in 1936.

For a derivation using the elegant Fock-Schwinger proper-time method one should consult Schwinger's superb paper of 1951 [5] or reproduce the corresponding pages of his monograph [6] on source theory.

The result of the effective Lagrangian when constant electric and magnetic fields are present simultanously is given then by

$$\begin{aligned}\mathcal{L}^{(1)} &= \frac{1}{8\pi^2}\int_0^\infty \frac{ds}{s^3}\,e^{-im^2s}\left\{(es)^2|\mathcal{G}|\cot\left[es\left(\sqrt{\mathcal{F}^2+\mathcal{G}^2}+\mathcal{F}\right)^{1/2}\right]\right.\\ &\left.\times\coth\left[es\left(\sqrt{\mathcal{F}^2+\mathcal{G}^2}-\mathcal{F}\right)^{1/2}\right]+\frac{2}{3}(es)^2\mathcal{F}-1\right\}. \qquad (14.6)\end{aligned}$$

An expansion of the integral for small values of e corresponds to a weak-field approximation, $F^{\mu\nu}\ll m^2$, and reduces equation (14.6) to

$$\begin{aligned}\mathcal{L}^{(1)} &= \frac{8}{45}\frac{\alpha^2}{m^4}\mathcal{F}^2+\frac{14}{45}\frac{\alpha^2}{m^4}\mathcal{G}^2\ldots\\ &= \frac{2\alpha^2}{45m^4}\left[\left(\vec{E}^2-\vec{B}^2\right)^2+7\left(\vec{E}\cdot\vec{B}\right)^2\right]. \qquad (14.7)\end{aligned}$$

At last we have identified our unknown constants a and b in (14.1).

Comparing our ansatz (14.1) with the result (14.7), we can write down the explicit values for the constants a and b, namely

$$a=\frac{2\alpha^2}{45m^4}, \qquad b=7a. \qquad (14.8)$$

14.5 Photon-Photon Scattering: More Details and Outlook

The matching procedure of the previous chapter that set in relation the Lagrangian of pure QED to the phenomenological Lagrangian (14.1) was performed in the frame of external-field QED. J. Schwinger, in his monograph [6], has given another way to identify the constants a and b. Using the fundamental QED-Lagrangian, he calculates directly the vacuum persistence amplitude for the photon-photon scattering, the box graph of Fig. 14.5 with the four in- and outgoing on-shell photons.

$$\langle 0_+|0_-\rangle = \exp\left\{ i\, W_{04}^{(1)} \right\} . \tag{14.9}$$

The index 0 appended to the action stands for no external electrons and 4 stands for the two incoming and outgoing photons. He then shows that for low-energy photons the generally non-local photon-photon interaction (form factor for the box) becomes inevitably local and that the vacuum amplitude in (14.9) can be expressed as a space-time integral of a local Lagrange function. In particular, the choice of parallel ($||$) and perpendicular ($\perp$) polarization of the initial and the final photons (the polarization vectors of the photons do not change in the low-energy collision process) yield the following amplitude:

$$\langle 0_+|0_-\rangle = \exp\left\{ i \int d^4x\, \mathcal{L}_{04}^{(1)}\,(x) \right\}$$

with

$$\mathcal{L}_{04}^{(1)}\,(x) = \mathcal{L}_{||}\,(x) + \mathcal{L}_{\perp}\,(x)\ ,$$

where (!)

$$\begin{aligned} \mathcal{L}_{||}\,(x) &= \frac{2\alpha^2}{45m^4}\left(\vec{E}^2 - \vec{B}^2\right)^2 \\ \mathcal{L}_{\perp}\,(x) &= \frac{2\alpha^2}{45m^4}\,7\left(\vec{E}\cdot\vec{B}\right)^2 , \end{aligned} \tag{14.10}$$

which is precisely our earlier matching-result arrived at by using a one-loop external-field calculation.

Given the Lagrangians of (14.10) we are now able to compute various elastic scattering cross sections for the collision process, Fig. 14.7.

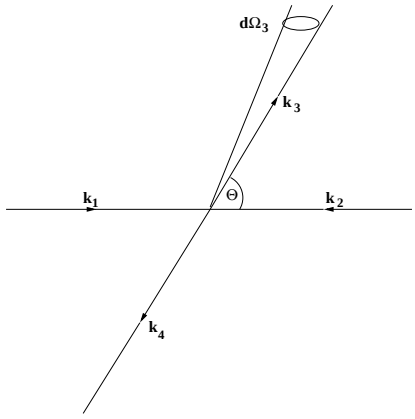


Figure 14.7: Elastic $\gamma - \gamma$ scattering in C.M. system

Since generally for polarized photons

$$\left(\frac{d\sigma}{d\Omega}\right)^{\gamma\gamma}_{\text{pol}} = \frac{1}{(8\pi)^2}\,\alpha^4\,\frac{1}{\omega^2}\,|M|^2 = r_0^2\left(\frac{\alpha}{2\pi}\right)^2\,\frac{m^2}{16}\,\frac{1}{\omega^2}\,|M|^2\,. \tag{14.11}$$

we obtain for non-forward scattering ($\Theta \neq 0$) $\begin{array}{cccccc} \uparrow & \uparrow & \rightarrow & \uparrow & \uparrow \\ 1 & 2 & & 3 & 4 \end{array}$

$$M_{||\,||\,||\,||} = \frac{16}{\alpha^2}\,4\omega^4\,(2a+b)\left(3+\cos^2\Theta\right)\;,\; 2a+b = \frac{4}{180}\,\frac{\alpha^2}{4m^4} = \left(\frac{16}{45}\right)\left(\frac{\omega}{m}\right)^4\left(3+\cos^2\Theta\right)\,. \tag{14.12}$$

Hence, according to (14.11) we find

$$\left(\frac{d\sigma\,(\omega,\Theta)}{d\Omega}\right)_{||\,||\,||\,||} = r_0^2\left(\frac{\alpha}{2\pi}\right)^2\,\frac{16}{(45)^2}\left(\frac{\omega}{m}\right)^6\left(3+\cos^2\Theta\right)^2$$

where $r_0 = \dfrac{\alpha}{m} = 2.8\times 10^{-13}\,\text{cm}$ and $\alpha \approx \dfrac{1}{137}\,,$ (14.13)

and for forward scattering ($\Theta = 0$):

$$\left(\frac{d\sigma\,(0)}{d\Omega}\right)_{||\,||\,||\,||} = \frac{\alpha^2}{m^2}\left(\frac{\alpha}{2\pi}\right)^2\left(\frac{4}{45}\right)^2\left(\frac{\omega}{m}\right)^6 \times 4^2 = \left|f_{||}\,(0,\omega)\right|^2$$

with

$$f_{||}\,(0,\omega) = \left(\frac{32}{45}\right)\frac{\alpha^2}{4\pi}\,\frac{\omega^3}{m^4}\,. \tag{14.14}$$

The corresponding quantities for the scattering amplitude $\overset{\uparrow^1}{\swarrow_2} \;\rightarrow\; \overset{\uparrow^3}{\swarrow_4}$ are given by

$$\begin{aligned}
M_{\perp\,||\,\perp\,||} &= \frac{2}{45}\,\frac{1}{m^4}\left(31 + 22\cos\Theta + 3\cos^2\Theta\right) \\
M^{\Theta=0}_{\perp\,||\,\perp\,||} &= \frac{112}{45}\left(\frac{\omega}{m}\right)^4 \qquad (14.15)\\
\left(\frac{d\sigma\,(\Theta,\omega)}{d\Omega}\right)_{\perp\,||\,\perp\,||} &= \left(\frac{\alpha}{2\pi}\right)^2 r_0^2\,\frac{1}{(90)^2}\left(\frac{\omega}{m}\right)^6\left(31 + 22\cos\Theta + 3\cos^2\Theta\right)^2 \\
\Theta = 0: \quad \left(\frac{d\sigma\,(0)}{d\Omega}\right)_{\perp\,||\,\perp\,||} &= \left(\frac{\alpha}{2\pi}\right)^2 r_0^2\left(\frac{56}{90}\right)^2\left(\frac{\omega}{m}\right)^6 = \left|f_\perp\,(0,\omega)\right|^2
\end{aligned}$$

with

$$f_\perp\,(0,\omega) = \left(\frac{56}{45}\right)\frac{\alpha^2}{4\pi}\,\frac{\omega^3}{m^4}\,. \tag{14.16}$$

Given the results (14.14) and (14.16), we find for the index of refraction:

$$n_{\substack{||\\ \perp}} = 1 + \frac{2\pi}{\omega^2}\,N\,f_{\substack{||\\ \perp}}\,(0,\omega) \tag{14.17}$$

where $N = \frac{\perp}{4\omega}\,E^2$ is the average number density of centers of scattering, so that

$$n_{\substack{||\\ \perp}} = 1 + \frac{1}{45}\,\alpha^2\left\{\begin{array}{c} 8 \\ 14 \end{array}\right\}\frac{E^2}{m^4}\,, \qquad \textit{(Kerr effect)} \tag{14.18}$$

and with the substitution $E \rightarrow B$ and $8 \leftrightarrow 14$:

$$n_{\substack{||\\ \perp}} = 1 + \frac{1}{45}\,\alpha^2\left\{\begin{array}{c} 14 \\ 8 \end{array}\right\}\frac{B^2}{m^4}\,. \qquad \textit{(Cotton-Mouton effect)} \tag{14.19}$$

A plane wave field (laser beam) with fields $\vec{e}$, $\vec{b}$ and wave vector $\vec{k}$ passing through a constant B-field and laser polarization $\begin{array}{lll} \perp\text{-mode:} & \hat{e} = \hat{k} \times \hat{B}, & \left|\hat{k}\right| = 1 \\ ||\text{-mode:} & \hat{b} = \hat{k} \times \hat{B} & \end{array}$
yields for the Cotton-Mouton effect a slightly modified formula of (14.19):

$$n_{\substack{|| \\ \perp}} = 1 + \frac{1}{45}\,\alpha^2 \left\{ \begin{array}{c} 14 \\ 8 \end{array} \right\} \frac{B^2}{m^4} \sin^2\Theta\,, \qquad \Theta = \angle\left(\hat{B}, \hat{k}\right)\,. \tag{14.20}$$

Except for propagation along the external B-field direction ($\sin\Theta = 0$), the vacuum polarized by an applied constant magnetic field acts like a birefringent medium [7]. In particular, if we consider light of wavelength λ traversing a path length L normal to the B field, then the ellipticity from vacuum birefringence is given by

$$\psi^{\text{QED}} = \pi\,\frac{L\left(n_{||} - n_{\perp}\right)}{\lambda} = 1.6 \times 10^{-11}\,\text{rad.} \tag{14.21}$$

This theoretical numerical value from spinor QED was evaluated with PVLAS parameters [8].

$$\begin{array}{rcl} B \sim 5T & = & 5 \times 10^4\ \text{G} \\ \ell & = & 1\ \text{m} \\ N & = & 44000\ \text{passes} \\ L & = & 1\ \text{m} \times 4.4 \times 10^4 = 4.4 \times 10^6\ \text{cm} \\ \lambda & = & 1064\ \text{nm} = 1.064\ \text{cm} \\ \hbar\omega & = & 1.2\ \text{eV} \end{array}$$

Let me remark that there are other ways to discuss dispersion effects for low-frequency photons. I have chosen the shortest way so as to limit my contribution to a reasonable number of pages. If the reader is interested in a more elaborate representation, which to a large extent makes use of the brilliant papers by Tom and Wu-yan-Tsai [9], he (she) should consult either their original papers or the monographs [4].

Clearly the number given in (14.21) is one of the smallest numbers ever to be measured in a high-precission experiment. The PVLAS group in the Legnaro laboratory in Italy, formerly under the leadership of the deceased E. Zavattini, has spent many years detecting and measuring this tiny ellipticity. So far they have been able to provide only an experimental upper bound for $\psi^{\text{QED}}_{\text{exp}}$ of the order of 10^{-8} rad.

Before I conclude my contribution to the Festschrift for my colleague and friend Tom Erber, I want to mention that at present there are numerous activities going on worldwide to measure photon-photon scattering directly with the

aid of strong laser fields. To appreciate these formidable tasks, let me remind you of the minute numbers that are to be measured.

For unpolarized photons with $\omega \ll m$ in the C.M. system, it has been known for a long time that for unpolarized photons

$$\left(\frac{d\bar{\sigma}\left(\Theta,\omega\right)}{d\Omega}\right)_{\gamma\gamma} = r_0^2\,\frac{139}{\left(180\right)^2}\,\frac{\alpha^2}{\pi^2}\left(\frac{\omega}{m}\right)^6\left(3+\cos^2\Theta\right)^2\,,$$

which for $\Theta = 0$ and photons of frequency $\omega = 51$ keV yields

$$\left(\frac{d\bar{\sigma}\left(0\right)}{d\Omega}\right)_{\gamma\gamma} = 2.94\times 18^{-38}\,\mathrm{cm}^2$$

In comparison to Compton scattering, we obtain

$$\left(\frac{d\bar{\sigma}\left(0\right)}{d\Omega}\right)_{\mathrm{Compton}} = r_0^2 = 7.95\times 10^{-26}\,\mathrm{cm}^2\,.$$

The total cross section, $\bar{\sigma}_t\left(\omega\right) = \int d\Omega\,\frac{d\bar{\sigma}(\Theta,\omega)}{d\Omega}$, produces for unpolarized light

$$\bar{\sigma}_t^{\gamma\gamma}\left(\omega\right) = \frac{1}{\left(45\right)^2}\,\frac{973}{5\pi}\,\alpha^4\,\frac{1}{m^2}\left(\frac{\omega}{m}\right)^6\,.$$

Limits from ellipsometric data of the PVLAS group can be directly translated to $\gamma-\gamma$ scattering limits.

The total cross section for nonpolarized light with wave length $\lambda = 514.5$ nm is theoretically given by

$$\bar{\sigma}_t^{\mathrm{QED}} = 1.44\times 10^{-67}\,\mathrm{m}^2$$

where the measured upper bound given by the PVLAS experiment [8] is about 2×10^7 larger than predicted from QED. This can only mean that the PVLAS data need to be improved or that another, different experimental design has to be invented.

Maybe the laser group at the university in Jena [10] will surprise us one day with a confirmation of the theoretical results for which the entire QED community has been waiting for so long.

References

[1] H. Euler, Ann. Physik **26** (1936) 398

[2] H. Euler and B. Kockel, Naturwissenschaften **23** (1935) 246
W. Heisenberg and H. Euler, Z. Phys. **98** (1936) 714

[3] V. Weisskopf, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **14** (1936) 1

[4] W. Dittrich and M. Reuter, *Effective Lagrangians in Quantum Electrodynamics*, Lectures Notes in Physics, Vol. 220, Springer, Berlin, Heidelberg (1985)
W. Dittrich, H. Gies, *Probing the Quantum Vacuum*, Springer Tracts in Mod. Phys. **166** (2000) 1

[5] J. Schwinger, Phys. Rev. **82** (1951) 664

[6] J. Schwinger, *Particles, Sources and Fields*, Vol. 2, pp. 123-144, Addison-Wesley, Reading, MA (1973)

[7] R. Baier and P. Breitenlohner, Acta Phys. Austriaca **25** (1967) 212; Nuovo Cimento **A47** (1967) 261
Z. Bialynicka-Birual, Phys. Rev. **D2** (1970) 2341
S.L. Adler, Ann. Phys. (N.Y.) **67** (1971) 599

[8] E. Zavattinin et al., Phys. Rev. **D77** (2008) 032006

[9] W. Tsai and T. Erber, Phys. Rev. **D10** (1974) 492; Phys. Rev. **D12** (1975) 1132; Acta Phys. Austriaca **45**(1976) 245

[10] H. Gies, Eur. Phys. J. **D55** (2009) 311-317

15

For Professor Thomas Erber: in Recognition of His 80^{TH} Birthday

Michael E. Fisher
University of Maryland

Dear Tom:

I was delighted to learn that your colleagues were assembling a Festschift in your honor for this year, the tenth of a new century, when you will celebrate your 80^{TH} Birthday.

As you and I have lived our scientific lives, our paths have crossed only occasionally over the years. But, from decades ago, I still remember vividly our first meeting when you spoke with your characterstic dynamism at one of Joel Lebowitz's "Yeshiva" Meetings, long since transferred to Rutgers but always encouraging the liveliest and most original scientists. And, almost needless to say, your topic was Hysterisis that long-time love of your scientific career.[1] You brought novelty and insights to an old but oft neglected phenomenon. And I remember the magnetic models you invoked that resonated with my own theoretical pictures and modes of thought.

Then you wrote, for my own Festschrift in 1991, a stimulating article with Dave Gavelek revealing the iterative asymptotics of functions generating "random sequences".[2] In truth, "What is a random number?" is a question that first intrigued me in my high school days when I conducted my own 5-card ESP extrasensory perception experiments. My personal scientific love, the lsing

[1] See articles [28], [32-34], [43,44], [73,74], [77] and [84], and also [20] in Tom's Publication List et seq.

[2] See publication [80] and the related articles [40], [58], [61], [63], and [83].

model, had since shown how "nonrandom", indeed, were many computerized random-number generators that people used.

Most recently, only half a decade ago, I was charmed by your Reflections on Parity,[3] by how you brought in the always fascinating problem of "best packings" for your favorite case of interacting charges on a sphere[4] – and finally, how you showed that all was undermined by Smale's inspired discovery of sphere eversion![5] I am still promising myself to make time for an opportunity to hunt down and see the computer animations of the process constructed by Max and by Francis[6] and their coworkers. For my own part, I still prize the play-dough model I made in which the surface of a Möbius strip is deformed so that its edge lies on a circle in a plane. Have you ever seen that?

In the same vein, I hope you will accept as a small token of appreciation, a copy of a "wheel-full" pattern I constructed to adorn a doorway to a gallery my wife set up for a show for local artists. (see Figure 2.) The symmetry of the elements, or their lack thereof, might amuse you.

With my warm best wishes for your next adventures.

Michael

[3] See [95] on the Publication List.
[4] Illustrated by articles [82] and [88].
[5] See Refs. [57,58] in publication [95].
[6] See Refs. [59-63] in publication [95].

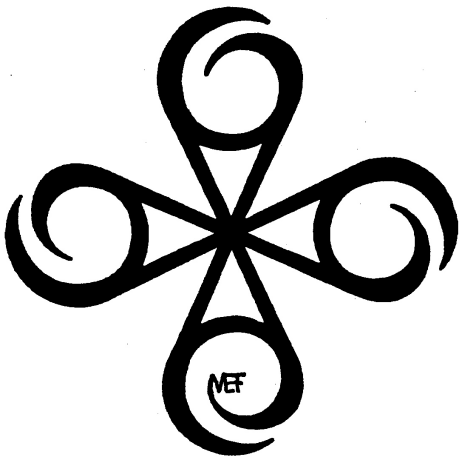


Figure 15.1: Four wheel figure

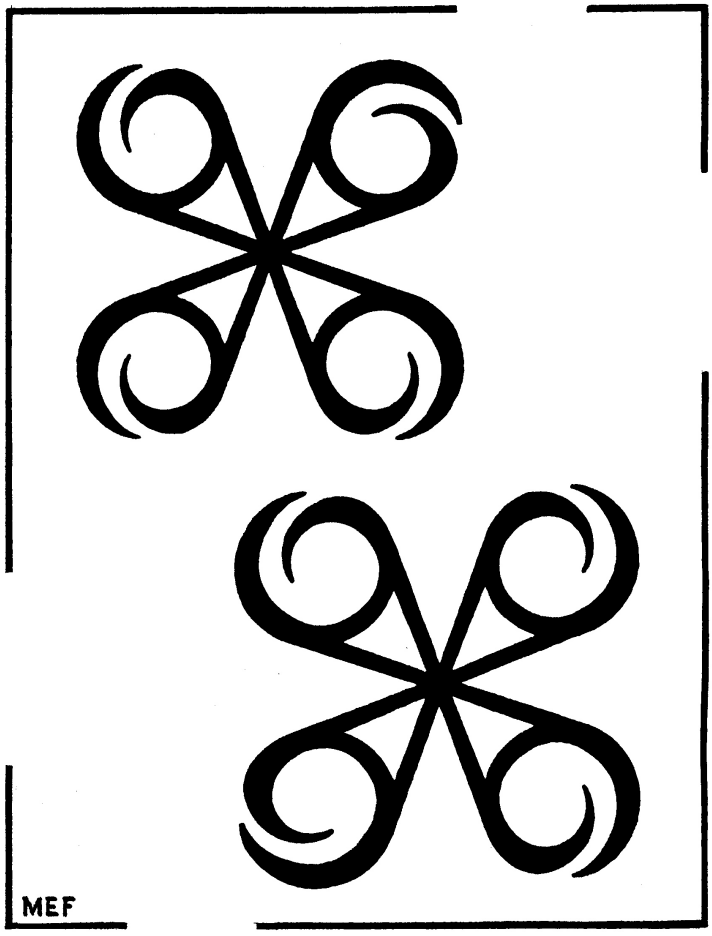


Figure 15.2: “Wheel-full” Pattern

16

Majorana Fermions in Fermi Superfluids: A Pedagogical Note

A. J. Leggett
Department of Physics
University of Illinois at Urbana-Champaign

Abstract

I analyze the concept of a "Majorana fermion" in a Fermi superfluid within the framework of an explicitly particle-conserving formulation, with the conclusion that this apparently exotic object is nothing but a quantum superposition of a real ("Dirac-Bogoliubov") fermion and an operator that annihilates the groundstate identically. I briefly sketch how Majorana fermions can arise in an exactly soluble toy model.

One of Tom Erber's signature contributions over his career has been his ability to extract interesting physics from problems that at first glance might seem to be trivial. It is in this spirit that I offer this note, which asks and attempts to answer a simple question: what exactly is a "Majorana fermion," as introduced in condensed matter physics, in terms of the original electrons?

In the context of quantum field theory, in which the idea was originally defined (see Ref. 1), a "Majorana fermion" is a fermion that is its own antiparticle. Since beyond this there seems to be some ambiguity in the literature about the precise definition (and we shall not need it in the present condensed-matter context) I just note that instead of the "canonical" Fermi commutation

relations

$$\{\psi(\mathbf{x}), \psi(\mathbf{x}')\} = 0, \left\{\psi(\mathbf{x}), \psi^\dagger(\mathbf{x}')\right\} = \delta(\mathbf{x} - \mathbf{x}') \tag{16.1}$$

the creation and destruction operators of a Majorana fermion satisfy the relations

$$\{\psi_M(\mathbf{x}),\ \psi_M(\mathbf{x}')\} \equiv \{\psi_M(\mathbf{x}),\ \psi_M^\dagger(\mathbf{x}')\} = \delta(\mathbf{x} - \mathbf{x}') \tag{16.2}$$

as will be the case below.

In condensed matter physics, the context in which "Majorana fermions" have been most widely discussed is that of an inhomogeneous Fermi system in which Cooper pairs are formed (hereafter for brevity called simply a "Fermi superfluid"). The fermionic excitations of such systems are conventionally described by the Bogoliubov-deGennes (B dG) equations (see Ref. 2); I will discuss these in detail below, but for the present note that within this approach the operator that creates a single fermionic excitation has the general form

$$\gamma_n^\dagger = \int u_n(\mathbf{r})\psi^\dagger(\mathbf{r}) + v_n(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} \tag{16.3}$$

and satisfies the relation

$$\left[\hat{H}_{MF},\ \gamma_n^\dagger\right] = E_n\gamma_n^\dagger \tag{16.4}$$

where $\hat{H}_{MF}$ is the standard mean-field (BdG) Hamiltonian and $\psi(\mathbf{r})$ is the standard (Dirac) fermion field operator. (For simplicity of notation any relevant spin degree of freedom is subsumed in $\mathbf{r}$.) Thus, the elementary fermionic excitations ("Bogoliubov quasiparticles") are quantum superpositions of an extra Dirac fermion with amplitude $u_n(\mathbf{r})$ and a Dirac hole with amplitude $v_n(\mathbf{r})$. Now under certain circumstances it may turn out that there exists a solution of the BdG equations such that $u_n(\mathbf{r}) = v_n(\mathbf{r})$. In this case it follows immediately from (16.3) and (16.4) that for such a solution (which we label with the value 0 of n)

$$\gamma_0^\dagger \equiv \gamma_0,\ E_0 = 0 \tag{16.5}$$

i.e. it has the defining properties of a Majorana fermion (hereafter M.F.) and is usually called by that name. Generally speaking, such solutions are most interesting when they are localized in space (i.e., the functions' amplitude $u_0(\mathbf{r}) \equiv v_0(\mathbf{r})$ falls off exponentially with distance from a particular point); a very widely discussed case in point is the M.F. that is generally believed to be associated with a certain type of vortex in a 2D Fermi superfluid with the so-called "p + ip" pairing (see Refs. 3–5), and which has been proposed[6] as a basis for topologically protected quantum computing in such systems. Rather generally,

one expects the occurrence of M.F.'s to be associated with groundstates with a nontrivial degree of entanglement: see in particular Ref. 7.

To motivate the ensuing discussion, I note the following difference between the properties of M.F.'s as they occur in condensed-matter physics and in particle physics: In the latter, there seems no a priori reason against defining a field $\psi_M(\mathbf{x})$ that satisfies both of Eqns. (16.2); whether or not it originally arose from an ordinary (Dirac) field $\psi(\mathbf{x})$ as in Eq.(16.1) is unimportant. In the condensed-matter case, the operator $\psi_M(\mathbf{r}) \equiv \gamma_0 u_0(\mathbf{r})$ also satisfies both of Eqns. (16.2). However, this operator must in the last resort be defined in terms of real (Dirac) fermions, and the reason it then satisfies both of Eqns. (16.2) is that in the BdG approach the total number of real (Dirac) fermions is conserved only mod 2, because of the "spontaneous breaking of $U(1)$ symmetry," (see e.g. Ref. 8). While the latter is a convenient trick that facilitates calculations, it does not correspond to the true physical situation, and it is perfectly possible and indeed, in the present author's opinion, in the context of conceptual discussions desirable[9], to avoid it by reformulating everything in a completely particle-conserving language. When one does so, the coefficient $u_n(\mathbf{r})$ in Eq.(16.3) is replaced by $u_n(\mathbf{r})\hat{C}^\dagger$, where $\hat{C}^\dagger$ is a normalized creation operator for a Cooper pair (i.e. such that (up to a term of relation order $N^{-1/2}$) $\hat{C}^\dagger|N\rangle = |N+2\rangle$, where $|N\rangle$ indicates the groundstate of the $N(=\text{even})-$particle system). However, when we substitute the modified form of (16.3) with $n = 0$ in the anticommutators, we find that while the second of Eqn. (16.2) is preserved, the first becomes

$$\{\psi_M(\mathbf{r}),\ \psi_M(\mathbf{r}')\} = \delta(\mathbf{r}-\mathbf{r}')\hat{C} \tag{16.6}$$

and thus when projected on to the physical subspace ($N = \text{const.}$) the anticommutator is zero just as for a Dirac fermion. This then prompts the question: What do Majorana fermions look like in a particle-conserving approach?

For orientation let's briefly review, using a particle-conserving representation, the relation between the groundstate and the elementary fermionic excitations (Bogoliubov quasiparticles) in the simple BCS theory of a translationally invariant s-wave Fermi superfluid. Using the abbreviation $\mathbf{k} \equiv (\mathbf{k},\uparrow), -\mathbf{k} \equiv (-\mathbf{k},\downarrow)$ we can write the N-particle groundstate many-body wave function up to normalization in the form (for $N = \text{even}$)

$$\Psi_N = \left(\sum_{\mathbf{k}} c_{\mathbf{k}} a^\dagger_{\mathbf{k}} a^\dagger_{-\mathbf{k}}\right)^{N/2} |\text{vac}\rangle \tag{16.7}$$

where the coefficients $c_{\mathbf{k}}$ are related to the more familiar coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$

by

$$c_k = v_{\mathbf{k}}/u_{\mathbf{k}}, \qquad |u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 = 1 \tag{16.8}$$

For simple s-wave pairing it is possible and convenient to choose both $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ to be real, positive and to satisfy $u_{\mathbf{k}} = u_{-\mathbf{k}},\ v_{\mathbf{k}} = v_{-\mathbf{k}}$. Formally, let us define the operator

$$\hat{C}^\dagger_{(\mathbf{k})} \equiv \mathcal{N}_{(\mathbf{k})} {\sum_{\mathbf{k}'}}' c_{\mathbf{k}'} a^\dagger_{\mathbf{k}'} a^\dagger_{-\mathbf{k}'} \tag{16.9}$$

where the prime on the sum indicates that the state $\mathbf{k}$ is to be omitted, and the normalization $\mathcal{N}_{(\mathbf{k})}$ is such as to ensure that to relative order $N^{-1/2}$ we have $\hat{C}^\dagger_{(\mathbf{k})}\Psi_N = \Psi_{N+2}$. In view of the very weak $O(N^{-1/2})$ sensitivity of $C^\dagger_{(\mathbf{k})}$ to $\mathbf{k}$, we omit the suffix in what follows and treat $\hat{C}^\dagger$ as independent of $\mathbf{k}$. Then, if we focus on the occupation of the pair of states $(\mathbf{k},\ -\mathbf{k})$, the GSWF may be written in the (normalized)[1] form

$$\Psi_N = (u_{\mathbf{k}}|00\rangle_{\mathbf{k}} + v_{\mathbf{k}}|11\rangle_{\mathbf{k}}\hat{C})\,(\hat{C}^\dagger)^{N/2}\,|\text{vac}\rangle \tag{16.10}$$

where $|00\rangle_{\mathbf{k}}$ indicates that both $\mathbf{k}$ and $-\mathbf{k}$ are empty, etc.

Now consider the energy eigenstates of the $(N+1)$-particle system. The simplest such states are of the form

$$\begin{aligned} \Psi^{\mathbf{k}1}_{N+1} &= |10\rangle_{\mathbf{k}}(\hat{C}^\dagger)^{N/2}\,|\text{vac}\rangle \\ \Psi^{\mathbf{k}2}_{N+1} &= |01\rangle_{\mathbf{k}}(\hat{C}^\dagger)^{N/2}\,|\text{vac}\rangle \end{aligned} \tag{16.11}$$

These (normalized) states are generated from the groundstate (16.10) by application of the Bogoliubov quasiparticle creation operators, which in our particle-conserving representation take the form

$$\begin{aligned} \alpha^\dagger_{\mathbf{k}1} &\equiv u_{\mathbf{k}}a^\dagger_{\mathbf{k}} - v_{\mathbf{k}}a_{-\mathbf{k}}\hat{C}^\dagger \\ \alpha^\dagger_{\mathbf{k}2} &\equiv u_{\mathbf{k}}a^\dagger_{-\mathbf{k}} + v_{\mathbf{k}}a_{\mathbf{k}}\hat{C}^\dagger \end{aligned} \tag{16.12}$$

Similarly, the simplest $(N-1)$-particle energy eigenstates are obtained simply by multiplying both of Eqns. (16.12) by an overall factor of $\hat{C}$. Thus, in this simple BCS case there exists a very simple relationship between the structure of the (even-N) groundstate and that of the elementary fermionic excitations.

[1] There is a minor notational problem here, in that (16.10) really only makes sense if we implicitly take $\hat{C} \equiv \hat{C}_{\mathbf{k}}$; the same applies to (16.11) and (16.12).

However, we note that it is also possible to define a second linear combination of $a^\dagger_{\mathbf{k}}$ and $a_{-\mathbf{k}}$ which is orthogonal to the first operator in (16.12):

$$\beta^\dagger_{\mathbf{k}1} \equiv v_{\mathbf{k}} a^\dagger_{\mathbf{k}} + u_{\mathbf{k}} a_{-\mathbf{k}} C^\dagger \tag{16.13}$$

(and similarly for the second equation in (16.12)). This operator simply annihilates the N-particle groundstate identically (that is, it produces a state vector of zero length):

$$\beta^\dagger_{\mathbf{k}1}\Psi_N = \beta^\dagger_{\mathbf{k}2}\Psi_N = 0 \tag{16.14}$$

In many textbook presentations of BCS theory, the operators $\beta^\dagger_{\mathbf{k}1}$, $\beta^\dagger_{\mathbf{k}2}$ are said to create negative-energy fermion states which are supposed to be filled in the groundstate (in analogy to the "filled Dirac sea" of quantum field theory); however, my personal belief is that it is much less confusing to regard them as simply "pure annihilators" as indicated by (16.14). In particular, it is then a trivial observation that any linear combination of pure annihilators is itself a pure annihilator, irrespective of whether or not it corresponds to a (negative-energy) eigenstate of the Hamiltonian.

Now let us turn to the general case of a Fermi system described by the Hamiltonian (where as above we subsume the spin variable in $\mathbf{r}$, and add a term $-\mu\hat{N}$ to allow us to vary the particle number)

$$\begin{aligned}\hat{H} - \mu\hat{N} = \int dr \left\{ \frac{-\hbar^2}{2m}\psi^\dagger(\mathbf{r})\nabla^2\psi(\mathbf{r}) + \big(U(\mathbf{r}) - \mu\big)\psi^\dagger(\mathbf{r})\psi(\mathbf{r}) \right\} \\ + \frac{1}{2}\iint \mathbf{dr}\ \mathbf{dr}'\ V(\mathbf{r},\ \mathbf{r}')\psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r})\end{aligned} \tag{16.15}$$

with $V(\mathbf{r},\ \mathbf{r}') \equiv V(\mathbf{r}',\ \mathbf{r})$. Of course, without arguments specific to the form of the single-particle potential $U(\mathbf{r})$ and the two-particle interaction $V(\mathbf{r},\ \mathbf{r}')$ it is pretty much impossible to tell whether the groundstate of this Hamiltonian will correspond to a normal Fermi liquid, a crystalline solid, a ferromagnet... or many other possibilities. We will, however, be interested in the case where the "topology" of the many-body groundstate wave function for N =even is that of the "completely Cooper-paired" state

$$\Psi_N = \mathcal{N}_N \left\{ \iint \mathbf{dr}\ \mathbf{dr}'\ K(\mathbf{r},\ \mathbf{r}')\psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}') \right\}^{N/2} |\text{vac}\rangle \tag{16.16}$$

where $\mathcal{N}_N$ is an appropriate normalization factor and the function $K(\mathbf{r},\mathbf{r}')$ is antisymmetric under exchange of its indices. While it is very unlikely that the groundstate of any real system is exactly (16.16) (any more than the groundstate of any real translation-invariant metal is given by Eq.(16.5)!), it should be adequate for the qualitative analysis we give below.

We start our analysis with the observation that according to a standard theorem on antisymmetric matrices (proved for example in Ref. 10, appendix A), Eq.(16.16) can always be rewritten in the simpler form

$$\Psi_N = \mathcal{N}'_N \left(\sum_m c_m a^\dagger_m a^\dagger_{\bar{m}} \right)^{N/2} |\mathrm{vac}\rangle \tag{16.17}$$

where the constants c_m are in general complex, and where the operators $a^\dagger_m,\ a^\dagger_{\bar{m}}$ create states $\chi_m(\mathbf{r}),\ \varphi_m(\mathbf{r})$ which together form an orthonormal set, i.e. are such that

$$(\chi_m,\ \chi_{m'}) = (\varphi_m,\ \varphi_{m'}) = \delta_{mm'}, \qquad (\chi_m, \varphi_{m'}) = 0 \tag{16.18}$$

At this point it is amusing to digress for a moment from our main goal and ask: What happens if we try to determine the quantities $c_m,\ \chi_m(\mathbf{r})$ and $\varphi_m(\mathbf{r})$ by optimizing the expectation value of the Hamiltonian (16.16) with respect to them? Let us introduce the notation

$$c_m \equiv v_m/u_m,\ |u_m|^2 + |v_m|^2 = 1, \tag{16.19}$$

$$U_m(\mathbf{r}) \equiv u_m \chi_m(\mathbf{r}), \qquad V_m(\mathbf{r}) \equiv v_m \varphi_m(\mathbf{r}). \tag{16.20}$$

where it is convenient for the moment not to make any particular choice of the phase of u_m. Then it turns out that the result of the suggested procedure is to lead to a pair of equations for $U_m(\mathbf{r})$ and $V_m(\mathbf{r})$ which are exactly of the form of the standard BdG equations—with however, crucially, a set of auxiliary conditions corresponding to (16.18) (e.g. $(U_m(\mathbf{r}), U_{m'}(\mathbf{r})) = 0$ for $m \neq m'$). These conditions are much more stringent than those which follow from the BdG equations themselves and are routinely used in their analysis, namely (with the quantities $u_i(\mathbf{r}),\ v_i(\mathbf{r})$ defined in the standard way)

$$\big(u_i(\mathbf{r}), u_j(\mathbf{r})\big) + \big(v_i(\mathbf{r}), v_j(\mathbf{r})\big) = \delta_{ij} \tag{16.21}$$

Consequently, it is only in the case that the solutions of the BdG equations happen to satisfy the more stringent conditions corresponding to Eq.(16.18) (as in the simple BCS case) that there is any simple relationship between the

structure of the groundstate and that of the fermionic elementary excitations, which as we shall see is described by the BdG equations. In the general case a relationship exists, but is much more complicated, see Ref. 11, appendix A.

Returning to our main theme, let us introduce as in the BCS case a normalized Cooper pair creation operator $C_m^\dagger$, which to order $N^{-1/2}$ is independent of m and thus is written $C^\dagger$ below, use the definitions in Eq.(16.19) and consider, for given m, possible linear combinations of the operators $a_m^\dagger$ and $a_{\bar{m}}$ (or $a_{\bar{m}}^\dagger$ and a_m). Analogously to the BCS case (Eqs. (16.13) and (16.14)) the operators

$$\beta_{m1}^\dagger \equiv v_m a_m^\dagger + u_m a_{\bar{m}} C^\dagger, \qquad \beta_{m2}^\dagger \equiv v_m a_{\bar{m}}^\dagger - u_m a_m C^\dagger \tag{16.22}$$

are pure annihilators of the groundstate, and as in that case any linear combination of β_m's is also a pure annihilator. However, in contrast to the BCS case, the states created by the operators analogous to (12a,b) namely

$$\alpha_{m1}^\dagger \equiv u_m a_m^\dagger - v_m a_{\bar{m}} C^\dagger,\ \alpha_{m2}^\dagger \equiv u_m a_{\bar{m}}^\dagger + v_m a_m C^\dagger \tag{16.23}$$

while they are of course eigenstates of particle number with eigenvalues $N+1$, are not in general energy eigenstates. The operators $\Omega_j^\dagger$ that create the actual energy eigenstates of the $N+1$-particle system are linear combinations of the various $\alpha_{m1}^\dagger$ and $\alpha_{m2}^\dagger$ for different m:

$$\hat{\Omega}_j^\dagger = \sum_m (c_{mj}\alpha_{m1}^\dagger + d_{mj}\alpha_{m2}^\dagger) \tag{16.24}$$

The specific form of the coefficients c_{mj} and d_{mj} may be obtained by solving the equations

$$[\hat{H} - \mu\hat{N},\ \hat{\Omega}_j^\dagger] = E_j \hat{\Omega}_j^\dagger \tag{16.25}$$

with $E_j \geq 0$. When written out explicitly, the equations (16.25) are just the standard BdG equations written in the basis of the $\chi_m(\mathbf{r})$ and $\varphi_m(\mathbf{r})$, so we recover all the standard results for the fermionic excitations of the system; we will call these excitations "Dirac-Bogoliubov" (DB) fermions.

Suppose now that we find a solution of the BdG equations with (if we ignore the factors $C^\dagger$)

$$\hat{\Omega}_j = \hat{\Omega}_j^\dagger \tag{16.26}$$

(so that by (16.25) $E_j = 0$): denote this solution by $j = 0$. Then on substituting (16.24) and (16.25) into this relation, we find the relations, for each value of m,

$$c_{m0}u_m = d_{m0}^* v_m^*, \qquad c_{m0}v_m = -d_{m0}^* u_m^* \tag{16.27}$$

which have no nontrivial solution. Thus we conclude that no solution satisfying (16.26) can describe a pure $N+1$-particle energy eigenstate of the system.

So what has gone wrong? The key observation is that the special case of the BdG equations (16.25) corresponding to $E_j \equiv E = 0$, namely

$$[\hat{H} - \mu\hat{N},\ \hat{\Omega}_0^\dagger] = 0 \tag{16.28}$$

has two possible interpretations: (1) $\Omega_0^\dagger$ creates an energy eigenstate of the $N+1$-particle system with energy(relative to μ) exactly zero, or (2) $\Omega_0^\dagger$ simply annihilates the groundstate! Any solution of (16.28) satisfying (16.26) is easily seen to be a quantum superposition of those two types of solution (since an argument based on Eq.(16.22) and similar so that leading to (16.27) shows that it cannot be a pure annihilator either), with equal weight. The conclusion, therefore, is that the so-called "Majorana fermion" (M.F.) is simply a quantum superposition of a genuine (DB) $E = 0$ fermion and a pure annihilator.

A particularly interesting situation arises where the spinor wave function $\big(u_0(\mathbf{r}), v_0(\mathbf{r})\big)$ of the DB fermion is "split" into two parts localized around mutually distant points in space. In this case the corresponding pure annihilator will be similarly split, and by combining the two solutions of (16.28) with appropriate relative phases we can produce two M.F.'s each of which is entirely localized in one of the two regions. (In most of the literature on M.F.'s, e.g. Ref. (6), this maneuver is performed in reverse: one combines the two M.F.'s with the appropriate relative phase to form a single DB fermion, and the orthogonal combination that forms a pure annihilator is implicitly neglected.) In the last part of this paper I will give a very simple illustration of how this can happen.

The "toy" model I shall give as an illustration of how M.F.'s can arise is almost embarrassingly simple; I strongly suspect that it is a special case of a more general class of model that may well have been studied in the literature in more abstract terms[2], but I believe its pedagogical value is such that it may be worthwhile to explore it explicitly here. It may actually be regarded as

[2]After this manuscript had been submitted I became aware that the special case of this model corresponding to $X_i = \text{const.}$ had been studied in some detail by A. Yu Kitaev, cond-mat/0010440v2 (published in the Proceedings of the Conference on Mesoscopic and Strongly Correlated Electron Systems, Chernogolovka, Russia 2000). (I am grateful to Dr. G. Baskaran for drawing my attention to this.) While the results presented here do not go beyond those of this reference, I hope that some readers may find the somewhat more informal presentation helpful.

the "nearest one can get" to a model of a $(p+ip)$ Fermi superfluid using only one spatial dimension. Consider a set of spinless fermions on a 1D ring lattice labeled by n sites $0, 1, 2...j...n-1$, with sites n and 0 identified, coupled to a "large" superconducting bath with which they can exchange Cooper pairs. In the tight-binding limit a fairly generic Hamiltonian for such a system is

$$\hat{H} - \mu\hat{N} = \sum_{j=0}^{n-1} U_j a_j^\dagger a_j - \sum_{j=0}^{n-1} \left\{ (t_j a_{j-1}^\dagger a_j + \Delta_j a_{j-1}^\dagger a_j^\dagger \hat{C}) + H.c. \right\} \qquad (16.29)$$

where U_j, t_j and Δ_j represent respectively the on-site potential, the tight-binding hopping matrix element and the pairing potential due to the superconducting bath and $\hat{C}$ creates a (normalized) Cooper pair in the bath. It would of course be possible to study this model for arbitrary values of U_j, t_j and Δ_j, but for our purposes it suffices to make the special choice

$$U_j = 0, \qquad t_j = \Delta_j \equiv X_j = \text{real} \qquad (16.30)$$

Thus if we introduce the notation

$$b_j^\dagger \equiv a_j^\dagger - a_j, \qquad c_j^\dagger \equiv a_j^\dagger + a_j, \qquad \hat{K}_j \equiv b_{j-1}^\dagger c_j^\dagger \qquad (16.31)$$

and take it as read that the appropriate number of C's and $C^\dagger$'s are implicitly added to the various terms to ensure conservation of total (system plus bath) particle number, (i.e. in effect we follow the BCS particle-nonconserving convention), the Hamiltonian (16.29) can be rewritten in the simple form

$$\hat{H} - \mu\hat{N} = -\sum_{j=0}^{n-1} X_j \hat{K}_j \qquad (16.32)$$

The operator $\hat{K}_j$ basically acts on the link between sites j and $j-1$. The $\hat{K}_j$ have the following convenient properties:

[1] They are Hermitian.

[2] They are mutually commuting ($[\hat{K}_j,\ \hat{K}_{j'}] = 0$).

[3] $\hat{K}_j^2 = 1$, i.e. their eigenvalues are ± 1.

[4] They commute with the fermion number parity.

[5] The quantity $\prod_{j=0}^{n-1} \hat{K}_j = -\prod_{j=0}^{n-1}(1-2\hat{n}_j)$ is minus the total fermion number parity.

As a result of property (4), the parity is a good quantum number (though the fermion number of the "system" itself is of course not). If all the X_j are positive definite, the odd-parity groundstate actually lies lower than the even-parity one. To remedy this state of affairs, it is convenient to choose one of the X_j, say X_0, to be negative and to have absolute magnitude less than that of any of the rest. The explicit form of the even-parity groundstate is then easily seen to be

$$|0\rangle_+ = 2^{-n/2}(1-\hat{K}_0)\prod_{j=1}^{n-1}(1+\hat{K}_j)|\mathrm{vac}\rangle \tag{16.33}$$

which evidently satisfies

$$\hat{K}_j|0\rangle = (1-2\delta_{j,0})|0\rangle \tag{16.34}$$

The odd-parity groundstate, created by breaking the link 0, has excitation energy $2|X_0|$, and the "single-fermion" excited (odd-parity) states are each associated with breaking the links $j(\neq 0)$ and have positive definite energy $2X_j$ (it is irrelevant to our argument whether or not there is any degeneracy). Explicitly, if we define the operators

$$\hat{\Omega}_j^\dagger \equiv 2^{-1/2}\big(b_{j-1}^\dagger + (-1)^{\delta_{j0}}c_j^\dagger\big) \tag{16.35}$$

with $b_j^\dagger, c_j^\dagger$ defined by (16.31), then the single-fermion excitations are created by $\hat{\Omega}_j^\dagger$, while the operators $\hat{\Omega}_j$ are pure annihilators.

Imagine now that while keeping the sign of X_0 negative we gradually turn its magnitude down to zero. Nothing changes, so in the limit $|X_0| \to 0$ the even- and odd-parity groundstates become degenerate (i.e. we get a zero-energy DB fermion); however, they are still simply related by changing the state of the $(0, n-1)$ link with the operator $\Omega_0^\dagger \equiv 2^{-1/2}(b_{n-1}^\dagger - c_0^\dagger)$. Now comes the crucial step: Once there is no energy associated with the $(0, n-1)$ link, we may as well break it entirely and reorganize the ring of sites into a linear lattice! As a result, we still have a zero-energy DB fermion, but it is "split," i.e. generated by a linear combination of the operators $b_{n-1}^\dagger$ and $c_0^\dagger$, which now refer to sites at opposite ends of the lattice. The "pure annihilator" $\hat{\Omega}_0$ is likewise "split"; however, one can take linear combinations of the zero-energy DB fermion and the pure annihilator so as to form the operators $b_{n-1}^\dagger$ and $c_0^\dagger$, which are localized at different ends of the linear chain. These are nothing but the two Majorana fermions of the model.

I believe that this very simple toy model, with its obvious generalizations, obtained e.g. by allowing the phase of Δ_j to vary along the ring, may hold some potential for a more intuitive understanding of the behavior of a real-life $(p+ip)$ Fermi superfluid, and I intend to pursue this direction further; for the present, I would like to hope that the above discussion may remove at least a part of the "mystery" that some readers of the existing literature may feel attaches to the concept of a Majorana fermion.

Acknowledgements

It is a pleasure to dedicate this paper to Tom Erber and to wish him many more happy years of activity in physics. This material is based on work supported by the U.S. Department of Energy, Division of Materials Sciences under Award No. DE-FG02-07ER46453 through the Frederick Seitz Materials Research Laboratory at the University of Illinois at Urbana-Champaign. I am grateful to Mike Stone for a helpful discussion.

References

[1] E. Majorana, Nuovo Cimento, **14**, 171 (1937) (in Italian).

[2] See e.g. K. Halterman,O. T. Valls and P. Barsic, Phys. Rev. B, **77**, 174511 (2008).

[3] N. B. Kopnin and M. M. Salomaa , Phys. Rev. B, **44**, 9667 (1991).

[4] G. E. Volovik, JETP Letters, **70**, 609 (1999).

[5] N. Read and D. Green, Phys. Rev. B, **61**, 10267 (2000).

[6] D. Ivanov, Phys. Rev. Lett., **86** 268 (2001).

[7] A. Stern, F. von Oppen and E. Mariani, Phys. Rev., **70**, 205338 (2004).

[8] P. W. Anderson, Rev. Mod. Phys., **38**, 298 (1966).

[9] A. J. Leggett, Quantum Liquids: Bose Condensation and Cooper Pairing in Condensed Matter Systems, Oxford University Press, Oxford, 2006, ch. 2.

[10] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962).

[11] M. Stone and S-B. Chung, Phys. Rev. B, **73**, 014505 (2006).

17

"Running" Gravitational Constant?

A.C. Melissinos
Department of Physics and Astronomy
University of Rochester

Abstract

If the gravitational interaction is unified with the electro-weak and strong interactions at a mass $M = 10^{15}$ GeV, the evolution of Newton's constant must differ from its classical (general relativistic) form. We can model such behavior by introducing an ad hoc dependence on $\ell n(s/4m^2)$, where s is the usual cm energy between two protons. We can then predict the observable effects for relativistic collisions ($\sqrt{s} \sim 1.4 \times 10^4$ GeV) as well as for the case of low velocity motion ($\beta^2 \sim 10^{-5}$).

It is well known that the dimensionless coupling constants of the three gauge groups of the standard model $SU(3) \times U(2) \times U(1)$ vary with the momentum transfer of the interaction [1]. This effect which is due to the polarization of the vacuum was first recognized for the electromagnetic field. It is most prominent in the case of the color field and leads to asymptotic freedom.

Extrapolation to higher energies is governed by the equations of the renormalization group, and it is customary to consider the inverse coupling constants[1]

[1]Throughout this note we express particle masses by their equivalent energy, ie $m_z = M_z c^2$ etc

$$\begin{aligned}
\frac{1}{\alpha_1(\sqrt{s})} &= \frac{1}{\alpha_e(m_Z)}\frac{3}{5}\cos^2\theta_W - \frac{1}{12\pi}(4n_f)\ell n\left(\frac{s}{m_Z^2}\right) \\
\frac{1}{\alpha_2(\sqrt{s})} &= \frac{1}{\alpha_e(m_Z)}\sin^2\theta_W + \frac{1}{12\pi}\left(22 - 4n_f - \frac{1}{2}\right)\ell n\left(\frac{s}{m_Z^2}\right) \\
\frac{1}{\alpha_3(\sqrt{s})} &= \frac{1}{\alpha_s(m_Z)} + \frac{1}{12\pi}\left(33 - 4n_f\right)\ell n\left(\frac{s}{m_Z^2}\right)
\end{aligned}$$

The above expressions have been normalized at a cm energy $\sqrt{s} = m_Z$ where the couplings are given by

$$\alpha_e = \frac{e^2}{(4\pi\epsilon_0)\hbar c} = \frac{1}{128}$$

$$\alpha_3 = \alpha_s = \frac{g_s^2}{\hbar c} = 0.118$$

$$\sin^2\theta_W(m_Z) = 0.2315$$

and n_f is the number of quark/lepton families.

The inverse couplings are plotted in Fig.1 as a function of $\sqrt{s}$. As observed by Georgi and Glashow [2] all three couplings seem to reach the same value at an energy $\sqrt{s} \simeq 10^{14}$ GeV which is referred to as the "Grand Unification Scale". If the couplings evolve according to the minimal supersymmetric model (MSSM) much better agreement is obtained, and within present uncertainties, the constants meet exactly at $\sqrt{s} = 10^{15}$ GeV [3].

The gravitational constant depends on the interaction energy as well. Consider two protons moving against each other in the laboratory frame with velocity β and energy γm_p. The gravitational coupling in this case takes the form

$$\alpha_G(\sqrt{s}) = \frac{G_N m_p^2}{\hbar c}(2\gamma^2 - 1) \tag{17.1}$$

The factor of $2\gamma^2$ arises because both energy γm_p and momentum $\gamma\beta m_p$ couple; see for instance [4]. The c.m. collision energy squared is $s = 4\gamma^2 m_p^2$, so Eq.(1) can be written as

$$\alpha_G(\sqrt{s}) = \frac{G_N m_p^2}{\hbar c}\left[\frac{s}{2m_p^2} - 1\right] \tag{17.2}$$

valid for $s \geq 4m_p^2$. Numerically

$$\frac{G_N m_p^2}{\hbar c} = \left(\frac{m_p}{M_P}\right)^2 = 0.59 \times 10^{-38} \tag{17.3}$$

where M_P is the Planck Mass. When $\sqrt{s/2} = M_P$, then α_G becomes unity.

If all four forces can be derived from a single symmetry group then the gravitational coupling should reach the common value

$$1/\alpha_G = 1/\alpha_1 = 1/\alpha_2 = 1/\alpha_3 \simeq 42 \tag{17.4}$$

at the unification scale $\sqrt{s} = 10^{15}$ GeV. [5]. We can achieve this by modifying the classical evolution of $\alpha_G(s)$ by a logarithmic term, of the form

$$\alpha_G(\sqrt{s}) = \left(\frac{m_p}{M_P}\right)^2 \left[\left(\frac{s}{2m_p^2}\right)^{1+b\ell n(s/4m^2)} - 1\right] \tag{17.5}$$

The coefficient b is found, by the requirement of Eq.(4), to have the value

$$b = 0.00340 \tag{17.6}$$

and the result is shown in Fig.2. We can now obtain the gravitational coupling at any given interaction energy. There are two possibilities for testing this hypothesis:

In the first case one tries to measure the gravitational effect at LHC energies. Here there is a significant difference in the couplings

$$\alpha_{GC}(\sqrt{s} = 1.4 \times 10^4 \text{ GeV}) = 6.6 \times 10^{-31} \qquad \text{classical}$$

$$\alpha_{GR}(\sqrt{s} = 1.4 \times 10^4 \text{ GeV}) = 2.0 \times 10^{-30} \qquad \text{running}$$

However the gravitational effects are extremely small and are dominated by the much larger electromagnetic force [6].

The other possibility is to test the deviation from the classical behavior at low energies but with macroscopic bodies. The classical correction in this case leads to

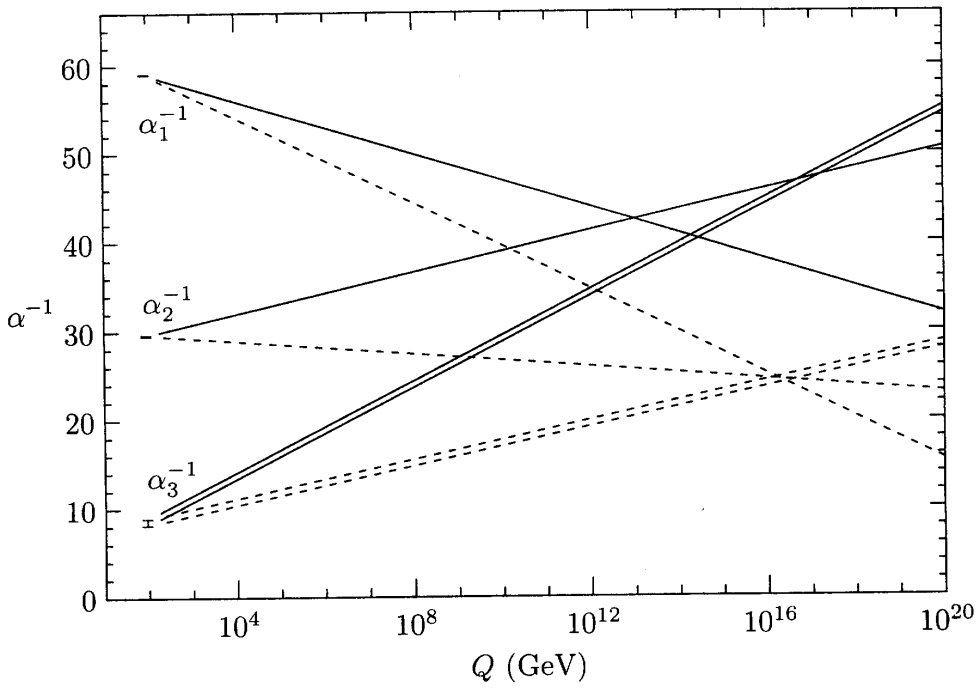


Figure 17.1: Extrapolation of the coupling constants of the standard model. Solid lines are calculated using the β-functions corresponding to the known elementary particles. Dashed lines are according to the MSSM. From Ref. [1].

$$\alpha_{GC}(\beta) = \left(\frac{m_p}{M_P}\right)^2 (1 + 2\beta^2) \tag{17.7}$$

while for the model of Eq.(5)

$$\begin{aligned} \alpha_{GR}(\beta) &\simeq \left(\frac{m_p}{M_P}\right)^2 \left\{ \left[2(1+\beta^2) \right]^{1+b\beta^2} - 1 \right\} \\ &\simeq \alpha_{GC}(\beta)[1+x] \end{aligned} \tag{17.8}$$

with

$$x = 2^{1+b\beta^2} - 2 \sim 2(\ell n 2) b\beta^2 \tag{17.9}$$

For a satellite in earth orbit

$$\beta^2 = 0.7 \times 10^{-9}$$

whereas for a close solar orbit

$$\beta^2 = 0.2 \times 10^{-5}$$

In this latter case, the effect of the running coupling constant is

$$x \simeq 10^{-8}$$

which could be measurable as a difference in the predicted orbital dynamics, beyond the effects of classical general relativity.

In a recent publication [8] an apparent increase in the gravitational attraction during the Earth fly-by of the NEAR mission was reported. For $\beta = 4 \times 10^{-5}$ it is found that $\Delta G/G \sim 2 \times 10^{-6}$, which however, is significantly larger than the prediction of Eq.(8).

References

[1] M.E. Peskin and D.V. Schroeder "An Introduction to Quantum Field Theory", Addison-Wesley (1995).

[2] H. Georgi and S.L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

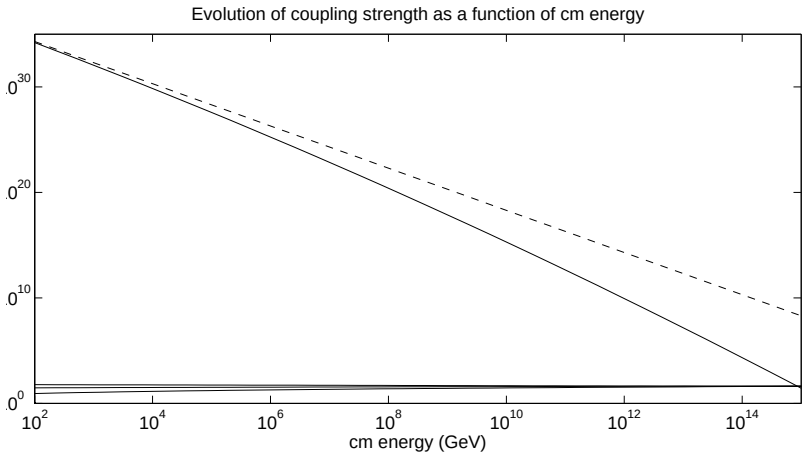


Figure 17.2: Extrapolation in cm energy of the inverse gravitational coupling, $1/a_{GR}$, when Eq.(5) of the text is used so that all four constants have the same value at $\sqrt{s} = 10^{15}$ GeV; solid curve. The dashed curve represents the classical (general relativity) evolution of the gravitational coupling, Eq.(2) of the text. The evolution of the electroweak and strong couplings is also shown.

[3] F. Wilczek in "Critical Problems in Physics", Princeton University Press (1997).

[4] A.C. Melissinos, Il Nuovo Cimento, **62B**, 190 (1981).

[5] I heard of this suggestion in a talk by F. Wilczek at the symposium in honor of N.P. Samios held at BNL, May 2002.

[6] P. Reiner et al., Physics Letters **B176**, 233 (1986).

[7] Ph. Bernard et al., Review of Scientific Instruments **72**, 2428 (2001).

[8] J.D. Anderson et al., Phys. Rev. Lett. **100**, 091102 (2008).

18

Quantum Fields in a Dielectric: Langevin and Exact Diagonalization Approaches

F. S. S. Rosa, D. A. R. Dalvit, and P. W. Milonni
Theoretical Division
Los Alamos National Laboratory

18.1 Introduction

Professor Erber has made important contributions to several areas of both pure and applied physics, making it easy to identify topics about which one can write to celebrate his work; approaching such topics at his level of rigor and insight is far more difficult! His contributions to fundamental electromagnetic theory and quantum mechanics in particular include papers on electromagnetic energy density in dispersive media [1], synchrotron-Cerenkov radiation [2], radiation reaction [3], and quantum jumps [4], to cite only a few of those with which we are familiar. The first two papers cited, among others, deal with electromagnetic processes in dielectric media, and a small part of that subject will be addressed here. Specifically, this article is concerned with the quantized electromagnetic field in a dispersive and dissipative dielectric medium, and with the energy density in such a medium. Erber's work has also dealt with nonlinear processes in such media as well as in vacuum, but we will restrict ourselves here to linear and idealized, homogeneous media.

This is an important subject about which much has been written, although in most of the literature it has been assumed that the medium can be assumed to be non-dissipative at field frequencies of interest. The theory ignoring dissipation is not without value; it can be used to describe, for example, the spontaneous emission by an atom in a host dielectric that does not absorb radiation at the atom's transition frequency. But a dispersive medium cannot be non-absorbing at all frequencies. What happens, for instance, if the medium in our example strongly absorbs radiation at the transition frequency? And what about situations in which we cannot exclude any frequencies *a priori* and therefore cannot ignore absorption? In the calculation of the van der Waals force between two neutral dielectric bodies, for example, all field frequencies can in principle contribute to the force. For two perfectly conducting parallel plates at zero temperature, similarly, Casimir [5] discovered, as a consequence of the zero-point electromagnetic energy of every mode, that there is an attractive force per unit area between the plates. His original method involving changes in zero-point field energy was later extended to dielectrics by van Kampen *et al.* [6] and others [7]. As emphasized by Ginzburg [8], however, these theories invoking changes in zero-point energy ignore absorption: " ... oddly enough there is no mention that they consider directly only transparent media" [9]. An entirely different route, based on the calculation of the force via the stress tensor, was taken by Lifshitz [10]; he accounts for absorption through the fluctuation-dissipation relation between the quantum fluctuations of the polarization density and the imaginary (absorptive) part of the permittivity.

The intent here is to derive, in probably the simplest way possible, expressions for the quantum electromagnetic field in a dispersive and dissipative dielectric medium, treating the medium as a continuum. The derivation might seem superfluous in the sense that correct expressions for the electric and magnetic fields in such a medium are already available [11]. However, the diagonalization procedure by which these expressions are obtained is not easily applied to general, inhomogeneous media, whereas the "Langevin approach" presented here can be applied more or less straightforwardly when extended and formulated via Green functions [12].

We begin in the following section with the simple model of an oscillator A coupled to a reservoir R of other oscillators, the R oscillators causing a damping of A described by a Langevin equation. In Section 18.3 we review the Fano procedure for the diagonalization of the Hamiltonian of this system, and compare the diagonalization and Langevin-equation approaches insofar as they describe the time evolution of A. In the limit of zero temperature, or at any finite temperature, the two descriptions are shown to be equivalent. In Section 18.4 we

generalize these considerations, following the Langevin approach, to the model of a homogeneous dielectric medium in which each atom is treated as a harmonic oscillator coupled to its own reservoir. The Langevin forces acting on the atoms give rise to a noise polarization determined by the reservoir operators, and the fluctuating electromagnetic field caused by this noise polarization can therefore be expressed in terms these operators. In similar fashion to the model of Sections 18.2 and 18.3, the quantized electric and magnetic fields obtained in this way have exactly the same form as obtained by Fano diagonalization. It is shown explicitly in Section 18.5 that the zero-point energy per mode of frequency ω is $(1/2)\hbar\omega$ regardless of whether there is absorption at that frequency [13].

18.2 An Oscillator and a Reservoir

Consider an oscillator A of frequency ω_0 coupled to a reservoir R of other oscillators, a well-studied model for dissipation in quantum theory. To make things as simple as possible we will assume that the A-R coupling involves only energy-conserving processes, and choose this coupling such that the Hamiltonian is

$$\hat{H} = \hbar\omega_0 \hat{a}^\dagger \hat{A} + \int_0^\infty d\omega \hbar\omega \hat{b}^\dagger(\omega)\hat{b}(\omega) + \hbar\sqrt{\gamma/\pi}\int_0^\infty d\omega[\hat{a}^\dagger \hat{b}(\omega) + \hat{b}^\dagger(\omega)a], \quad (18.1)$$

with $[\hat{A}, \hat{a}^\dagger] = 1$, $[\hat{b}(\omega), \hat{b}(\omega')] = 0$, $[\hat{b}(\omega), \hat{b}^\dagger(\omega')] = \delta(\omega - \omega')$. (We use the circumflex to denote operators.) Because only energy-conserving processes are included, our model is consistent with the so-called "rotating-wave approximation" (RWA). The coupling we have chosen results in a frictional damping rate γ in the time evolution of A, as follows.

The Heisenberg equations of motion for $\hat{A}$ and $\hat{b}(\omega)$ are

$$\dot{\hat{A}} = -i\omega_0 a - i\sqrt{\gamma/\pi}\int_0^\infty d\omega \hat{b}(\omega), \quad (18.2)$$

$$\dot{\hat{b}}(\omega) = -i\omega\hat{b}(\omega) - i\sqrt{\gamma/\pi}\hat{A}. \quad (18.3)$$

Using the formal solution

$$\hat{b}(\omega, t) = \hat{b}_0(\omega)e^{-i\omega t} - i\sqrt{\gamma/\pi}\int_0^t dt' \hat{A}(t')e^{i\omega(t'-t)} \quad (18.4)$$

of equation (18.3) in equation (18.2), and defining $\hat{b}_0(\omega) \equiv \hat{b}(\omega, 0)$, we obtain

$$\begin{aligned} \dot{\hat{A}}(t) + i\omega_0 \hat{A}(t) &+ (\gamma/\pi) \int_0^\infty d\omega \int_0^t dt' \hat{A}(t') e^{i\omega(t'-t)} \\ &= -i\sqrt{\gamma/\pi} \int_0^\infty d\omega \hat{b}_0(\omega) e^{-i\omega t}, \end{aligned} \tag{18.5}$$

the operator on the right-hand side being a quantum Langevin force. We solve this equation for "steady state" ($\gamma t \gg 1$) by first writing

$$\hat{A}(t) = \int_0^\infty d\Omega A(\Omega) \hat{b}_0(\Omega) e^{-i\Omega t}. \tag{18.6}$$

Then, using the approximation

$$\int_0^t dt' e^{i(\omega-\Omega)(t'-t)} \cong \pi\delta(\omega - \Omega) - iP\frac{1}{\omega} - \Omega \tag{18.7}$$

for times t such that $\Omega t \gg 1$ for frequencies Ω that make a significant contribution to the time evolution of $\hat{A}(t)$, we obtain

$$A(\Omega) = \frac{\sqrt{\gamma/\pi}}{\Omega} - \omega_0 + \Delta(\Omega) + i\gamma. \tag{18.8}$$

The frequency shift

$$\Delta(\Omega) = \frac{\gamma}{\pi} P \int_0^\infty \frac{d\omega}{\omega} - \Omega \tag{18.9}$$

obviously diverges in our model, and so the upper limit of integration must be appropriately cut off; for our purposes there is no need to explicitly indicate any cutoff. Then our solution for $\hat{A}(t)$ is

$$\hat{A}(t) = \sqrt{\frac{\gamma}{\pi}} \int_0^\infty d\Omega \hat{b}_0(\Omega) e^{-i\Omega t}/\Omega - \omega_0 + \Delta(\Omega) + i\gamma. \tag{18.10}$$

Note that, under the assumption that $\omega_0 \gg \gamma$, consistent with the RWA, we have

$$[\hat{A}(t), \hat{a}^\dagger(t)] = \frac{\gamma}{\pi} \int_0^\infty \frac{d\omega}{(\omega - \omega_0)^2 + \gamma^2} \cong 1, \tag{18.11}$$

as required for the validity of the RWA.

18.3 Fano Diagonalization

A method of diagonalizing a Hamiltonian for coupled oscillators, used many years ago by Fano [14], has been employed in seminal work by Huttner and Barnett [11] to obtain expressions for the quantized fields in a *dissipative* dielectric medium. Here we briefly review the method in the case of the model Hamiltonian (18.1), and compare it to the approach of the preceding section.

We define an operator

$$\hat{B}(\Omega) = \alpha(\Omega)\hat{A} + \int_0^\infty d\omega \beta(\Omega, \omega)\hat{b}(\omega) \tag{18.12}$$

that we require to satisfy

$$[\hat{B}(\Omega), \hat{B}^\dagger(\Omega')] = \delta(\Omega - \Omega'), \quad [\hat{B}(\Omega), \hat{B}(\Omega')] = 0, \tag{18.13}$$

and also require that the Hamiltonian (18.1) takes the diagonal form

$$\hat{H} = \int_0^\infty d\Omega \hbar\Omega \hat{B}^\dagger(\Omega)\hat{B}(\Omega). \tag{18.14}$$

From $[\hat{B}(\Omega), \hat{H}] = \hbar\Omega\hat{B}(\Omega)$ and the definition (18.12) we deduce equations relating the coefficients $\alpha(\Omega)$ and $\beta(\Omega, \omega)$:

$$(\omega_0 - \Omega)\alpha(\Omega) = -\sqrt{\frac{\gamma}{\pi}} \int_0^\infty d\omega \beta(\Omega, \omega), \tag{18.15}$$

$$\int_0^\infty d\omega \left[(\omega - \Omega)\beta(\Omega, \omega) + \sqrt{\frac{\gamma}{\pi}}\alpha(\Omega) \right] \hat{b}(\omega) = 0. \tag{18.16}$$

The last equation leads us to write

$$\beta(\Omega, \omega) = \alpha(\Omega) f(\Omega)\delta(\omega - \Omega) - \sqrt{\frac{\gamma}{\pi}} \frac{\alpha(\Omega)}{\omega} - \Omega, \tag{18.17}$$

and it follows from (18.15) that

$$f(\Omega) = \sqrt{\frac{\pi}{\gamma}}[\Omega - \omega_0 + \Delta(\Omega)], \tag{18.18}$$

i.e.,

$$\beta(\Omega, \omega) = \sqrt{\frac{\pi}{\gamma}}\alpha(\Omega)[\Omega - \omega_0 + \Delta(\Omega)]\delta(\omega - \Omega) - \sqrt{\frac{\gamma}{\pi}} \frac{\alpha(\Omega)}{\omega} - \Omega, \tag{18.19}$$

where $\Delta(\Omega)$ is defined by (18.9).

To determine $\alpha(\Omega)$ we impose the requirement that the commutation relations (18.13) be satisfied. From the commutation relations stated earlier for the $\hat{A}$ and $\hat{b}$ operators we obtain

$$\begin{aligned}[\hat{B}(\Omega),\hat{B}^\dagger(\Omega')] &= \alpha(\Omega)\alpha^*(\Omega') \\ &+ \int_0^\infty d\omega \int_0^\infty d\omega' \beta(\Omega,\omega)\beta^*(\Omega',\omega')\delta(\omega-\omega') \quad (18.20)\end{aligned}$$

or, from (18.19) and some straightforward algebra,

$$\begin{aligned}[\hat{B}(\Omega),\hat{B}^\dagger(\Omega')] &= \frac{\pi}{\gamma}\alpha(\Omega)\alpha^*(\Omega') \times \left\{ [\Omega-\omega_0+\Delta(\Omega)]^2\delta(\Omega-\Omega') \right. \\ &\left. + \frac{\gamma}{\pi}\frac{\Delta(\Omega')-\Delta(\Omega)}{\Omega'-\Omega} + \frac{\gamma^2}{\pi^2} P\int_0^\infty \frac{d\omega}{\omega-\Omega} P\int_0^\infty \frac{d\omega'}{\omega'-\Omega'}\delta(\omega-\omega') \right\}. \quad (18.21)\end{aligned}$$

Using

$$P\frac{1}{\omega-\Omega} = \frac{1}{\omega-\Omega+i\epsilon} + i\pi\delta(\omega-\Omega) \quad (\epsilon\to 0^+) \qquad (18.22)$$

and partial fractions we obtain

$$[\hat{B}(\Omega),\hat{B}^\dagger(\Omega')] = \frac{\pi}{\gamma}\alpha(\Omega)\alpha^*(\Omega')\left(\gamma^2+[\Omega-\omega_0+\Delta(\Omega)]^2\right)\delta(\Omega-\Omega'). \qquad (18.23)$$

Therefore we can satisfy (18.13) by taking

$$\alpha(\Omega) = \frac{\sqrt{\gamma/\pi}}{\Omega-\omega_0+\Delta(\Omega)-i\gamma}. \qquad (18.24)$$

Then

$$\hat{A}(t) = \int_0^\infty d\Omega\alpha^*(\Omega)\hat{B}(\Omega,t) = \sqrt{\frac{\gamma}{\pi}}\int_0^\infty d\Omega \frac{\hat{B}(\Omega,0)e^{-i\Omega t}}{\Omega-\omega_0+\Delta(\Omega)+i\gamma}, \qquad (18.25)$$

since $\hat{B}(\Omega,t) = \hat{B}(\Omega,0)e^{-i\Omega t}$.

The expressions (18.10) and (18.25) obtained respectively in the Langevin and Fano approaches look formally the same in the sense that $[\hat{B}(\Omega,0),\hat{B}^\dagger(\Omega',0)] = [\hat{b}_0(\Omega),\hat{b}_0^\dagger(\Omega')] = \delta(\Omega-\Omega')$ and $[\hat{B}(\Omega,0),\hat{B}(\Omega',0)] = [\hat{b}_0(\Omega),\hat{b}_0(\Omega')] = 0$. They differ in that $\hat{B}(\Omega,0)$ in (18.25) is a linear combination of A and R operators, whereas only R operators determine $\hat{A}(t)$ in (18.10). Suppose, however, that at

$t = 0$ the A and R oscillators are all in their ground states. This state $|\Psi\rangle$ is the *exact* ground state of the coupled A-R system in the RWA:

$$\hat{B}^\dagger(\Omega, t)\hat{B}(\Omega, t)|\Psi\rangle = 0. \tag{18.26}$$

In this case the properties of A derived from (18.10) are trivially equivalent to those obtained from (18.25). If the system is not initially in an eigenstate of $\hat{B}^\dagger\hat{B}$, it will nevertheless approach after a time $\gg \gamma^{-1}$ an equilibrium state for which the long-term solution (18.10) for $\hat{A}(t)$ is applicable, i.e., transient effects associated with $\hat{A}(0)$ at some initial time $t = 0$ ultimately play no role in the evolution of A. Equilibrium values of correlation functions involving products of the $\hat{B}$ operators are determined solely by the reservoir operators; in thermal equilibrium, for example, $\langle\hat{B}^\dagger(\Omega, t)\hat{B}(\Omega', t)\rangle = \langle\hat{b}^\dagger(\Omega, t)\hat{b}(\Omega', t)\rangle = [\exp(\hbar\Omega/k_BT) - 1]^{-1}\delta(\Omega - \Omega')$ and it follows from either (18.10) or (18.25) that $\langle\hat{a}^\dagger(t)\hat{A}(t)\rangle = [\exp(\hbar\omega_0/k_BT) - 1]^{-1}$ when we invoke the condition $\omega_0 \gg \gamma$ as in (18.11). In other words, in any state of equilibrium the solutions (18.10) and (18.25) provide equivalent descriptions of A. This equivalence holds more generally beyond the RWA (see below).

18.4 Fields in a Dielectric Continuum

Aside from the need to introduce oscillator strengths in order to obtain correct numerical results, we can model a dielectric medium in which atoms remain with high probability in their ground states as a collection of harmonic oscillators. We take each oscillator "atom" to have a mass m and a natural frequency ω_0. We assume each of these material oscillators is coupled to a reservoir of other harmonic oscillators responsible for the damping of its oscillations and line broadening. For the Hamiltonian, including the electromagnetic field and its (electric-dipole) coupling to the material oscillators, we write

$$\begin{aligned}\hat{H} &= \frac{1}{8\pi}\int d\mathbf{r}(\hat{\mathbf{E}}^2 + \hat{\mathbf{H}}^2) + \sum_j \left(\frac{1}{2m}[\hat{\mathbf{p}}_j - \frac{e}{c}\hat{\mathbf{A}}(\mathbf{r}_j)]^2\right. \\ &\quad \left. + \frac{1}{2}m\omega_0^2\hat{\mathbf{x}}_j^2\right) + \int_0^\infty d\omega\hbar\omega\sum_j\left[\hat{\mathbf{b}}_j^\dagger(\omega)\cdot\hat{\mathbf{b}}_j(\omega) + \frac{1}{2}\right] \\ &\quad - i\int_0^\infty d\omega\Lambda(\omega)\sum_j\hat{\mathbf{x}}_j\cdot[\hat{\mathbf{b}}_j(\omega) - \hat{\mathbf{b}}_j^\dagger(\omega)].\end{aligned} \tag{18.27}$$

The first two terms are the Hamiltonian operators for the electromagnetic field, the material oscillators (atoms), and their coupling via the (Coulomb-gauge)

vector potential $\hat{\mathbf{A}}(\mathbf{r}_j)$, $\mathbf{r}_j$ being the position of the jth atom. The third and fourth terms are respectively the Hamiltonian operators for the reservoir oscillators and their interaction with the atoms. The reservoir oscillators satisfy the commutation relations

$$[\hat{b}_{i\mu}(\omega), \hat{b}^\dagger_{j\nu}(\omega')] = \delta_{ij}\delta_{\mu\nu}\delta(\omega-\omega'), \quad [\hat{b}_{i\mu}(\omega), \hat{b}_{j\nu}(\omega')] = 0, \tag{18.28}$$

where we use Greek letters to denote Cartesian components of vectors. We choose the atom-reservoir coupling constant to be

$$\Lambda(\omega) = \left(\frac{m\hbar\gamma\omega}{\pi}\right)^{1/2}. \tag{18.29}$$

Then, as shown below, each atom's oscillations are damped at the rate γ. Note that no rotating-wave approximation is made in writing (18.27). The operators $\hat{\mathbf{A}}$ and $\hat{\mathbf{E}}$ satisfy the usual canonical commutation relations for the electromagnetic field.

From (18.28) and $[\hat{x}_{i\mu}, \hat{p}_{j\nu}] = i\hbar\delta_{ij}\delta_{\mu\nu}$ we obtain the Heisenberg equations of motion

$$\ddot{\hat{\mathbf{x}}}_j + \omega_0^2\hat{\mathbf{x}}_j = \frac{e}{m}\hat{\mathbf{E}}(\mathbf{r}_j) + \frac{i}{m}\int_0^\infty d\omega\Lambda(\omega)[\hat{\mathbf{b}}_j(\omega,t) - \hat{\mathbf{b}}_j^\dagger(\omega,t)], \tag{18.30}$$

$$\dot{\hat{\mathbf{b}}}_j(\omega,t) = -i\omega\hat{\mathbf{b}}_j(\omega,t) + \frac{1}{\hbar}\Lambda(\omega)\hat{\mathbf{x}}_j\ . \tag{18.31}$$

Using the formal solution of (18.31) in (18.30), we write

$$\begin{aligned}\ddot{\hat{\mathbf{x}}}_j + \omega_0^2\hat{\mathbf{x}}_j &= \frac{e}{m}\hat{\mathbf{E}}(\mathbf{r}_j) + \frac{1}{m}\hat{\mathbf{F}}_{Lj}(t) \\ &+ \frac{i}{m\hbar}\int_0^\infty d\omega\Lambda^2(\omega)\int_0^t dt'\hat{\mathbf{x}}_j(t')[2i\sin\omega(t'-t)],\end{aligned} \tag{18.32}$$

where the Langevin force operator $\hat{\mathbf{F}}_{Lj}(t)$ acting on the jth atom is

$$\hat{\mathbf{F}}_{Lj}(t) = i\int_0^\infty d\omega\Lambda(\omega)[\hat{\mathbf{b}}_j(\omega,0)e^{-i\omega t} - \hat{\mathbf{b}}_j^\dagger(\omega,0)e^{i\omega t}]. \tag{18.33}$$

The third term on the right-hand side of (18.32) is

$$\begin{aligned}-\frac{2}{m\hbar}\int_0^\infty d\omega\Lambda^2(\omega)\int_0^t dt'\hat{\mathbf{x}}_j(t')\sin\omega(t'-t) &= -\frac{2\gamma}{\pi}\int_0^t dt'\hat{\mathbf{x}}_j(t') \\ \int_0^\infty d\omega\omega\sin\omega(t'-t) = 2\gamma\int_0^t dt'\hat{\mathbf{x}}_j(t')\frac{\partial}{\partial t'}\delta(t'-t) &= -\gamma\dot{\hat{\mathbf{x}}}_j(t)\ .\end{aligned} \tag{18.34}$$

We ignore a divergent frequency shift which, as in the model considered in the preceding sections, can be made finite by introducing a form factor or a high-frequency cutoff; the (finite) shift can be assumed to be contained in the definition of ω_0. Equation (18.32) then has the form of a quantum Langevin equation:

$$\ddot{\hat{\mathbf{x}}}_j + \gamma\dot{\hat{\mathbf{x}}}_j + \omega_0^2\hat{\mathbf{x}}_j = \frac{e}{m}\hat{\mathbf{E}}(\mathbf{r}_j) + \frac{1}{m}\hat{\mathbf{F}}_{Lj}(t). \tag{18.35}$$

In the absence of coupling to the electromagnetic field we have, for times $t \gg \gamma^{-1}$,

$$\hat{\mathbf{p}}_j(t) = m\dot{\hat{\mathbf{x}}}_j(t) = \int_0^\infty d\omega\omega\Lambda^2(\omega)\left[\frac{\hat{\mathbf{b}}_j(\omega)e^{-i\omega t}}{\omega_0^2-\omega^2-i\gamma\omega} + \frac{\hat{\mathbf{b}}_j^\dagger(\omega)e^{i\omega t}}{\omega_0^2-\omega^2+i\gamma\omega}\right]. \tag{18.36}$$

(We now write $\hat{\mathbf{b}}_j(\omega)$ in place of $\hat{\mathbf{b}}_j(\omega,0)$.) Similarly, using (18.28), we obtain

$$\begin{aligned}[\hat{x}_{i\mu}(t),\hat{p}_{j\nu}(t')] &= \delta_{ij}\delta_{\mu\nu}\frac{2i\hbar\gamma}{\pi}\int_0^\infty \frac{d\omega\omega^2\cos\omega(t'-t)}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2} \\ &= i\hbar\delta_{ij}\delta_{\mu\nu}\left[\cos\omega_1(t'-t) - \frac{\gamma}{2\omega_1}\sin\omega_1|t'-t|\right]e^{-\gamma|t'-t|/2},\end{aligned} \tag{18.37}$$

where $\omega_1 \equiv [\omega_0^2-\gamma^2/4]^{1/2}$. The canonical commutation relation $[\hat{x}_{i\mu}(t),\hat{p}_{j\nu}(t)] = i\hbar\delta_{ij}\delta_{\mu\nu}$ is therefore preserved in the coupling of each atom to its reservoir.

Since we are working in the Heisenberg picture, expectation values are over the initial state of the coupled system of oscillators. If we assume that the reservoir is in an initial state of thermal equilibrium at temperature T, then

$$\begin{aligned}\langle\hat{b}_{i\mu}^\dagger(\omega)\hat{b}_{j\nu}(\omega')\rangle &= \langle\hat{b}_{i\mu}(\omega)\hat{b}_{j\nu}^\dagger(\omega')\rangle - \delta_{ij}\delta_{\mu\nu}\delta(\omega-\omega') \\ &= \frac{1}{e^{\hbar\omega/k_BT}-1}\delta_{ij}\delta_{\mu\nu}\delta(\omega-\omega').\end{aligned} \tag{18.38}$$

The Heisenberg equations of motion for the electric and magnetic fields that follow from the Hamiltonian (18.27) and the canonical commutation relations for the field operators have exactly the same form as their classical (Maxwell) counterparts:

$$\begin{aligned}\nabla\times\hat{\mathbf{E}} &= -\frac{1}{c}\frac{\partial\hat{\mathbf{B}}}{\partial t}, \\ \nabla\times\hat{\mathbf{H}} &= \frac{4\pi}{c}\hat{\mathbf{J}} + \frac{1}{c}\frac{\partial\hat{\mathbf{E}}}{\partial t}.\end{aligned} \tag{18.39}$$

For a charge-free medium, furthermore, $\nabla \cdot \hat{\mathbf{B}} = \nabla \cdot \hat{\mathbf{D}} = 0$, where

$$\begin{aligned}
&\hat{\mathbf{D}} = \hat{\mathbf{E}} + 4\pi\hat{\mathbf{P}}, \\
&\hat{\mathbf{J}}(\mathbf{r},t) = \frac{\partial\hat{\mathbf{P}}(\mathbf{r},t)}{\partial t}, \\
&\hat{\mathbf{P}}(\mathbf{r},t) = e\sum_j \hat{\mathbf{x}}_j(t)\delta^3(\mathbf{r}-\mathbf{r}_j),
\end{aligned} \tag{18.40}$$

with $\hat{\mathbf{B}} = \hat{\mathbf{H}}$ in our model.

It is convenient to work in the frequency domain and write

$$\begin{aligned}
\hat{\mathbf{E}}(\mathbf{r},t) &= \int_0^\infty d\omega[\hat{\mathbf{E}}(\mathbf{r},\omega)e^{-i\omega t} + \hat{\mathbf{E}}^\dagger(\mathbf{r},\omega)e^{i\omega t}], \\
\hat{\mathbf{H}}(\mathbf{r},t) &= \int_0^\infty d\omega[\hat{\mathbf{H}}(\mathbf{r},\omega)e^{-i\omega t} + \hat{\mathbf{H}}^\dagger(\mathbf{r},\omega)e^{i\omega t}], \\
\hat{\mathbf{P}}(\mathbf{r},t) &= \int_0^\infty d\omega[\hat{\mathbf{P}}(\mathbf{r},\omega)e^{-i\omega t} + \hat{\mathbf{P}}^\dagger(\mathbf{r},\omega)e^{i\omega t}].
\end{aligned} \tag{18.41}$$

The Fourier transform of the polarization density may be written as

$$\hat{\mathbf{P}}(\mathbf{r},\omega) = e\sum_j \hat{\mathbf{x}}_j(\omega)\delta^3(\mathbf{r}-\mathbf{r}_j), \tag{18.42}$$

$$\hat{\mathbf{x}}_j(t) = \int_0^\infty d\omega[\hat{\mathbf{x}}_j(\omega)e^{-i\omega t} + \hat{\mathbf{x}}_j^\dagger(\omega)e^{i\omega t}], \tag{18.43}$$

and it follows from (18.35) that

$$\begin{aligned}
\hat{\mathbf{P}}(\mathbf{r},\omega) &= \frac{e^2/m}{\omega_0^2-\omega^2-i\gamma\omega}\sum_j \hat{\mathbf{E}}(\mathbf{r}_j,\omega)\delta^3(\mathbf{r}-\mathbf{r}_j) \\
&\quad + \frac{ie/m}{\omega_0^2-\omega^2-i\gamma\omega}\Lambda(\omega)\sum_j \hat{\mathbf{b}}_j(\omega)\delta^3(\mathbf{r}-\mathbf{r}_j) \\
&\to \frac{Ne^2/m}{\omega_0^2-\omega^2-i\gamma\omega}\hat{\mathbf{E}}(\mathbf{r},\omega) + \frac{iNe/m}{\omega_0^2-\omega^2-i\gamma\omega}\Lambda_c(\omega)\hat{\mathbf{b}}(\mathbf{r},\omega)
\end{aligned} \tag{18.44}$$

in the approximation in which we assume the atoms are continuously distributed with a density N and we define $\Lambda_c(\omega) = \sqrt{\rho_m\hbar\gamma\omega/\pi}$, with $\rho_m = m/N$.

From Maxwell's equations and (18.44) we obtain

$$\nabla^2\hat{\mathbf{E}}(\mathbf{r},\omega) + \frac{\omega^2}{c^2}\epsilon(\omega)\hat{\mathbf{E}}(\mathbf{r},\omega) = -\frac{\omega^2}{c^2}\hat{\mathbf{K}}(\mathbf{r},\omega), \tag{18.45}$$

where the complex permittivity is defined as

$$\epsilon(\omega) = 1 - \frac{4\pi Ne^2/m}{\omega^2 - \omega_0^2 + i\gamma\omega} \equiv 1 - \frac{\omega_p^2}{\omega^2 - \omega_0^2 + i\gamma\omega} = \epsilon_R(\omega) + i\epsilon_I(\omega). \quad (18.46)$$

We have also defined the "noise polarization" at frequency ω:

$$\hat{\mathbf{K}}(\mathbf{r},\omega) = \frac{4\pi iNe/m}{\omega_0^2 - \omega^2 - i\gamma\omega}\Lambda(\omega)\hat{\mathbf{b}}(\mathbf{r},\omega). \quad (18.47)$$

This noise polarization obviously stems from the Langevin force $\hat{\mathbf{F}}_{Lj}(t)$ in the quantum Langevin equation (18.35). Its principal properties for our purposes are the thermal equilibrium expectation values

$$\begin{aligned} \langle \hat{K}_\mu(\mathbf{r},\omega)\rangle &= \langle \hat{K}_\mu^\dagger(\mathbf{r},\omega)\rangle = 0, \\ \langle \hat{K}_\mu(\mathbf{r},\omega)\hat{K}_\nu(\mathbf{r}',\omega')\rangle &= \langle \hat{K}_\mu^\dagger(\mathbf{r},\omega)\hat{K}_\nu^\dagger(\mathbf{r}',\omega')\rangle = 0, \end{aligned} \quad (18.48)$$

and

$$\langle \hat{K}_\mu^\dagger(\mathbf{r},\omega)\hat{K}_\nu(\mathbf{r}',\omega')\rangle = 4\hbar\epsilon_I(\omega)\delta_{\mu\nu}\delta(\omega-\omega')\delta^3(\mathbf{r}-\mathbf{r}')\frac{1}{e^{\hbar\omega/k_BT}-1}, \quad (18.49)$$

$$\langle \hat{K}_\mu(\mathbf{r},\omega)\hat{K}_\nu^\dagger(\mathbf{r}',\omega')\rangle = 4\hbar\epsilon_I(\omega)\delta_{\mu\nu}\delta(\omega-\omega')\delta^3(\mathbf{r}-\mathbf{r}')\left[\frac{1}{e^{\hbar\omega/k_BT}-1}+1\right], \quad (18.50)$$

all of which follow from (18.38) and $\langle \hat{b}_{i\mu}(\omega)\hat{b}_{j\nu}(\omega')\rangle = 0$. Equations (18.49) and (18.50) are the well-known fluctuation-dissipation relations between the correlation functions of a noise polarization and the imaginary part of the dielectric function [10, 15].

Next we define operators $\hat{g}_\lambda(\mathbf{k},\omega)$ by writing

$$\hat{\mathbf{K}}(\mathbf{r},\omega) = \int d^3k \sum_{\lambda=1,2} \hat{g}_\lambda(\mathbf{k},\omega)\mathbf{e}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (18.51)$$

Since $\nabla\cdot\hat{\mathbf{K}}(\mathbf{r},\omega) = 0$ we can choose the vectors $\mathbf{e}_{\mathbf{k}\lambda}$ such that $\mathbf{k}\cdot\mathbf{e}_{\mathbf{k}\lambda} = 0$, $\mathbf{e}_{\mathbf{k}\lambda}\cdot\mathbf{e}_{\mathbf{k}\lambda'} = 0$, $\lambda = 1,2$; we also take the $\mathbf{e}_{\mathbf{k}\lambda}$ to be real. Then

$$\begin{aligned} \hat{g}_\lambda(\mathbf{k},\omega) &= \left(\frac{1}{2\pi}\right)^3 \int d^3r\, \hat{\mathbf{K}}(\mathbf{r},\omega)\cdot\mathbf{e}_{\mathbf{k}\lambda}e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &\equiv \left(\frac{1}{2\pi}\right)^3 \int d^3r \hat{K}_\lambda(\mathbf{r},\omega)e^{-i\mathbf{k}\cdot\mathbf{r}}, \end{aligned} \quad (18.52)$$

and Eqs. (18.47) and (18.28) imply the commutation relation

$$[\hat{g}_\lambda(\mathbf{k},\omega), \hat{g}^\dagger_{\lambda'}(\mathbf{k}',\omega')] = \frac{\hbar}{2\pi^3}\epsilon_I(\omega)\delta_{\lambda\lambda'}\delta(\omega-\omega')\delta^3(\mathbf{k}-\mathbf{k}'). \tag{18.53}$$

We also define operators

$$\hat{c}_\lambda(\mathbf{k},\omega) \equiv [\hbar\epsilon_I(\omega)/2\pi^3]^{-1/2}\hat{g}_\lambda(\mathbf{k},\omega) \tag{18.54}$$

satisfying

$$[\hat{c}_\lambda(\mathbf{k},\omega), \hat{c}^\dagger_{\lambda'}(\mathbf{k}',\omega')] = \delta_{\lambda\lambda'}\delta(\omega-\omega')\delta^3(\mathbf{k}-\mathbf{k}'). \tag{18.55}$$

Finally an expression for the quantized electric field follows from (18.41), (18.45), (18.51), and (18.54):

$$\begin{aligned}\hat{\mathbf{E}}(\mathbf{r},t) &= \int d^3k \sum_\lambda \int_0^\infty d\omega \sqrt{\hbar\epsilon_I(\omega)/2\pi^3}\,\frac{\omega^2/c^2}{k^2-\epsilon(\omega)\omega^2/c^2}\hat{c}_\lambda(\mathbf{k},\omega)\mathbf{e}_{\mathbf{k}\lambda} \\ &\quad e^{-i(\omega t-\mathbf{k}\cdot\mathbf{r})} + \text{h.c.}\end{aligned} \tag{18.56}$$

From $\nabla\times\hat{\mathbf{E}} = -(1/c)\partial\hat{\mathbf{B}}/\partial t$ we also obtain

$$\begin{aligned}\hat{\mathbf{H}}(\mathbf{r},t) &= i\int d^3k \sum_\lambda \int_0^\infty d\omega \sqrt{\hbar\epsilon_I(\omega)/2\pi^3}\,\frac{\omega/c}{k^2-\epsilon(\omega)\omega^2/c^2}\hat{c}_\lambda(\mathbf{k},\omega)\,(\mathbf{k}\times\mathbf{e}_{\mathbf{k}\lambda}) \\ &\quad e^{-i(\omega t-\mathbf{k}\cdot\mathbf{r})} + \text{h.c.}\end{aligned} \tag{18.57}$$

These expressions have the same form as the corresponding ones obtained by Huttner and Barnett [11] by Fano diagonalization of the entire system of coupled harmonic oscillators (EM field, dielectric oscillators, and bath oscillators). Their equations for the quantized electric and magnetic fields, however, involve annihilation and creation operators $\hat{C}_\lambda(\mathbf{k},\omega)$ and $\hat{C}^\dagger_\lambda(\mathbf{k},\omega)$ for the exactly diagonalized Hamiltonian, instead of the reservoir annihilation and creation operators $\hat{c}_\lambda(\mathbf{k},\omega)$ and $\hat{c}^\dagger_\lambda(\mathbf{k},\omega)$ appearing in our expressions (18.56) and (18.57). Their diagonalized Hamiltonian, including the zero-point energy, is

$$H = \int d^3k \sum_\lambda \int_0^\infty d\omega\hbar\omega\left[\hat{C}^\dagger_\lambda(\mathbf{k},\omega)\hat{C}_\lambda(\mathbf{k},\omega) + 1/2\right]. \tag{18.58}$$

The situation here parallels that for the simple model employed in Sections 18.2 and 18.3, except that no rotating-wave approximation has been made, and that one deals with three coupled subsystems instead of two coupled subsystems: we

again arrive at results by a straightforward "Langevin" approach that are equivalent to those obtained by diagonalizing the complete Hamiltonian. Equations (18.56) and (18.57) are analogous to equation (18.10) obtained in the Langevin approach to the single oscillator coupled to a reservoir, whereas the Huttner-Barnett equations for the fields are analogous to equation (18.25) obtained by exact diagonalization. As in the model of Sections 18.2 and 18.3, results such as (18.56) and (18.57) obtained by the Langevin approach will reproduce those obtained by exact diagonalization for dielectric media in thermal equilibrium. To illustrate this we show in the next section that the *total* zero-point energy appearing in (18.58) follows exactly from our approach; the calculation also sheds light on some of the physics involved, and in particular on the role of the Langevin forces in maintaining equilibrium.

18.5 Energy Density

We consider now the total energy density of the system of dielectric atoms, their reservoirs, and the electromagnetic field, focusing for simplicity on the limit of zero temperature. We start from Poynting's theorem in the conventional notation, using the symmetrized Poynting operator $\hat{\mathbf{S}} = (c/8\pi)[\hat{\mathbf{E}} \times \hat{\mathbf{H}} - \hat{\mathbf{H}} \times \hat{\mathbf{E}}]$ and taking expectation values over the initial state of the system consisting of the field, the dielectric atoms, and the reservoir:

$$\oint \langle \hat{\mathbf{S}} \rangle \cdot \mathbf{n} da = -\frac{1}{8\pi} \int \langle \hat{\mathbf{E}} \cdot \frac{\partial \hat{\mathbf{D}}}{\partial t} + \frac{\partial \hat{\mathbf{D}}}{\partial t} \cdot \hat{\mathbf{E}} \rangle dV - \frac{1}{8\pi} \int \langle \hat{\mathbf{H}} \cdot \frac{\partial \hat{\mathbf{H}}}{\partial t} + \frac{\partial \hat{\mathbf{H}}}{\partial t} \cdot \hat{\mathbf{H}} \rangle dV. \quad (18.59)$$

The left-hand side gives the energy flux through a closed surface S and, given that we are assuming thermal equilibrium, must vanish. We identify the rate of change of the expectation value of the total energy density W as

$$\frac{\partial W}{\partial t} = \frac{1}{8\pi} \langle \hat{\mathbf{E}} \cdot \frac{\partial \hat{\mathbf{D}}}{\partial t} + \frac{\partial \hat{\mathbf{D}}}{\partial t} \cdot \hat{\mathbf{E}} \rangle + \frac{1}{8\pi} \frac{\partial}{\partial t} \langle \hat{\mathbf{H}}^2 \rangle, \quad (18.60)$$

and the assumption of thermal equilibrium implies that this must also vanish. In the case of interest $\hat{\mathbf{D}} = \hat{\mathbf{E}} + 4\pi \hat{\mathbf{P}}_\epsilon + \hat{\mathbf{K}}$, where $\hat{\mathbf{P}}_\epsilon$ is the part of the polarization giving rise to the dielectric permittivity $\epsilon(\omega)$ and $\hat{\mathbf{K}}$ is the noise polarization defined by (18.47). Thus $\hat{\mathbf{D}} = \hat{\mathbf{D}}_\epsilon + \hat{\mathbf{K}}$ and

$$\frac{\partial W}{\partial t} = \frac{\partial W_1}{\partial t} + \frac{\partial W_2}{\partial t}, \quad (18.61)$$

where

$$\frac{\partial W_1}{\partial t} = \frac{1}{8\pi}\langle \hat{\mathbf{E}} \cdot \frac{\partial \hat{\mathbf{D}}_\epsilon}{\partial t} + \frac{\partial \hat{\mathbf{D}}_\epsilon}{\partial t} \cdot \hat{\mathbf{E}} \rangle + \frac{1}{8\pi}\frac{\partial}{\partial t}\langle \hat{\mathbf{H}}^2 \rangle \tag{18.62}$$

and

$$\frac{\partial W_2}{\partial t} = \frac{1}{8\pi}\langle \hat{\mathbf{E}} \cdot \frac{\partial \hat{\mathbf{K}}}{\partial t} + \frac{\partial \hat{\mathbf{K}}}{\partial t} \cdot \hat{\mathbf{E}} \rangle. \tag{18.63}$$

Before proceeding with the calculation of W we note the following identity expressing conservation of energy:

$$\begin{aligned} \frac{\partial W}{\partial t} &= \langle \frac{\partial}{\partial t} \sum_j \left[\frac{1}{2} m \dot{\hat{\mathbf{x}}}_j^2 + \frac{1}{2} m\omega_0^2 \hat{\mathbf{x}}_j^2 \right] \delta^3(\mathbf{r} - \mathbf{r}_j) + \frac{1}{4\pi}\frac{\partial}{\partial t}\left[\hat{\mathbf{E}}^2 + \hat{\mathbf{H}}^2 \right] \\ &\quad + \sum_j [2\gamma(\frac{1}{2} m \dot{\hat{\mathbf{x}}}_j^2) - \dot{\hat{\mathbf{x}}}_j \cdot \mathbf{F}_{Lj}] \delta^3(\mathbf{r} - \mathbf{r}_j) \rangle. \end{aligned} \tag{18.64}$$

The first term is the rate of change of the energy density (kinetic plus potential) of the oscillators of the dielectric, while the second term is the rate of change of the energy density of the electromagnetic field. If there were no dissipation ($\gamma = 0$ and therefore $\mathbf{F}_{Lj} = 0$), the third term on the right would vanish, and W would be just the matter-plus-field energy density. The third term gives the rate of change of the energy density in the reservoirs; $2\gamma \sum_j (\frac{1}{2} m \dot{\hat{\mathbf{x}}}_j^2) \delta^3(\mathbf{r} - \mathbf{r}_j)$ is the rate of increase of energy density in the reservoir, equal to the rate at which the energy density of the dielectric oscillators decreases due to their coupling to their reservoirs, while $\sum_j \dot{\hat{\mathbf{x}}}_j \cdot \mathbf{F}_{Lj} \delta^3(\mathbf{r} - \mathbf{r}_j)$ is the rate of work per unit volume done by the Langevin forces of the reservoirs on the dielectric oscillators.

Using (18.41) and

$$\frac{\partial \hat{\mathbf{D}}_\epsilon}{\partial t} = -i \int_0^\infty d\omega \omega [\epsilon(\omega) \hat{\mathbf{E}}(\mathbf{r}, \omega) e^{-i\omega t} - \epsilon^*(\omega) \hat{\mathbf{E}}^\dagger(\mathbf{r}, \omega) e^{+i\omega t}], \tag{18.65}$$

and integrating over t, we obtain

$$\begin{aligned} W_1(\mathbf{r}, t) &= \frac{1}{8\pi} \sum_\lambda \int_0^\infty d\omega' \int_0^\infty d\omega \frac{\omega' \epsilon^*(\omega') - \omega \epsilon(\omega)}{\omega' - \omega} \langle \hat{\mathbf{E}}_\lambda(\mathbf{r}, \omega) \cdot \hat{\mathbf{E}}_\lambda^\dagger(\mathbf{r}, \omega') \rangle \\ &\quad e^{-i(\omega - \omega')t} + \frac{1}{8\pi} \langle \hat{\mathbf{H}}^2(\mathbf{r}, t) \rangle, \end{aligned} \tag{18.66}$$

since the zero-temperature expectation value $\langle \hat{\mathbf{E}}_\lambda^\dagger(\mathbf{r}, \omega) \cdot \hat{\mathbf{E}}_{\lambda'}(\mathbf{r}, \omega') \rangle = 0$ and $\langle \hat{\mathbf{E}}_\lambda(\mathbf{r}, \omega) \cdot \hat{\mathbf{E}}_{\lambda'}^\dagger(\mathbf{r}, \omega') \rangle = 0$ unless $\lambda = \lambda'$ and $\omega = \omega'$. It is convenient to rewrite

(18.66) as a sum of two identical terms and to interchange ω and ω' in one of these terms; this allows us to write

$$\begin{aligned} W_1(\mathbf{r},t) &= \frac{1}{8\pi}\sum_\lambda \int_0^\infty d\omega' \int_0^\infty d\omega \frac{\omega'\epsilon_R(\omega')-\omega\epsilon_R(\omega)}{\omega'-\omega}\langle \hat{\mathbf{E}}_\lambda(\mathbf{r},\omega)\cdot\hat{\mathbf{E}}^\dagger_\lambda(\mathbf{r},\omega')\rangle \\ & e^{-i(\omega-\omega')t} - \frac{i}{8\pi}\sum_\lambda \int_0^\infty d\omega' \int_0^\infty d\omega\,(\omega'\epsilon_I(\omega')+\omega\epsilon_I(\omega)) \\ & \frac{\langle \hat{\mathbf{E}}_\lambda(\mathbf{r},\omega)\cdot\hat{\mathbf{E}}^\dagger_\lambda(\mathbf{r},\omega')\rangle e^{-i(\omega-\omega')t} - \langle \hat{\mathbf{E}}_\lambda(\mathbf{r},\omega')\cdot\hat{\mathbf{E}}^\dagger_\lambda(\mathbf{r},\omega)\rangle e^{i(\omega-\omega')t}}{2(\omega'-\omega)} \\ & + \frac{1}{8\pi}\langle \hat{\mathbf{H}}^2(\mathbf{r},t)\rangle. \end{aligned} \tag{18.67}$$

It follows from (18.55) and (18.56) that the vacuum (zero temperature) expectation value is

$$\begin{aligned} \langle \hat{\mathbf{E}}_\lambda(\mathbf{r},\omega)\cdot\hat{\mathbf{E}}^\dagger_\lambda(\mathbf{r},\omega')\rangle &= \langle \hat{\mathbf{E}}_\lambda(\mathbf{r},\omega')\cdot\hat{\mathbf{E}}^\dagger_\lambda(\mathbf{r},\omega)\rangle \\ &= \frac{\hbar}{2\pi^3}\epsilon_I(\omega)\frac{\omega^4}{c^4}\int d^3k \frac{1}{|k^2-\epsilon(\omega)\omega^2/c^2|^2}\delta(\omega-\omega'). \end{aligned} \tag{18.68}$$

The first term in (18.67) is now evaluated using

$$\lim_{\omega'\to\omega}\frac{\omega\epsilon_R(\omega)-\omega'\epsilon_R(\omega')}{\omega-\omega'} = \frac{d}{d\omega}[\omega\epsilon_R(\omega)]. \tag{18.69}$$

We evaluate the second term by noting that the zeroth-order contributions in $(\omega-\omega')$ in the numerator cancel each other, while the first-order terms result in a contribution linear in the elapsed time t:

$$\lim_{\omega'\to\omega}\frac{e^{-i(\omega-\omega')t}-e^{i(\omega-\omega')t}}{2(\omega'-\omega)} = it. \tag{18.70}$$

Therefore

$$\begin{aligned} W_1(\mathbf{r},t) &= \frac{1}{8\pi}\frac{\hbar}{2\pi^3c^4}\sum_\lambda\int_0^\infty d\omega\left(\frac{d}{d\omega}[\omega\epsilon_R]+2t\omega\epsilon_I\right)\omega^4\epsilon_I \\ \int d^3k\frac{1}{|k^2-\epsilon\omega^2/c^2|^2} &+ \frac{1}{8\pi}\langle\hat{\mathbf{H}}^2(\mathbf{r},t)\rangle = \frac{\hbar}{8\pi^2c^3}\sum_\lambda\int_0^\infty d\omega\omega^3 n_R\frac{d}{d\omega}[\omega\epsilon_R] \\ &+\frac{1}{8\pi}\langle\hat{\mathbf{H}}^2(\mathbf{r},t)\rangle + t\frac{\hbar}{4\pi^2c^3}\sum_\lambda\int_0^\infty d\omega\omega^4 n_R\epsilon_I, \end{aligned} \tag{18.71}$$

where we have used the relations $\epsilon_R = n_R^2 - n_I^2$ and $\epsilon_I = 2n_R n_I$ between the real and imaginary parts of the permittivity $\epsilon(\omega)$ and the refractive index $n(\omega)$.

For the evaluation of W_2 it is convenient to define $\hat{\mathbf{K}}(\mathbf{k},\omega)$ by writing

$$\hat{\mathbf{K}}(\mathbf{r},t) = \int_0^\infty d\omega \int d^3k \sum_\lambda [\hat{\mathbf{K}}_\lambda(\mathbf{k},\omega) e^{-i\omega t} e^{i\mathbf{k}\cdot\mathbf{r}} + \hat{\mathbf{K}}_\lambda^\dagger(\mathbf{k},\omega) e^{i\omega t} e^{-i\mathbf{k}\cdot\mathbf{r}}], \quad (18.72)$$

and using (18.51), (18.54), and (18.56) to relate $\hat{\mathbf{K}}_\lambda(\mathbf{k},\omega)$ and $\hat{\mathbf{E}}_\lambda(\mathbf{k},\omega)$:

$$\hat{\mathbf{K}}_\lambda(\mathbf{k},\omega) = \frac{c^2}{\omega^2}[k^2 - \epsilon(\omega)\omega^2/c^2]\hat{\mathbf{E}}_\lambda(\mathbf{k},\omega). \quad (18.73)$$

After inserting (18.72) and (18.73) in (18.63) and performing some algebra we get

$$\begin{aligned} W_2(\mathbf{r},t) &= -\frac{\hbar}{16\pi^4 c^2} \sum_\lambda \int_0^\infty d\omega' \int_0^\infty d\omega \frac{\omega^2\omega'}{\omega - \omega'} \sqrt{\epsilon_I(\omega)\epsilon_I(\omega')}\,\delta(\omega - \omega') \\ &\quad \int d^3k \left[\frac{e^{-i(\omega-\omega')t}}{k^2 - \epsilon(\omega)\omega^2/c^2} + \frac{e^{i(\omega-\omega')t}}{k^2 - \epsilon^*(\omega)\omega^2/c^2} \right], \end{aligned} \quad (18.74)$$

and, proceeding as in the evaluation of W_1,

$$\begin{aligned} W_2(\mathbf{r},t) &= -\frac{\hbar}{8\pi^4 c^2} \sum_\lambda \mathrm{Re} \int_0^\infty d\omega' \int_0^\infty d\omega \frac{\omega^2\omega'}{\omega - \omega'} \sqrt{\epsilon_I(\omega)\epsilon_I(\omega')}\,\delta(\omega - \omega') \\ &\quad \times \int d^3k \frac{1}{k^2 - \epsilon(\omega')\omega'^2/c^2} - t\frac{\hbar}{4\pi^2 c^3} \sum_\lambda \int_0^\infty d\omega \omega^4 n_R\, \epsilon_I. \end{aligned} \quad (18.75)$$

We see that the time-dependent term in $W_2(t)$ exactly cancels the time-dependent term in $W_1(t)$.

The total energy density is obtained by adding (18.71) and (18.75) [13]:

$$\begin{aligned} W &= \frac{\hbar}{8\pi^2 c^3} \sum_\lambda \int_0^\infty d\omega \omega^3 \Big\{ \mathrm{Re}\left[n_R \frac{d}{d\omega}(\omega\epsilon) + \epsilon^{3/2} \right] \\ &\quad + \frac{1}{\omega}\epsilon_I \mathrm{Im} \frac{d}{d\omega}(\omega^2 \epsilon^{1/2}) \Big\}. \end{aligned} \quad (18.76)$$

Finally, using $\epsilon(\omega) = n^2(\omega)$ and the following relations

$$\begin{aligned} n_R \frac{d}{d\omega}(\omega\epsilon_R) &= (n_R^2 - n_I^2)n_R + \omega n_R \left(2n_R \frac{dn_R}{d\omega} - 2n_I \frac{dn_I}{d\omega}\right), \\ \mathrm{Re}\,\epsilon^{3/2} &= n_R^3 - 3n_R n_I^2, \\ \frac{\epsilon_I}{\omega}\mathrm{Im}\frac{d}{d\omega}(\omega^2\sqrt{\epsilon}) &= 4n_R n_I^2 + 2n_R n_I \omega \frac{dn_I}{d\omega}, \end{aligned} \tag{18.77}$$

and summing over polarizations, we obtain the vacuum expectation value of the total energy density:

$$\begin{aligned} W &= \frac{\hbar}{2\pi^2 c^3}\int_0^\infty d\omega\omega^3 n_R^2(\omega)\left(n_R + \omega\frac{dn_R}{d\omega}\right) \\ &= \frac{\hbar}{2\pi^2 c^3}\int_0^\infty d\omega\omega^3 n_R^2(\omega)\frac{d}{d\omega}[\omega n_R(\omega)]. \end{aligned} \tag{18.78}$$

This has the exactly the form expected had we ignored absorption entirely and simply posited that each mode of frequency ω and wavenumber $k = n_R(\omega)\omega/c$ has a zero-point energy $(1/2)\hbar\omega$, so that the energy density summed over all modes is

$$W = 2\left(\frac{1}{2\pi}\right)^3 \int d^3k \frac{1}{2}\hbar\omega. \tag{18.79}$$

The physical interpretation of this result is that the loss of energy due to absorption is balanced by the work done by the Langevin forces that maintain the canonical commutation relations of the dielectric oscillators.

18.6 Summary

Based on the simple model of a harmonic oscillator coupled to a reservoir, we showed how the Heisenberg equations of motion leading to a Langevin equation for the oscillator can give results equivalent to those obtained from the exact (Fano) diagonalization of the complete oscillator-reservoir system. We then used the model of a dielectric medium as a collection of harmonic oscillators, each oscillator coupled to a reservoir responsible for dissipation and a Langevin force as well as to the electromagnetic field, to derive the fluctuation-dissipation relation between the noise polarization arising from the Langevin forces and the imaginary part of the permittivity of the dielectric medium.

The simple oscillator-reservoir model we considered would suggest that the solutions for the electric and magnetic fields in a dielectric medium, with the

noise polarization as a source, might have the same form as obtained when the complete system of dielectric oscillators, reservoirs, and the electromagnetic field is diagonalized. We showed that this is in fact the case. Then we considered the *total* energy density of the complete system and showed explicitly that it is given by Eq. (18.78), which is exactly the form of the energy density obtained when absorption is ignored. In particular, we showed that a positive energy rate $\dot{W}_1 > 0$ arising from the interaction of the electromagnetic field with the dielectric oscillators is exactly canceled by a corresponding negative energy rate coming from the interaction of the system with the reservoir, $\dot{W}_2 = -\dot{W}_1 < 0$.

Acknowledgements

We thank S.M. Barnett, L.S. Brown, S.Y. Buhmann, I.E. Dzyaloshinskii, J.H. Eberly and R.F. O'Connell for helpful comments relating to this research. This work was funded by DARPA/MTO's Casimir Effect Enhancement program under DOE/NNSA Contract DE-AC52-06NA25396.

References

[1] T. Erber, "Poynting Vector and Energy Density in Dispersive Media," Bull. Classe des Sciences, Acad. Royale de Belgique **50**, 328 (1964).

[2] J. Schwinger, W.Y. Tsai, and T. Erber, "Classical and Quantum Theory of Synergic Synchrotron-Cerenkov Radiation," Ann. Phys. (N.Y.) **96**, 303 (1976); **281**, 1019 (2000).

[3] T. Erber, "The Classical Theories of Radiation Reaction," Fort. der Physik **9**, 343 (1961).

[4] T. Erber, "Testing the Randomness of Quantum Mechanics: Nature's Ultimate Cryptogram?", Ann. N.Y. Acad. Sci. **755**, 748 (1995).

[5] H.B.G. Casimir, "On the Attraction between Two Perfectly Conducting Plates," Proc. K. Ned. Akad. Wet. **51**, 793 (1948).

[6] N.G. van Kampen, B.R.A. Nijboer, and K. Schram, "On the Macroscopic Theory of van der Waals Forces," Phys. Lett. **26A**, 307 (1968).

[7] See, for instance, P.W. Milonni, *The Quantum Vacuum. An Introduction to Quantum Electrodynamics* (Academic, San Diego, 1994), and references therein.

[8] V.L. Ginzburg, *Theoretical Physics and Astrophysics* (Pergamon, Oxford, 1979), p. 309.

[9] Correct results (in agreement with Lifshitz) are nevertheless obtained: an integral over frequency is analytically continued in such a way that the permittivity appears as a function of purely imaginary frequencies, at which the *complex* permittivity, like the permittivity in the absence of absorption, is purely real.

[10] E.M. Lifshitz, "The Theory of Molecular Attractive Forces between Solids," Sov. Phys. JETP **2**, 73 (1956).

[11] B. Huttner and S.M. Barnett, Phys. Rev. A **46**, 4306 (1992).

[12] F.S.S. Rosa, S.Y. Buhmann, D.A.R. Dalvit, and P.W. Milonni, in preparation.

[13] Details of some of the material presented here are provided in a preprint by F.S.S. Rosa, D.A.R. Dalvit, and P.W. Milonni, "Electromagnetic Energy, Absorption, and Casimir Forces. I. Uniform Dielectric Media in Thermal Equilibrium.", arXiv:0911.2736.

[14] U. Fano, "Atomic Theory of Electromagnetic Interactions in Dense Materials," Phys. Rev. **103**, 1202 (1956).

[15] See Reference [13] for a discussion of the relation between the form of fluctuation-dissipation relation obtained here and that in Lifshitz's paper.

19

Exact Casimir Energies at Nonzero Temperature: Validity of Proximity Force Approximation and Interaction of Semitransparent Spheres

Kimball A. Milton[1] Prachi Parashar[2], Jef Wagner[3]
Oklahoma Center for High Energy Physics
Homer L. Dodge Department of Physics and Astronomy,
and K. V. Shajesh[4 5]
University of Oklahoma

PACS: 03.70.+k, 03.65.Nk, 11.80.La, 42.50.Lc

Abstract

In this paper, dedicated to the career of Tom Erber, we consider the Casimir interaction between weakly coupled bodies at nonzero temperature. For the case of semitransparent bodies, that is, ones described by δ-function potentials, we first examine the interaction

[1]email: milton@nhn.ou.edu, homepage: http://www.nhn.ou.edu/%7Emilton
[2]email: prachi@nhn.ou.edu
[3]email: wagner@nhn.ou.edu
[4]email: shajesh@nhn.ou.edu
[5]St. Edwards School, Vero Beach, FL, 32963-2699

between an infinite plane and an arbitrary curved surface. In weak coupling, such an interaction energy coincides with the exact form of the proximity force approximation obtained by summing the interaction between opposite surface elements at arbitrary temperature. This result generalizes a theorem proved recently by Decca et al. We also obtain exact closed-form results for the Casimir energy at arbitrary temperature for weakly-coupled semitransparent spheres.

19.1 Introduction

Since the earliest calculations of fluctuation forces between bodies [1], that is, Casimir or quantum vacuum forces, multiple scattering methods have been employed. Rather belatedly, it has been realized that such methods could be used to obtain accurate numerical results in many cases [2 - 5]. These results allow us to transcend the limitations of the proximity force theorem (PFT) [6, 7], and so make better comparison with experiment, which typically involve curved surfaces. (For a review of the experimental situation, see Ref. [8].)

The multiple scattering formalism, which is in principle exact, dates back at least into the 1950s [9 - 10]. Particularly noteworthy is the seminal work of Balian and Duplantier [11]. (For more complete references see Ref. [12].) This technique, which has been brought to a high state of perfection by Emig et al. [5], has concentrated on numerical results for the Casimir forces between conducting and dielectric bodies such as spheres and cylinders. For recent impressive numerical results for metals and dielectrics see Refs. [13 - 14]. Our group has noticed that the multiple-scattering method can yield exact, closed-form results for bodies that are weakly coupled to the quantum field [12, 15]. (That is, we are carrying out first-order perturbation theory in the background potential. For early examples of this in the Casimir context, see Ref. [16].) This allows an exact assessment of the range of applicability of the PFT. The calculations there, however, as those in recent extensions of our methodology [17] have been restricted to scalar fields with δ-function potentials, so-called semitransparent bodies. (These are closely related to plasma shell models [3, 18 - 21].) The technique was recently extended to dielectric bodies [22, 23], characterized by a permittivity ε. Strong coupling would mean a perfect metal, $\varepsilon \to \infty$, while weak coupling means that ε is close to unity.

In this paper we will extend the weak-coupling formalism to the situation of nonzero temperature. This extension is extremely straightforward. We then apply the general formula to the case of an arbitrarily curved semitransparent

surface above an infinite semitransparent plane. Remarkably, the result coincides with the use of the so-called proximity force approximation (PFA), which in its general form is exact in this case for all separations between the surfaces and for all temperatures. We also obtain exact closed-form results for the forces between separated spherical shells for all temperatures. In the Appendix we discuss exact formulas for arbitrary positive and negative powers of the distances between points on two spheres, needed for such calculations.

19.2 Multiple Scattering Derivation of Vacuum Energy between Weakly Coupled Potentials

The quantum vacuum energy for the interaction mediated by a massless scalar field between two nonoverlapping potentials $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ is

$$E = -\frac{i}{2}\mathrm{Tr}\,\ln(1 - V_1 G_1 V_2 G_2), \tag{19.1}$$

in terms of the single potential Green's functions

$$G_i = (1 + G_0 V_i)^{-1} G_0. \tag{19.2}$$

The free Green's function, satisfying

$$-\partial^2 G_0(x - x') = \delta(x - x'), \tag{19.3}$$

has the explicit form

$$G_0(x - x') = \int \frac{d\omega}{2\pi} \mathcal{G}_0(\mathbf{r} - \mathbf{r}', \omega) e^{-i\omega(t-t')}, \tag{19.4}$$

where the time-Fourier transform is

$$\mathcal{G}_0(\mathbf{r} - \mathbf{r}', i\zeta) = \frac{e^{-|\zeta||\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|}, \tag{19.5}$$

where we have performed the Euclidean rotation $\omega \to i\zeta$.

For weak potentials, the energy (19.1) simplifies dramatically:

$$E \approx \frac{i}{2}\mathrm{Tr}\, V_1 G_0 V_2 G_0 = -\frac{1}{64\pi^3} \int (d\mathbf{r})(d\mathbf{r}') \frac{V_1(\mathbf{r}) V_2(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{19.6}$$

At finite temperature the integral over imaginary frequency becomes the Matsubara sum:

$$\int_{-\infty}^{\infty} \frac{d\zeta}{2\pi} \frac{e^{-2|\zeta||\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2} \to T \sum_{m=-\infty}^{\infty} \frac{e^{-4\pi T|m||\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2} = \frac{T}{|\mathbf{r}-\mathbf{r}'|^2} \coth 2\pi T|\mathbf{r}-\mathbf{r}'|, \tag{19.7}$$

so the interaction energy becomes

$$E_T = -\frac{T}{32\pi^2} \int (d\mathbf{r})(d\mathbf{r}') V_1(\mathbf{r}) V_2(\mathbf{r}') \frac{\coth 2\pi T|\mathbf{r}-\mathbf{r}'|}{|\mathbf{r}-\mathbf{r}'|^2}, \tag{19.8}$$

which evidently reduces to Eq. (19.6) for $T = 0$.

19.3 Parallel plates

For parallel, semitransparent plates, separated by a distance a, where the potentials are

$$V_1(\mathbf{r}) = \lambda_1 \delta(z),\ V_2(\mathbf{r}) = \lambda_2 \delta(z-a), \tag{19.9}$$

the integrals in Eq. (19.6) are readily carried out, with the resulting energy per unit area A, $\mathcal{E} = E/A$:

$$\mathcal{E} = -\frac{\lambda_1 \lambda_2}{32\pi^2 a}. \tag{19.10}$$

This well-known result holds even if one of the plates has a finite area A. At finite temperature the result is

$$\mathcal{E}_T = -\frac{\lambda_1 \lambda_2 T}{16\pi} \int_{2\pi T a}^{\infty} \frac{dx}{x} \coth x. \tag{19.11}$$

The energy is ambiguous because it depends on the arbitrarily chosen upper limit. However, it corresponds to a well-defined pressure between the plates,

$$P_T = -\frac{\partial}{\partial a} \mathcal{E}_T = -\frac{\lambda_1 \lambda_2 T}{16\pi a} \coth 2\pi T a. \tag{19.12}$$

19.4 Interaction between an Infinite Plane and an Arbitrarily Curved Surface: PFA

Now consider the interaction between a semitransparent plane, described by the potential

$$V_1(\mathbf{r}) = \lambda_1 \delta(z), \tag{19.13}$$

and an arbitrary curved surface S, which does not intersect the plane $z = 0$, which corresponds to the potential

$$V_2(\mathbf{r}) = \lambda_2 \delta(z - s(x, y)), \tag{19.14}$$

where $z = s(x, y)$ is the equation of the surface. Then, from Eq. (19.8) it is immediate that the energy is (the upper limit of the x integration is again physically irrelevant)

$$E_T = -\frac{\lambda_1 \lambda_2 T}{16\pi} \int dS \int_{2\pi T z(S)} dx \frac{\coth x}{x}, \tag{19.15}$$

where the area integral is over the curved surface. This is precisely what one means by the proximity force approximation, where one sums energies between adjacent elements treated as parallel plates:

$$E_{\mathrm{PFA}} = \int dS \mathcal{E}_{\parallel}(z(S)), \tag{19.16}$$

in view of Eq. (19.11). This is in fact just the theorem proved by Decca et al. [24], who were considering gravitational and Yukawa type forces, but we see it applies to any central force.

For example, the above, exact formula for weakly-coupled semitransparent surfaces says that the force on such a sphere, of radius a, the center of which is a distance Z above a semitransparent plane is

$$F_T = -\frac{\partial E_T}{\partial Z} = -\frac{\lambda_1 \lambda_2 a T}{8} \int_{2\pi T(Z-a)}^{2\pi T(Z+a)} \frac{du}{u} \coth u. \tag{19.17}$$

The zero-temperature limit of this is

$$F = -\frac{\lambda_1 \lambda_2}{8\pi} \frac{a^2}{Z^2 - a^2}, \tag{19.18}$$

which may be alternatively derived from the zero-temperature energy

$$E = -\frac{\lambda_1 \lambda_2 a^2}{16\pi} \int_{-1}^{1} \frac{d\cos\theta}{Z + a\cos\theta} = -\frac{\lambda_1 \lambda_2 a}{16\pi} \ln \frac{Z + a}{Z - a}, \tag{19.19}$$

again, the exact PFA result.

19.5 Interaction between Two Semitransparent Spheres at Nonzero Temperature

Consider now two spheres, of radius a and b, respectively, with a distance between their centers $R > a + b$. In terms of local coordinates with origins at the centers of the two spheres, the semitransparent potentials are

$$V_1 = \lambda_1 \delta(r - a),\ V_2 = \lambda_2 \delta(r' - b), \tag{19.20}$$

and let us further suppose that $\mathbf{R}$ lies along the z axis of both coordinate systems. Then the squared distance between points on the spheres is

$$|\mathbf{r} - \mathbf{r}'|^2 = R^2 + a^2 + b^2 - 2ab\cos\gamma - 2R(a\cos\theta - b\cos\theta'), \tag{19.21}$$

in terms of polar angles in the two spheres, where the cosine of the angle between the two radial vectors locating the points is

$$\cos\gamma = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi'). \tag{19.22}$$

We insert this into the expression for the energy (19.8), obtaining

$$E = -\frac{\lambda_1\lambda_2 T}{32\pi^2} a^2 b^2 \int d\Omega\, d\Omega' \frac{\coth 2\pi T|\mathbf{r} - \mathbf{r}'|}{|\mathbf{r} - \mathbf{r}'|^2}. \tag{19.23}$$

It seems difficult to proceed in general, but we can work out a low temperature expansion using

$$\coth y = \sum_{n=0}^{\infty} \frac{2^{2n} B_{2n}}{(2n)!} y^{2n-1} = \frac{1}{y} + \frac{1}{3}y - \frac{1}{45}y^3 + \ldots,\ (|y| < \pi) \tag{19.24}$$

which will give rise to an expansion of the form

$$E_T = E_0 + T^2 E_2 + T^4 E_4 + \ldots. \tag{19.25}$$

The zero temperature result was worked out, by inspection, in Ref. [12]:

$$E_0 = -\frac{\lambda_1\lambda_2 ab}{16\pi R} \ln \frac{1 - (a-b)^2/R^2}{1 + (a+b)^2/R^2}. \tag{19.26}$$

The T^2 term is trivial because it is evaluated by Newton's theorem that a Coulomb potential exterior to a spherically symmetric charge distribution is as though the charge were concentrated at the center:

$$E_2 = -\frac{\lambda_1\lambda_2\pi}{3} \frac{a^2 b^2}{R}. \tag{19.27}$$

The T^{2n} term, $n > 1$ however, is slightly nontrivial:

$$E_{2n} = -\frac{\lambda_1\lambda_2}{64\pi^3}a^2b^2\frac{(4\pi)^{2n}B_{2n}}{(2n)!}\int d\Omega\, d\Omega' |\mathbf{r}-\mathbf{r}'|^{2n-3}. \tag{19.28}$$

We may evaluate the integrals by expanding in powers of $\hat{a} = a/R$ and $\hat{b} = b/R$:

$$\begin{aligned}
\int d\Omega\, d\Omega' |\mathbf{r}-\mathbf{r}'| &= (4\pi)^2 R\left[1 + \frac{1}{3}(\hat{a}^2+\hat{b}^2)\right] \\
\int d\Omega\, d\Omega' |\mathbf{r}-\mathbf{r}'|^3 &= (4\pi)^2 R^3\left[1 + 2(\hat{a}^2+\hat{b}^2) + \frac{1}{5}\hat{a}^4 + \frac{2}{3}\hat{a}^2\hat{b}^2 + \frac{1}{5}\hat{b}^4\right] \\
\int d\Omega\, d\Omega' |\mathbf{r}-\mathbf{r}'|^5 &= (4\pi)^2 R^5\Bigg[1 + 5(\hat{a}^2+\hat{b}^2) + 3\hat{a}^4 + 10\hat{a}^2\hat{b}^2 + 3\hat{b}^4 \\
&\quad + \frac{1}{7}\hat{a}^6 + \hat{a}^2\hat{b}^2(\hat{a}^2+\hat{b}^2) + \frac{1}{7}\hat{b}^6\Bigg],
\end{aligned} \tag{19.29}$$

and so on. The reason these are polynomials is evident when one considers the multipole expansion of the Coulomb potential—See, for example, Chap. 22 of Ref. [25]. For general formulas for such moments, see the Appendix.

By computing further terms in the sequence of polynomials, we are able to recognize the pattern:

$$\frac{1}{(4\pi)^2 R^{2n+1}}\int d\Omega\, d\Omega' |\mathbf{r}-\mathbf{r}'|^{2n+1} = \sum_{p=0}^{n+1}\sum_{q=0}^{p} A(n,p,q)\hat{a}^{2(p-q)}\hat{b}^{2q}, \tag{19.30}$$

where

$$A(n,p,q) = \frac{(2n+2)!}{(2n-2p+2)!(2p-2q+1)!(2q+1)!}. \tag{19.31}$$

When this is inserted into the low temperature expansion, we can remarkably sum the series:

$$\begin{aligned}
E_T &= -\frac{\lambda_1\lambda_2}{16\pi}\frac{ab}{R}\Bigg\{\ln\frac{1-(a-b)^2/R^2}{1-(a+b)^2/R^2} + f(2\pi T(R+a+b)) \\
&+ f(2\pi T(R-a-b)) - f(2\pi T(R-a+b)) - f(2\pi T(R+a-b))\Bigg\},
\end{aligned} \tag{19.32}$$

where f is

$$f(y) = \sum_{n=1}^{\infty}\frac{2^{2n}B_{2n}}{2n(2n-1)(2n)!}y^{2n}, \tag{19.33}$$

which is obtained from the second antiderivative of the hyperbolic cotangent:

$$y\frac{d^2}{dy^2}f(y) = \coth y - \frac{1}{y},\ f(0) = f'(0) = 0. \tag{19.34}$$

Although the power series expansion (19.33) is valid only for sufficiently low temperatures $2T(R+a+b) < 1$, the solution of the differential equation is valid for all values of T.

For sufficiently high temperatures we can replace the hyperbolic cotangent in the differential equation by 1, and then

$$f(y) \sim y\ln y + \ln y + Ay + B,\ y \gg 1, \tag{19.35}$$

where A and B are integration constants that do not contribute to Eq. (19.32). When this asymptotic solution is inserted into Eq. (19.32) the zero temperature logarithm cancels out, and we are left with

$$\begin{aligned} E_T \quad &\sim \quad -\frac{\lambda_1\lambda_2 ab}{8}T\left[\ln\frac{R^2-(a+b)^2}{R^2-(a-b)^2}\right. \\ &\left.+\frac{a}{R}\ln\frac{(R+b)^2-a^2}{(R-b)^2-a^2} + \frac{b}{R}\ln\frac{(R+a)^2-b^2}{(R-a)^2-b^2}\right],\ T\to\infty. \end{aligned} \tag{19.36}$$

This result may be derived directly from the high-temperature form

$$E_T \sim -\frac{\lambda_1\lambda_2 T a^2 b^2}{32\pi^2}\int d\Omega\, d\Omega' \frac{1}{|\mathbf{r}-\mathbf{r}'|^2},\ T\to\infty. \tag{19.37}$$

This again may be worked out by expanding in powers of the radii of the spheres. Computing the first several terms reveals the pattern:

$$\int d\Omega\, d\Omega' \frac{1}{|\mathbf{r}-\mathbf{r}'|^2} = \frac{(4\pi)^2}{R^2}\sum_{n=0}^{\infty}\frac{1}{(n+1)(2n+1)}\sum_{m=0}^{n}\frac{1}{2}\binom{2n+2}{2m+1}\hat{a}^{2(n-m)}\hat{b}^{2m}. \tag{19.38}$$

This sum is almost identical to that found for spheres at zero temperature, as seen in Eq. (6.15) of Ref. [12], which led to Eq. (19.26), except for the appearance of $1/(2n+1)$ here. Therefore, the former series must be obtained from the present series by differentiation. Denoting the double sum in Eq. (19.38) by S, it must be true that

$$R^2\frac{\partial}{\partial R}\frac{S}{R} = \frac{R^2}{4ab}\ln\left(\frac{1-(a+b)^2/R^2}{1-(a-b)^2/R^2}\right), \tag{19.39}$$

where S is $R^2/4ab$ times the square-bracketed quantity in Eq. (19.36). This equality is, in fact, easily verified. See the Appendix for the generalization of this result.

We compare the general form [obtained by numerically integrating Eq. (19.34)] and the high-temperature limiting form (19.36) in Fig. 19.1.

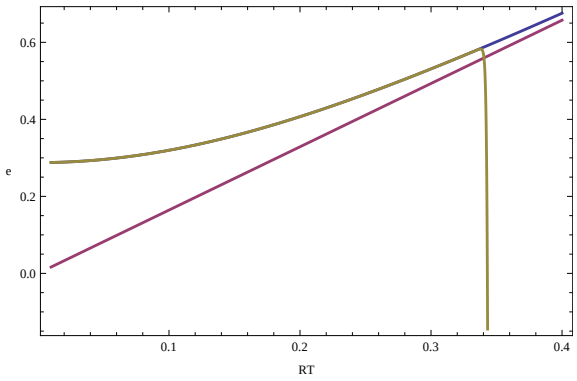


Figure 19.1: Comparison between the general and high temperature forms of the energy, as a function of RT. Energies are shown for $a = b = R/4$. The high temperature result is linear in T. Also shown is the power series expansion [Eq. (19.33) truncated at 200 terms], which diverges in this case at $RT = 1/3$. Plotted is $e = -16\pi RE/(\lambda_1\lambda_2 a^2)$.

19.6 Conclusions

We have shown that exact results may be found in weak coupling for the quantum vacuum forces between nontrivial bodies not only at zero temperature, but at finite temperature. We have shown that the exact form of the proximity force approximation holds exactly for all temperatures for the force between an infinite plane surface and an arbitrarily curved one. We have also computed the force between two semitransparent spheres at arbitrary temperatures, and obtain remarkably simple, closed-form expressions. The PFA equivalence evidently will hold for tenuous dielectric bodies in electromagnetism, and closed-form finite temperature results may be easily obtained between dielectric bodies as well.

Acknowledgements

K.A.M. first met Tom Erber at UCLA during one of Tom's many visits to Schwinger's group there in the 1970s. Thus it is a great pleasure to dedicate this paper to him. We thank the US National Science Foundation (Grant No. PHY-0554926) and the US Department of Energy (Grants Nos. DE-FG02-04ER41305 and DE-FG02-04ER-46140) for partially funding this research. We thank Nima Pourtolami and Elom Abalo for collaborative assistance.

19.7 Appendix: Mean Powers of Distances between Points on Spheres

In Sec. 6.5 we used exact evaluations of mean distances, defined by

$$\int d\Omega\, d\Omega' |\mathbf{r} - \mathbf{r}'|^p = (4\pi)^2 R^p P_p(\hat{a}, \hat{b}), \tag{19.40}$$

where R is the distance between the centers of the two nonoverlapping spheres, of radii a and b, respectively. Here $\hat{a} = a/R$ and $\hat{b} = b/R$, and $P_p(\hat{a}, \hat{b})$ can in general be represented by the infinite series

$$P_p(\hat{a}, \hat{b}) = \sum_{n=0}^{\infty} \frac{2}{(2n+2)!} \frac{\Gamma(2n-p-1)}{\Gamma(-p-1)} Q_n(\hat{a}, \hat{b}). \tag{19.41}$$

Here the homogeneous polynomials Q_n are

$$\begin{aligned} Q_0 &= 1, \\ Q_1 &= 2(\hat{a}^2 + \hat{b}^2), \\ Q_2 &= 3\hat{a}^4 + 10\hat{a}^2\hat{b}^2 + 3\hat{b}^4, \\ Q_3 &= 4\hat{a}^6 + 28\hat{a}^4 b^2 + 28\hat{a}^2\hat{b}^4 + 4\hat{b}^6, \end{aligned} \tag{19.42}$$

or in general,

$$Q_n = \frac{1}{2} \sum_{m=0}^{n} \binom{2n+2}{2m+1} \hat{a}^{2(n-m)} \hat{b}^{2m}. \tag{19.43}$$

We can easily see the following recursion relation holds:

$$P_{p-1}(\hat{a}, \hat{b}) = \frac{R^{-p}}{1+p} \frac{\partial}{\partial R} R^{1+p} P_p(\hat{a}, \hat{b}), \tag{19.44}$$

since Q_n is homogeneous in R of degree $-2n$.

For p a non-negative integer, P_p is a polynomial of degree $2\lceil p/2 \rceil$, and we can immediately find

$$\begin{aligned} P_p(\hat{a},\hat{b}) &= \frac{1}{4\hat{a}\hat{b}}\frac{1}{(p+2)(p+3)}\Big[(1+\hat{a}+\hat{b})^{p+3}+(1-\hat{a}-\hat{b})^{p+3} \\ &-(1-\hat{a}+\hat{b})^{p+3}-(1+\hat{a}-\hat{b})^{p+3}\Big],\ p=0,1,2,\ldots. \end{aligned} \tag{19.45}$$

For p a negative integer, we have

$$\begin{aligned} P_{-1} &= 1, \\ P_{-2} &= \frac{1}{4\hat{a}\hat{b}}\left[\ln\frac{1-(\hat{a}+\hat{b})^2}{1-(\hat{a}-\hat{b})^2}+\hat{a}\ln\frac{(1+\hat{b})^2-\hat{a}^2}{(1-\hat{b})^2-\hat{a}^2}+\hat{b}\ln\frac{(1+\hat{a})^2-\hat{b}^2}{(1-\hat{a})^2-\hat{b}^2}\right], \\ P_{-3} &= -\frac{1}{4\hat{a}\hat{b}}\ln\frac{1-(\hat{a}+\hat{b})^2}{1-(\hat{a}-\hat{b})^2}, \\ P_{-4} &= \frac{1}{[1-(\hat{a}+\hat{b})^2][1-(\hat{a}-\hat{b})^2]}, \end{aligned} \tag{19.46}$$

and further expressions can be obtained by use of Eq. (19.44).

References

[1] H. B. G. Casimir, "On the attraction between two perfectly conducting plates," Kon. Ned. Akad. Wetensch. Proc. **51**, 793 (1948).

[2] A. Wirzba, "The Casimir effect as scattering problem," J. Phys. A **41**, 164003 (2008) [arXiv:0711.2395 [quant-ph]].

[3] M. Bordag and V. Nikolaev, "Casimir force for a sphere in front of a plane beyond proximity force approximation," J. Phys. A **41**, 164002 (2008) [arXiv:0802.3633 [hep-th]].

[4] P. A. Maia Neto, A. Lambrecht, S. Reynaud, "Casimir energy between a plane and a sphere in electromagnetic vacuum," Phys. Rev. A **78**, 012115 (2008) [arXiv:0803.2444]; A. Canaguier-Durand, P. A. Maia Neto, I. Cavero-Peláez, A. Lambrecht, and S. Reynaud, "Casimir interaction between plane and spherical metallic surfaces," Phys. Rev. Lett. **102**, 230404 (2009).

[5] T. Emig, N. Graham, R. L. Jaffe and M. Kardar, "Casimir forces between arbitrary compact objects," Phys. Rev. Lett. **99**, 170403 (2007) [arXiv:0707.1862 [cond-mat.stat-mech]]; "Casimir forces between compact objects: I. The scalar case," Phys. Rev. D **77**, 025005 (2008) [arXiv:0710.3084 [cond-mat.stat-mech]]; T. Emig and R. L. Jaffe, "Casimir forces between arbitrary compact objects: Scalar and electromagnetic field," J. Phys. A **41**, 164001 (2008) [arXiv:0710.5104 [quant-ph]].

[6] J. Blocki, J. Randrup, W. J. Świątecki, and C. F. Tsang, "Proximity forces," Ann. Phys. (N.Y.) **105**, 427 (1977).

[7] B. V. Deryagin, "Untersuchung über die Reibung und Adhäsion. IV," Kolloid Z. **69**, 155 (1934).

[8] R. Onofrio, "Casimir forces and non-Newtonian gravitation," New J. Phys. **8**, 237 (2006) [arXiv:hep-ph/0612234].

[9] M. G. Krein, "O formule sledov v teorii vozmushcheniy," Mat. Sb. (N.S.) **33**, 597 (1953); "Perturbation determinants and a formula for the traces of unitary and self-adjoint operators," Dokl. Akad. Nauk SSSR **144**, 268 (1962) [Sov. Math. Dokl. **3**, 707 (1962)]; M. Sh. Birman and M. G. Krein, "On the theory of wave operators and scattering operators," Dokl. Akad. Nauk SSSR **144**, 475 (1962) [Sov. Math. Dokl. **3**, 740 (1962)].

[10] M. J. Renne, "Microscopic theory of retarded van der Waals forces between macroscopic dielectric bodies," Physica **56**, 125 (1971).

[11] R. Balian and B. Duplantier, "Geometry of the Casimir effect," arXiv:quant-ph/0408124, in the proceedings of 15th SIGRAV Conference on General Relativity and Gravitational Physics, Rome, Italy, 9–12 September 2002; "Electromagnetic waves near perfect conductors. 2. Casimir effect," Ann. Phys. (N.Y.) **112**, 165 (1978); "Electromagnetic waves near perfect conductors. 1. Multiple scattering expansions, distribution of modes," Ann. Phys. (N.Y.) **104**, 300 (1977).

[12] K. A. Milton and J. Wagner, "Multiple scattering methods in Casimir calculations," J. Phys. A **41**, 155402 (2008) [arXiv:0712.3811 [hep-th]].

[13] S. J. Rahi, T. Emig, N. Graham, R. L. Jaffe, and M. Kardar, " Scattering theory approach to electrodynamic Casimir forces," arXiv:0908.2649.

[14] M. T. Homer Reid, A. W. Rodriguez, J. White, and S. G. Johnson, "Efficient computation of Casimir interactions between arbitrary 3D objects," Phys. Rev. Lett. **103**, 040401 (2009).

[15] K. A. Milton and J. Wagner, "Exact Casimir interaction between semitransparent spheres and cylinders," Phys. Rev. D **77**, 045005 (2008), [arXiv:0711.0774 [hep-th]].

[16] M. Bordag, "On the vacuum interaction of two parallel cosmic strings," Annalen d. Phys. **47** [**502**], 93 (1990); "The vacuum interaction of magnetic strings," Ann. Phys. (N.Y.) **206**, 257 (1991).

[17] I. Cavero-Peláez, K. A. Milton, P. Parashar and K. V. Shajesh, "Non-contact gears: I. Next-to-leading order contribution to lateral Casimir force between corrugated parallel plates," Phys. Rev. D **78**, 065018 (2008) [arXiv:0805.2776 [hep-th]; "Non-contact gears: II. Casimir torque between concentric corrugated cylinders for the scalar case," Phys. Rev. D **78**, 065019 (2008) arXiv:0805.2777 [hep-th]]; J. Wagner, K. A. Milton and P. Parashar, "Weak coupling Casimir energies for finite plate configurations," J. Phys. Conf. Ser. **161**, 012022 (2009) [arXiv:0811.2442 [hep-th]]; K. A. Milton, "Recent developments in the Casimir effect," J. Phys. Conf. Ser. **161**, 012001 (2009) [arXiv:0809.2564 [hep-th]].

[18] M. Bordag and N. Khusnutdinov, "On the vacuum energy of a spherical plasma shell," Phys. Rev. D **77**, 085026 (2008) [arXiv:0801.2062 [hep-th]].

[19] M. Bordag, "Interaction of a charge with a thin plasma sheet," Phys. Rev. D **76**, 065011 (2007) [arXiv:0704.3845].

[20] G. Barton, "Casimir effects for a flat plasma sheet: I. Energies," J. Phys. A **38**, 2997 (2005).

[21] M. Bordag, "Generalized Lifshitz formula for a cylindrical plasma sheet in front of a plane beyond proximity force approximation," Phys. Rev. D **75**, 065003 (2007) [arXiv:quant-ph/0611243].

[22] K. A. Milton, P. Parashar and J. Wagner, "Exact results for Casimir interactions between dielectric bodies: The weak-coupling or van der Waals Limit," Phys. Rev. Lett. **101**, 160402 (2008) [arXiv:0806.2880 [hep-th]].

[23] K. A. Milton, P. Parashar and J. Wagner, "From multiple scattering to van der Waals interactions: Exact results for eccentric cylinders," in The

Casimir effect and cosmology, ed. S. D. Odintsov, E. Elizalde, and O. B. Gorbunova, in honor of Iver Brevik (Tomsk State Pedagogical University) pp. 107-116 (2009), arXiv:0811.0128 [math-ph].

[24] R. S. Decca, E. Fischbach, G. L. Klimchitskaya, D. E. Krause, D. Lopez and V. M. Mostepanenko, "Application of the proximity force approximation to gravitational and Yukawa-type forces," Phys. Rev. D **79**, 124021 (2009) [arXiv:0903.1299 [quant-ph]].

[25] J. Schwinger, L. L. DeRaad, Jr., K. A. Milton, and W.-y. Tsai, Classical Electrodynamics (Perseus/Westview, New York, 1998).

20

Contribution to Tom Erber Festschrift

Randall D. Peters
Professor and Chair Emeritus, Physics Department
Mercer University[1]

I first met Tom approximately twenty years ago. Following the creation of a capacitive sensor in 1987, I began using my invention to study mechanical oscillators. Subsequently, one of my abstracts titled "Non-classical Pendulum" was published in Proceedings of the Texas Section of the American Physical Society. Following its distribution by APS, Tom sent me a letter, requesting details of what I had measured. His letter came as a surprise to me, especially since this one who was interested in my experiment was a distinguished theorist. Moreover, in the tradition of some historically significant physicists–it appeared that Tom had already chosen to 'get down and dirty' through personal involvement with some experiments that seemed to concern only engineers.

I provided Tom with a description of my pendulum and elaborated on the free-decay patterns that it had exhibited. At low energy levels, these were no longer exponential, but rather they exhibited discrete energy states that appeared to be separated from one another by multiples of about $11\,pJ$. Tom responded to my letter with a comment that I won't forget: "Randall, let me amuse you with some speculation". He proceeded then to tell me the following: Every good student of physics knows that a 'field' multiplied by an 'area' yields a flux. He then noted how (within a factor of π) the characteristic magnetic field strength of quantum electrodynamics — a scale of type Feynman would have

[1]Macon, Georgia 31207

recognized — when multiplied by the square of the Compton wavelength, yields the flux quantum of superconducting fame. Then he noted how a good student also knows that the same field, when squared and multiplied by a volume yields an energy. Upon squaring the same field and multiplying by the cube of the Compton wavelength, Tom obtained the expression (mc^2/α), where m is the mass of the electron, c is the speed of light, and α is the fine structure constant. In other words, he had obtained what a Nobel Laureate more recently indicated to me at the celebration honoring the 90^{th} birthday of Charles Townes–was the Compton energy of $11\,pJ$. Because these operations had been done with pre-system-international units with which I was not very familiar, I was unable to duplicate Tom's exercise in complete detail. Several years later I looked at the matter in a somewhat different manner (no doubt consistent with Tom's analysis). By the means described in the paper titled "Compton energy scale of friction quantization" I arrived at the same result.

I am especially impressed by the depth of Tom's understanding in so many seemingly disparate fields of physics and mathematics. Never in my previous experience had I encountered anybody with a PhD in a field as esoteric as Compton scattering; who would display even a short-lived fancy with the internal friction of metals. Yet it was that very connection that is responsible for the 'hysteron' that we have envisioned as a yet-to-be-proven mesoscale quantum of internal friction.

Multi-disciplinary expertise has allowed Tom to exhibit rare examples of creativity. An especially notable case is the topic of his lecture to the physics department at Mercer University in the Fall of 2004. His recognition of the importance of the mesoscale to magnetic phenomena had motivated him to propose and then participate with an engineering colleague in a seminal study of piezomagnetism. After the fact, it is obvious to many in applied physics and engineering, that magnetic field data should provide considerable insight concerning the incipient failure of a ferrous sample subjected to large cyclic stress. Although others may have thought about the matter, they evidently lacked the background and depth of experience with which to recognize two critical factors required for success–these being the use of

- Helmholtz-like coils to cancel the Earth's otherwise overwhelming field
- a sensitive flux-gate magnetometer to measure the field changes.

In the summer of 2005 I apprised Tom of a paper that I had written, which may be the best-to-date experimental evidence for the hysteron. The paper is titled "Mesoscale quantization and self organized stability". In his comments

concerning this article, Tom once again demonstrated the value of broad-yet-deep understanding of physics. He mentioned how "the measurement of specific heats is analogous to measuring the time decay of Q (the quality factor that I had measured)–but Q decay gives STRUCTURAL INFORMATION, whereas heat dissipation has no memory!"

There have been several occasions for Tom to both improve my understanding of physics, and also to recommend simple means to increase the probability that my publications would gain wider acceptance. An example of the latter involves nonlinear damping. By little and little, it is increasingly recognized as not just important, but pervasive; yet many scientists and engineers resist the idea of this type of nonlinearity. They prefer to cling to the oft-meaningless viscous model of the harmonic oscillator, that is treated in popular textbooks as 'gospel'. It was Tom's recommendation of the title "Beyond the linear damping model for mechanical harmonic oscillators", that I chose to use for my paper on this subject. After its publication I began to more fully realize the wisdom of one's careful choice of words, especially when it comes to controversial subjects.

I am also very grateful for a number of stories that Tom has shared with me through the years. He has been privy to the details of adversity faced by several well-known physicists. To hear how even they faced unexpected resistance, to the new ideas they were presenting, was a source of considerable encouragement to me. I may have ceased work on some of my most successful projects, had Tom not provided me with such encouragement. On one occasion he exhorted me to 'keep on keeping on' by way of the following comment: "Randall, watch out for the steam-rollers when your ideas become mainstream". The most controversial ones have not yet become mainstream. If and when they do, I am confident that Tom's prediction will prove true; that a host of individuals will try to claim credit for ideas for which they did not 'pay the price'.

In the matter just described I clearly saw at work another of Tom's unusual attributes. Sometimes I have marveled at the low level of common-sense exhibited by some in our profession. Though they demonstrate, in the specialty of their profession, intelligence that is clear to all; they seem 'clueless' when it comes to issues of the social world that surrounds them. From my discussions with him on topics other than physics, I have concluded that Tom Erber stands in stark contrast to these people.

Less tangible benefits of my friendship with Tom are certainly to be inferred from these comments. Their detailed description, however, is more difficult for me to articulate than the technical benefits. Thus I will speak only to the ones derived from the following two occasions. I fondly remember a meal at a restaurant on the banks of the Hudson River, shared with Tom and Audrey

about the time of Easter 1996. This was during my two-year stint as a Visiting Professor at the U. S. Military Academy at West Point, and I had managed to get Tom to come and give a lecture. As with their years later trip to Macon in 2004, I was delighted that Audrey was able to accompany him.

21

Are the Navier Stokes Equations of Hydrodynamics an Example of a Gödel Theorem in Physics?

Seth Putterman
University of California, Los Angeles

One of the great theorems of mathematics of the 20th century is Gödel's theorem which states that within the framework of number theory there are an infinite number of propositions whose truth or falsity cannot be decided by application of the well known axioms. Even mathematicians are not in agreement as to the scope and 'meaning' of this theorem. By way of an example consider the continuation of the Riemann Zeta function to the complex plane $z = x + iy$:

$$\zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x} \rightarrow \zeta(z)$$

So far all the zeros of this function, and over 10^{13} have been found, lie on the line $x = 1/2$. To date there is no proof that this should be so. Could this be an example of a Gödel's theorem in number theory: a true statement that is itself an additional independent axiom? In a recent issue Notices of the American Mathematical Society [7] experts disagreed as to whether a well defined theorem about a function such as $\zeta(z)$ should fall under the scope of Gödel's theorem. Although there is no proof that Riemann's assertion is decidable [which is less stringent than proving its truth or falsity] some mathematicians expressed the opinion that it would someday be decided according to the basic axioms. Other

mathematicians say that even the issue of the decidability of Riemann's conjecture is up in the air. If the mathematicians can't agree on the decidability of such an interesting issue then it would appear to be ripe for physicists to ask similar questions about the natural world. Are there experimentally testable theories which cannot be reduced to Quantum Mechanics [or Newton's laws]? This issue surely underlies the paper "A Theory of Everything" by Laughlin and Pines [1].

The equations of hydrodynamics constitute 5 coupled nonlinear differential equations for five field variables:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) &= 0 \\ \rho\, T \frac{Ds}{Dt} - \kappa \nabla^2 T &= \Sigma > 0 \\ \rho \frac{Dv}{Dt} &= -\nabla p + \eta \nabla^2 T \\ p &= p(\rho, s) \\ \kappa (\nabla T)^2 + \eta \left[\frac{\partial v_i}{\partial r_j}\right]^2 &= \Sigma \end{aligned}$$

where $\rho(r,t)$, $v(r,t)$, $s(r,t)$ are the mass density, velocity and entropy per gram as a function of position and time. Other quantities such as temperature T, and pressure p are related to these variable via equations of state. The viscosity and thermal conductivity are η, κ and the entropy production is Σ; for compactness the bulk viscosity has been omitted. These equations follow uniquely from the assumptions that 5 variables are a complete set and that the energy depends on the local values of these variables [the continuum mechanics of paint would require nonlocal equations of state]. The derivation of these equations proceeds from conservation laws, Galilean covariance, the second law of thermodynamics and an approximation that the deviation from equilibrium is characterized [in the dynamical equations] by first order gradients such as ∇T. Although these equations are written in stone and although they are very heavily tested by experiments, they have never been derived from first principles in a controlled approximation. Even for a weakly interacting dilute gas, where the wavelength of an initial condition is large compared to the mean free path of the molecules, these equations have not been derived as say the leading order of a controlled expansion.

First, let us review the reason that the Boltzmann equation cannot be used as a basis for deriving the Navier-Stokes equations. By introducing σ = cross-

section for initial velocities to turn over into final velocities as a result of the binary collision: $v + v_1 \rightarrow v' + v_1'$ one writes down the Boltzmann equation:

$$\frac{\partial f}{\partial t} = -\int \sigma \, |v - v_1| \; f(v) \, f(v_1) \, dv_1 + \text{restituting collisions}$$

where $f(r, v, t)$ is the density of particles in position and velocity space. This is a closed equation which over time evolves to the Maxwell-Boltzmann distribution:

$$f(r, v, t) \rightarrow \exp\left[-\frac{M \, v^2}{2k \, T}\right] \text{ as } T \rightarrow \infty$$

and furthermore the last stages of this approach to equilibrium agree with the Navier- Stokes hydrodynamics when the proper moments of this equation are taken. The problem with this approach lies in what Ehrenfest labeled as the Stosszahl Ansatz: namely that the mathematically correct equation should be of the form

$$\frac{\partial f}{\partial t} \approx \int \sigma \, F_2 \, d \cdots$$

where $F_2(r, v, v_1, t)$ is the joint or 2 particle distribution function. In order to arrive at a closed equation Boltzmann assumed that the probability of the joint distribution is the product of the single particle probabilities or:

$$F_2(v, v_1) = f(v) \, f(v_1)$$

The equation for F_2 involves F_3:

$$\frac{\partial F_2}{\partial t} \approx F_3$$

and so on, so that unless an assumption is made a closed equation is not achieved See for example Uhlenbeck and Ford, Lectures in Statistical Mechanics [2].

The closure assumption is equivalent to assuming the validity of the Navier-Stokes equations of hydrodynamics. To appreciate the physical origin of the interconnectivity between the levels of description consider a sphere [of radius R] oscillating in a fluid with a velocity given by $u(t)$. According to Landau and Lifshitz: Fluid Mechanics [3], the force as a function of time is:

$$\text{Force}(t) = -6\pi \, \eta \, R \, u + AR^2 \sqrt{\eta \, \rho} \int_{-\infty}^{t} \frac{du}{d\tau} \frac{d\tau}{\sqrt{t - \tau}}$$

The long time memory indicated by the fluid equations as expressed by the last term on the right hand side of this equation leads to a velocity autocorrelation

that also has a long time memory and has appeared in molecular dynamic simulations of Alder and Wainwright [4]:

$$< u(t)\, u(0) > \approx A \exp[-t/\tau] + B\, \frac{1}{t^{3/2}}$$

This means that when we let R shrink to atomic dimensions the motion of a molecule maintains a long term memory of its past motion The response of a molecule depends on the usual collision time τ as well as the hydrodynamic times. Or, that in order to describe the motion of an individual molecule one needs the equations of hydrodynamics which one was supposedly about to derive. Kadanoff [5] said it this way: "these long time tails remained a perplexing mystery for a long time but now they are pretty well understood to be a consequence of the hydrodynamic motion of the fluid as it flows past its own molecules." That is: one cannot say that the micro world is characterized only by short time scales or fast processes. So one cannot derive the macro world by averaging over the fast processes which take place at the micro-scale..

The prize offered by the Clay institute [6] for the proof of the existence of a solution to the Navier Stokes equations under certain conditions implies that a derivation of these equation from Newton's laws does not exist. If such a derivation existed one could use it to map hydrodynamics onto a linear set of equations whose solution is then guaranteed. The interconnectedness of the micro and macro scales even as the number of particles goes to infinity suggests that the conditions for Gödel's insights could be at work in hydrodynamics. I propose that it is interesting to determine the variety of situations and causes which lead to underivable yet experimentally testable theories of physics. Perhaps new axioms of Physics are all around us.

References

[1] R. E. Laughlin and D. Pines "The Theory of Everything", Proc. Nat. Acad. Sci. 97 (2000) 28-31.

[2] G. E. Uhlenbeck and G. V. Ford, Lectures in Statistical Mechanics, Amer. Math. Soc. (1963).

[3] L.D. Landau and E. M. Lifshitz, Fluid Mechanics: Course in Theoretical Physics, Vol. 6, Butterworth-Hummerman (2000).

[4] B. J. Alder and T. G. Wainwright, "Decay of Velocity Autocorrelation Function", Phys. Rev A1 (1970) 18-21.

[5] L. P. Kadanoff, "Excellence in Computer Simulations", Computing in Science and Engineering 6 (2001) 27-37.

[6] Clay Mathematics Institute: Millenium Problems: Navier-Stokes Equations:

http://www.claymath.org/millennium/Navier-Stokes_Equations/

[7] Notices of the American Mathematical Society, Vol. 53, no. 4, April 2006, "A Tribute to Kurt Gödel".

22

Completion of Classical Dynamics of Charges

Fritz Rohrlich
Department of Physics
Syracuse University

This paper is written in honor of Thomas Erber, an old friend and colleague. His many valuable contributions to physics and specifically electrodynamics are much appreciated.

Abstract

The classical dynamics of charges is reviewed. The old equations by Abraham and Lorentz (1904) have long been known to be defective. The correct equations of motion are known as the Landau-Lifshitz equations. In first order perturbation expansion, quantum electrodynamics of point charges leads in its classical limit to these equations thus confirming them. Their applicability is however restricted to classical charges that can be approximated by point charges. Otherwise, the dynamics of charges cannot be described by differential equations, and integro-differential equations become necessary. The theory is now problem-free.

22.1 History

The complete classical description of the interaction between electric charges and electromagnetic fields requires the simultaneous solution of two systems

of equations. The first system is the set of Maxwell equations for the fields that are incident on the charges as well as those that act between the moving charges. The second system consists of the equations of motion of the charges that are driven by these fields and consequently emit additional fields of their own. This difficult problem is necessarily broken up into two separate tasks: the solution of the field equations when the motions of the charges are given, and the solution of the charged particle equations when the impinging fields are given. While the field equations have been well understood since Maxwell, the particle equations have been under dispute ever since the fundamental work by Abraham and Lorentz (1904) and the later manifestly covariant work by Dirac (1939) [1]. The 1904 work preceded the discovery of special relativity but resulted in equations that are relativistically correct.

The Abraham-Lorentz-Dirac equation (ALD) of 1904 and 1939 is known to have serious defects. The most obvious defect is that these differential equations are of third order in time derivatives and therefore require more conditions than the initial position and velocity of the particle that are necessary and sufficient for Newton's equation of motion .

Much of the work on charged particle dynamics until the end of the 20^{th} century was devoted to attempts to remedy the defects of the ALD equation. The equation of motion first published in a book by Landau and Lifshitz in 1951 [2], was mostly ignored. The otherwise very comprehensive review by Erber noted only its existence [3]. The Landau-Lifshitz equations (LL equations) attracted attention only after Spohn used them in 2000 as an example in his mathematical study of the ALD equations [4]. He pointed out that the solution space of the ALD equation is restricted to a submanifold by the condition that the acceleration vanishes asymptotically. That restriction is satisfied for external forces of finite duration by all the solutions of the LL equation. Later, I provided the physical conditions that lead from the ALD equations to the LL equations [5]. They correspond to Spohns restriction of the ALD solution manifold.

In the book by Landau and Lifshitz [2], the LL equation was presented as if it were an immediate and trivial consequence of the ALD equation. This is not the case. The lack of reasoning in their text was not corrected in any of their later editions. Below, I shall show how a careful derivation of the LL equation from the ALD equation leads to the specification of a domain of validity for the LL equation [5].

In local quantum electrodynamics (QED), a charged particle is represented by a point charge and the physical electron is empirically indeed of that nature. (Suitable renormalization techniques exist to deal with the Coulomb divergence.) But in classical electrodynamics (CED) an electric charge is phe-

nomenologically always part of a finite size body: there are no point charges in the classical world. The physically correct description of a classical charge is therefore a spatially extended charge distribution. This entails a sacrifice: the particle dynamics in general cannot be represented by a differential equation. Rather, an integro-differential equation of motion must be used for an extended charge. Such an equation has been derived, for example, by Medina [6]. Nevertheless, it is often mathematically convenient to treat a charged particle in classical physics as if it were a point charge. I shall present the conditions under which such an approximate treatment is justified.

22.2 Differential Equation of Motion

The ALD equation is

$$\begin{aligned} m\dot{v}^\mu &= F^\mu + m\,\tau_0\,P^{\mu\nu}\,\ddot{v}_\nu \qquad (22.1)\\ \text{where} \qquad P^{\mu\nu} &= \eta^{\mu\nu} + v^\mu\,v^\nu/c^2 \end{aligned}$$

is the projection operator into the hyperplane perpendicular to the four-velocity v^μ. The force, F^μ is the Lorentz force and/or any other force acting on the particle. The parameter τ_0 is defined by

$$c\tau_0 = \frac{2\,q^2}{3\,m\,c^2} \qquad (22.2)$$

The particle is characterized by its mass m and its charge q. The second term on the right in (22.1) consists of two parts, the Schott term, $\tau_0\ddot{v}^\mu$, and the loss due to radiation, the fourvector, $-v^\mu R$, where R is the invariant radiation rate (the Larmor formula) $R = m\,\tau_0\,a^\mu\,a_\mu = -m\,\tau_0\,v^\mu\,\ddot{v}_\mu$. From here on, I shall use units such that $c = 1$.

The ALD equations (22.1) are obviously physically incorrect: they contain a third time derivative of position that entails the need for a third initial condition, and they lead (among others) to physically meaningless solutions that violate causality (pre-acceleration and post-deceleration).

If one assumes the τ_0 term is so small that higher powers of τ_0 are negligible, one can formally replace $\ddot{v}^\mu$ on the right of (22.1) by $\dot{F}^\mu/m$. One then obtains the equation

$$m\,\dot{v}^\mu = F^\mu + \tau_0\,P^{\mu\nu}\,\dot{F}_\nu \qquad (22.3)$$

If the force depends on $x^\mu(\tau)$ rather than directly on τ, $F^\mu = F^\mu(x(\tau))$ then

$$\dot{F}^\mu = v_\alpha\,\partial^\alpha\,F^\mu$$

Equation (22.3) is the LL equation. If the force is the Lorentz force its dependence on the velocity requires [in the step from (22.1) to (22.3)] a replacement of the acceleration and of its derivative on the right hand side by F/m and $\dot{F}^\mu/m$, respectively. The LL equation can then be written in the form

$$m\,\dot{v}^\mu = F^\mu + \tau_0\,\left[v_\alpha\,\partial^\alpha\,(q\,F^{\alpha\beta})\,v_\beta + \frac{1}{m}\,(q\,F^{\mu\alpha}\,F_\alpha - v^\mu\,F^\alpha\,F_\alpha)\right] \qquad (22.4)$$

where the Lorentz force is $F^\mu = q\,F^{\mu\nu}\,v_\nu$ and the last term in the brackets is the rate of radiation loss. The first two terms in the square bracket combined correspond to the "Schott" term of the ALD equation (22.1). There are no accelerations on the right hand side. Actually, only this special case (when F is the Lorentz force) appears in the book by LL.

The validity of the LL equation (22.3) (or (22.4) in the Lorentz force case) must now be examined carefully. Its derivation from the ALD equation (22.1) is based on the assumption that the τ_0 term in (22.1) is so small that higher powers of τ_0 are negligible. That assumption was essential in the step from the ALD equation to the LL equation. Since the actual classical physical charge is necessarily extended, the assumption of a point particle is valid only under certain conditions: it must be assumed that the charge is of sufficiently small spatial dimension so that the external force, $\vec{F}$, varies negligibly over its volume. The particle can then be treated as a "pseudo-point" charge. Since the size of the particle is given, this requirement imposes a condition on the external force in (22.3) that reads in its instantaneous rest frame,

$$\left|\tau_0\,\frac{d\vec{F}}{dt}\right| \ll |\vec{F}\,| \qquad (22.5)$$

This is the condition for the charged particle to be treated as a pseudo-point charge. Only if (22.5) holds can one treat the last term of the LL Eq.(22.3) or (22.4) as a small correction. It follows that condition (22.5) on the external force is a necessary condition for the validity of the LL equation. In particular, Eq.(22.5) implies that the frequency of an external electromagnetic force proportional to $\exp(i\omega t)$ be limited by

$$\omega \ll \frac{3\,m\,\pi}{q^2}$$

In the literature, the condition (22.5) is often ignored. As a consequence, physically meaningless results are obtained. The above conditions, for instance, do

not permit a time dependence of the external forces that increases too fast. That implies that they cannot be step functions of time. The latter assumption is made incorrectly in some textbooks (see for instance [7]). I have given an example that shows explicitly how closely an external force can be permitted to simulate a step function [5].

22.3 Consequences for Classical Electrodynamics

For the purpose of the following general considerations, I shall first deal with the dynamics of pseudo-point charges and electromagnetic fields. The theory is based on the electromagnetic field equations (both free fields and fields emitted by pseudo-point charges), and the equations of motion of pseudo-point charges, i.e. Maxwell's equations and the LL equations of motion. Non-electromagnetic external forces may also be included. But this set of equations must now be supplemented by the validity conditions (22.5) for the LL equations.

It is unfortunate that most treatments of electromagnetic systems do not use the LL equations but only their Newtonian approximation that omits the τ_0 term. The presence of the τ_0 term makes an enormous difference in the dynamics of particles because it invalidates well-known claims commonly made in Newtonian classical mechanics. These differences will here be illustrated for the dynamics of a single charge but hold, of course, for almost all electrodynamic systems.

(a) Charged particle systems are open in the sense that whenever a force acts on a charge, electromagnetic radiation is emitted that escapes to infinity. Thus, a loss of energy and momentum due to radiation is a necessary consequence of any interaction involving charged particles. This implies, of course, that Newton's second law of motion is no longer valid. The additional terms in (22.5) express the generation and emission of radiation as a necessary by-product of the acceleration of a charge. The mathematical complications due to going beyond Newton's second law are considerable.

(b) In any approximation in which the dynamical description includes radiation emission, the symmetry of time inversion invariance is lost. Therefore, the complete set of equations, Maxwell's equations together with the equation of motion are not time inversion invariant because the latter is not. Only when radiation emission is ignored does the system revert to time reversal invariance.

Various generalizations are of interest. First, there are the equations for systems of n charges each with its own proper time τ_k $(k = 1, \cdots, n)$ including the interactions between the charges as well as the emission of radiation from each

of them. Such problems are likely to be solved best by computer. Then, there are equations for charges with spin and magnetic moment. These properties are less difficult to deal with. My review article [8] and my book [9] introduce these equations.

Consider next the more difficult problem of equations of motion for extended charges that cannot be described by pseudo-point charges. Their finite extension prohibits a differential equation of motion. An integro-differential equation becomes necessary because one must integrate over the finite size charge distribution. That problem was treated by Medina [6] and I shall not pursue it here. But it is an important consistency check that the equation for an extended charge reduces (in the pseudo-point limit) to the LL equation.

22.4 The Relation to Quantum Electrodynamics

Classical electrodynamics (CED) is of course a theory that is meant to be valid as the classical limit of quantum electrodynamics (QED). Mathematically, this means that the above equations for CED should be the classical limit of a set of fundamental equations of QED. Long ago, this problem was attacked by Moniz and Sharp [10] for the nonrelativistic case. Later, Krivitskii and Tsytovich [11] solved it by starting with the nonrelativistic quantum mechanical Hamiltonian for a charge in an electromagnetic field and in the classical limit deduced the Lorentz force equation with an additional τ_0 term. In this approximation, one obtains to lowest order the nonrelativistic form of the LL equation (22.3), $m\,\ddot{\vec{x}} + \vec{F} = \tau_0\,\dot{\vec{F}}$, where F becomes the Lorentz force,

$$m\,\ddot{\vec{x}} + \vec{F} = \tau_0 \left(e\vec{E} + \frac{e^2}{m}\,\vec{E} \times \vec{B} \right) \tag{22.6}$$

The LL equation is thus confirmed as the classical limit of QED at least in the nonrelativistic limit.

The basic problem of the equations of motion of a classical charge can now be considered solved. It took about a century since the original work by Abraham and Lorentz.

References

[1] M. Abraham, Zeitschrift der Physik 14, 236 (1904). This is the first publication of the ALD equation. It is relativistically correct even though it

preceded Einstein's relativity papers of 1905. Lorentz's work before that contains only the non-relativistic limit. Much later, a covariant derivation of the ALD equation was published by P.A.M. Dirac, Proc. Roy. Soc. (London) A 167, 148 (1938).

[2] L. Landau and E. Lifshitz, The Classical Theory of Fields (Addison-Wesley, Reading, MA 1951.)

[3] T. Erber, Fortschritte der Physik 9, 343 (1961).

[4] H. Spohn, Europhysics Letters 50, 287 (2000).

[5] F. Rohrlich, Phys. Rev. E 77, 046609 (2008).

[6] R. Medina, Journal of Physics A 39, 182 (1993).

[7] J. Schwinger, L. L. DeRaad, K. A. Milton, and W. Tsai, Classical Electrodynamics (Perseus Press, NY 1998), Chapter 37.

[8] F. Rohrlich, The Electron: Development of the First Elementary Particle Theory, in collection of papers The Physicist's Conception of Nature, edited by J. Mehra, (D. Reidel Publishing Company, Dordrecht-Holland, 1973), pp. 331 - 379.

[9] F. Rohrlich, Classical Charged Particles, 3rd Edition (World Scientific, New Jersey, 2007), Section 7-5.

[10] E.J. Moniz and D.H. Sharp, Phys. Rev. D 15, 2850 (1977).

[11] V. S. Krivitskii and V. N. Tsytovich, Soviet Physics Uspeki 34, 250-258 (1991).

23

The Mystery of Parity: In Honor of Tom Erber's 80th birthday

Jonathan L. Rosner[1]
Enrico Fermi Institute and Department of Physics,
University of Chicago[2]

EFI 09-36, arXiv:0912.1053, December 2009

"And should I not take pity on Nineveh, that great city, with more than a hundred and twenty thousand inhabitants who do not know their right hand from their left, and many beasts besides?"[3]

23.1 Introduction

Our world does not exhibit left-right symmetry at the level of familiar objects: biological [1] and polar [2] molecules, organic chemicals [3], human anatomy, and much more. However, before 1956 it was widely [4] (though not universally [5]) assumed that the fundamental laws of physics exhibited that symmetry (*parity invariance*). When it was called into question for the weak interactions [6], experiments [7, 8, 9] quickly showed that in fact the weak interactions had a definite handedness, involving left-handed particles and right-handed antiparticles.

[1] rosner@hep.uchicago.edu
[2] 5640 S. Ellis Avenue, Chicago, IL 60637
[3] Jonah 4:11

Could parity violation at the microscopic level be responsible for what we see in the macroscopic world? Despite calculations claiming this to be so (see, e.g., [10]), Tom Erber has pointed out that this asymmetry need not stem from the microscopic level, but can arise spontaneously in very simple systems. For N point charges arranged on the surface of a unit sphere, the lowest-energy state is mirror-symmetric for $2 \leq N \leq 10$, but asymmetric for $N = 11$ [2, 11] and specific higher values of N. In like manner (see also [12]), although the Coriolis force tends to send water down a drain counterclockwise in the Northern Hemisphere and clockwise in the Southern, initial conditions play a far more important role.

In this article we shall be concerned with *microscopic* parity invariance and a mystery which it presents. This consists of a marriage of internal and space-time symmetries, forbidden when the space-time symmetry consists of the whole Poincaré group [13] but permitted in this case because of the discrete nature of the parity transformation. We will argue that this marriage could point to regularities underlying the nature of quarks and leptons, and to extensions of particle interactions beyond those known today.

In Section 2 we briefly review the observed pattern of quarks and leptons, noting the great difference between the masses of the light neutrinos and the remaining fermions. In Section 3 we express this difference in group-theoretic terms, relying on an oft-employed five-dimensional geometric construction based on the group SO(10). The role of parity reversal in this language is extremely simple, consisting of reflection of one of the five coordinates. Possible consequences of this observation are given in Section 4, while Section 5 concludes.

23.2 Quark and Lepton Patterns

The observed quarks and leptons fall into three families. One distinguishes left-handed from right-handed states. Each left-handed family consists of a quark electroweak doublet [transforming as a triplet of color SU(3)], a lepton doublet [transforming as a singlet of color SU(3)], and the corresponding antiparticles which are all electroweak singlets. In each right-handed family the roles of the particles and antiparticles are reversed. For Dirac particles (the quarks and charged leptons) the left-handed and right-handed states and the corresponding antiparticles are combined into one four-component object with a specific "Dirac" mass. The possibility that a neutrino can be its own antiparticle allows for left-handed neutrinos and right-handed antineutrinos (the "active" participants in weak interactions) to have one "Majorana" mass while the "sterile"

left-handed antineutrinos and right-handed neutrinos have another. In Fig. 1 we show present information on quark and lepton masses, quoting Dirac masses for the charged fermions and direct upper limits on masses of "active" neutrinos which may or may not be of Majorana type.

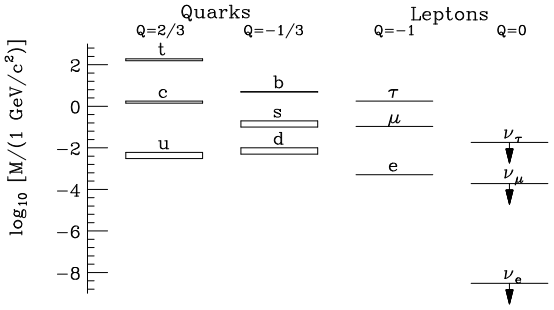


Figure 23.1: Masses of observed quarks and leptons on a logarithmic scale. The upper limits on neutrino masses are based on direct searches [14]; see text for much more stringent limits.

Neutrinos are known to mix with one another, so that the states of definite mass (denoted ν_1, ν_2, and ν_3) are linear combinations of the "flavor" eigenstates ν_e, ν_μ, and ν_τ. Neutrino oscillation experiments find $\Delta m^2_{21} \equiv m^2_2 - m^2_1 = (7.59^{+0.19}_{-0.21}) \times 10^{-5}$ eV^2, $\Delta m^2_{32} \equiv m^2_3 - m^2_2 = (2.43 \pm 0.13) \times 10^{-3}$ eV^2 [14]. If $m^2_1 \ll m^2_{(2,3)}$, then $m_1 \ll m_2 \simeq 9$ meV, $m_3 \simeq 50$ meV. However, all the neutrino masses could be larger and quasi-degenerate. In any case a cosmological bound implies that the sum of the (active) neutrino masses must not exceed 0.28 eV [15], far below the direct limits depicted in Fig. 1.

Fig. 1 represents one of the great puzzles of today's particle physics. Do the masses of the quarks and leptons (and their mixing under the weak interactions) represent some deep underlying structure (as in the Periodic Table of the Elements), or the solution of some anarchic dynamics (as in the Titius-Bode law describing planetary orbits)? For the present we bypass this question and discuss the structure of a single family, which we shall denote

$$F = \begin{pmatrix} u \\ d \\ \nu_e \\ e^- \end{pmatrix} . \tag{23.1}$$

23.3 Geometry of Grand Unified Groups

The strong interactions are described by an $SU(3)_C$ (C for color) Yang-Mills gauge theory, while the gauge symmetry of the electroweak interactions is $SU(2)_L \otimes U(1)_{Y_W}$, where the subscript L indicates that the interaction applies to left-handed fermions (and right-handed antifermions), while the subscript Y_W denotes weak hypercharge. Georgi and Glashow [16] found an ingenious way to unify $SU(3) \otimes SU(2) \otimes U(1)$ into an SU(5) group; the 15 observed left-handed quarks and leptons (excluding left-handed antineutrinos) of each family are apportioned into 5-dimensional and 10^*-dimensional representations of SU(5). However, the pattern becomes much simpler when SU(5) is included into the group SO(10) [17, 18]. The 5- and 10^*- dimensional representations of SU(5) combine with an SU(5) singlet, the right-handed neutrino, into a single 16^*-dimensional spinor representation of SO(10), a group of rank 5 whose representation members may be identified by their coordinates in a 5-dimensional vector space. The spinor consists of next-to-nearest neighbors on the vertices of a 5-dimensional hypercube in this space. Its members may be identified by vectors of the form

$$\left(\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}\right) \ , \tag{23.2}$$

with an odd number of + signs for 16^* and an even number for its conjugate 16 representation [19]. Other representations of SO(10) have simple depictions in this language: For example, members of the vector 10-plet of SO(10) are denoted by

$$(\pm 1, 0, 0, 0, 0) + \text{permutations} \ . \tag{23.3}$$

The group SO(10) has rank 5, so there are five mutually commuting observables which may be represented in it. As the color SU(3) subgroup of SO(10) has rank two, one may take color isospin I_{3C} and hypercharge Y_C as two of the observables. For $SU(2)_L$ one takes its third component I_{3L}, while weak hypercharge will be denoted by Y_W. The electromagnetic charge is $Q = I_{3L} + Y_W/2$. A fifth observable Q_χ, lying in SO(10) but outside SU(5), will be defined shortly.

One may now measure the value of any observable for an SO(10) representation member by taking its projection along a specific five-dimensional vector,

e.g.:

$$V(I_{3C}) = \left(+\frac{1}{2}, -\frac{1}{2}, 0, 0, 0\right) \quad ; \quad V(Y_C) = \left(+\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}, 0, 0\right) \ ,$$
$$V(I_{3L}) = \left(0, 0, 0, +\frac{1}{2}, -\frac{1}{2}\right) \quad ; \quad V(Y_W) = \left(-\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3}, 1, 1\right) \ ;$$
$$V(Q) = \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 1, 0\right) \quad .$$

An additional observable is represented by $V(Q_\chi) = (1, 1, 1, 1, 1)/\sqrt{10}$, where we have chosen to normalize Q_χ in the same way as I_{3C} or I_{3L}. In the 16-plet of SO(10), 5-plet SU(5) members have $Q_\chi = -3/\sqrt{40}$, 10^*-plet members have $Q_\chi = 1/\sqrt{40}$, and the SU(5) singlet has $Q_\chi = 5/\sqrt{40}$.

The specific members of the left-handed family (23.1) may be denoted by the following spinors. A subscript $1, 2, 3$ denotes the color label; we display only one color of each quark. We shall adopt the shorthand $\pm$ for coordinates $\pm 1/2$ [20]. We shall also put a vertical bar between the first three indices, denoting color SU(3), and the last two, denoting weak SU(2). We then have

$$u_{L1} = (+--|+-) \quad ; \quad d_{L1} = (+--|-+) \ ;$$
$$\nu_L = (+++|+-) \quad ; \quad e_L^- = (+++|-+) \ . \tag{23.4}$$

Each of these is a weak doublet with $I_{3L} = \pm 1/2$, as the last two indices are unequal. The corresponding antiparticles, obtained by reversing the signs of the first four indices, are

$$\bar{u}_{L1} = (-++|--) \quad ; \quad \bar{d}_{L1} = (-++|++) \ ;$$
$$\bar{\nu}_L = (---|--) \quad ; \quad e_L^+ = (---|++) \ . \tag{23.5}$$

The $\bar{\nu}_L$ has no charges within the Standard $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_{Y_W}$ Model; it is *sterile*.

An interesting three-dimensional projection of the five-dimensional space may be obtained by defining the horizontal plane to be the two-dimensional vector space describing color SU(3) and the vertical axis to be electric charge. The members of an SO(10) 16-plet may then be represented as two cubes stacked corner-to-corner, as shown in Fig. 2.

So far we have not discussed the right-handed states. These, it turns out, are related to the corresponding left-handed states by a simple reversal of the

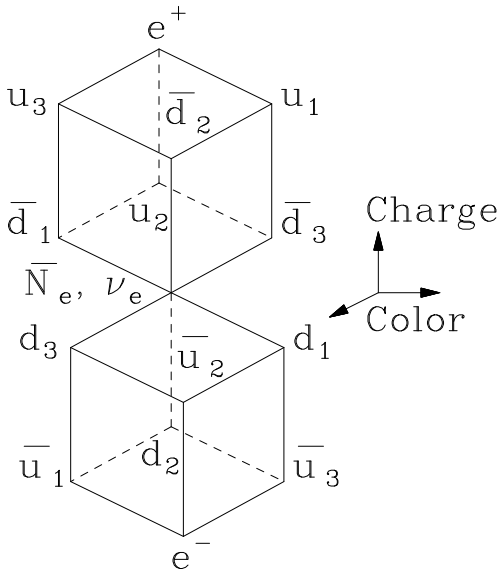


Figure 23.2: Projection of SO(10) 16-plet describing a quark-lepton family into the space of color (horizontal plane) $\otimes$ electric charge (vertical axis)

fifth index. Thus, for right-handed particles we have

$$\begin{aligned} u_{R1} &= (+--|++) \quad ; \quad d_{R1} = (+--|--) \; ; \\ \nu_R &= (+++|++) \quad ; \quad e_R^- = (+++|--) \; , \end{aligned} \tag{23.6}$$

while for right-handed antiparticles we have

$$\begin{aligned} \bar{u}_{R1} &= (-++|-+) \quad ; \quad \bar{d}_{R1} = (-++|+-) \; ; \\ \bar{\nu}_R &= (---|-+) \quad ; \quad e_R^+ = (---|+-) \; . \end{aligned} \tag{23.7}$$

It is now the right-handed *particles* which are electroweak singlets, while the right-handed *antiparticles* are electroweak doublets. In particular, the right-handed neutrino ν_R is sterile with respect to Standard Model charges.

All this is familiar to practitioners of grand unified theories. Indeed, the unbroken SO(10) symmetry is left-right symmetric [18, 21]; it is a non-zero expectation value of the charge Q_χ which destroys this symmetry. This could

arise at any mass scale from a Higgs mechanism. If the scale is several TeV or less, one might be able to observe the corresponding neutral gauge boson (a "Z_χ" [22, 23]) at the CERN Large Hadron Collider (LHC). A very large mass scale, however, could be associated with a large Majorana mass of right-handed neutrinos.

One can also envision the breaking of SO(10) as proceeding first through its subgroup SO(6) $\otimes$ SO(4) (easily illustrated on the fingers of two hands). The SO(6) is isomorphic to an SU(4) group which may be thought of as extended color, regarding leptons as the fourth color [18]. Its subgroup containing color is $\mathrm{SU}(3)_C \otimes \mathrm{U}(1)_{B-L}$, where B and L are baryon and lepton number. The SO(4) is isomorphic to $\mathrm{SU}(2)_L \otimes \mathrm{SU}(2)_R$. The subsequent breaking of $\mathrm{SU}(2)_R$ would be responsible for parity-noninvariance of the electroweak theory. A handy expression for electric charge, instead of the uninspiring relation involving weak hypercharge, is $Q = I_{3L} + I_{3R} + (B-L)/2$. The vectors projecting out I_{3R} and $B-L$ are

$$V(I_{3R}) = \left(0, 0, 0, +\frac{1}{2}, \frac{1}{2}\right) \ ; \ \ V(B-L) = \left(-\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3}, 0, 0\right) \ . \qquad (23.8)$$

What is puzzling about the parity operation is that, although it is a transformation of the Poincaré group, it corresponds to a simple operation in the SO(10) five-dimensional vector space. Because it is a discrete transformation, it evades the Coleman-Mandula theorem [13] which forbids the combination of internal and Poincaré symmetries except as a direct product. Its violation is also deeply implicated in how the Standard Model arises from some higher symmetry. In the next section we argue that such a symmetry is likely to exist on the basis of our very incomplete knowledge about the nature of matter in the Universe.

23.4 Expanded Symmetries

Ordinary matter makes up a small fraction of the known energy density of the Universe; dark matter comprises about five times as much [24]. We have little clue as to its nature.

Imagine a TeV-scale effective symmetry SU(3) $\otimes$ SU(2) $\otimes$ U(1) $\otimes$ G, where the beyond-Standard-Model (BSM) group G could be any number of extensions currently on the market. One can classify the new types of matter very generally as shown in Table 23.1 [25]:

Grand unified theories well beyond SO(10) have been proposed. The E_8' of $\mathrm{E}_8 \otimes \mathrm{E}_8'$ in the heterotic string [26] could play a role of "shadow matter" which

Table 23.1: Possible types of matter classified according to SM and BSM (G) transformation.

Type of matter	Std. Model	G	Example(s)
Ordinary	Non-singlet	Singlet	Quarks, leptons
Mixed	Non-singlet	Non-singlet	Superpartners
Shadow	Singlet	Non-singlet	E_8' of $E_8 \otimes E_8'$

communicates only weakly with our world. The fifth coordinate in the SO(10) description, whose reversal we have shown induces parity reflections, could play a wider role in an extended vector space of more than five dimensions.

The spinors of SO(2N) groups may be represented as alternate vertices of hypercubes in N dimensions. These spinors and their conjugates each have 2^{N-1} members. If $N > 5$, one can ask what fraction of those have the form $(+++++|a_{N-5} \dots a_N)$ or $(-----|a_{N-5} \dots a_N)$, where the first five indices refer to the SO(10) subgroup of SO(2N), and hence would be "sterile" under charges of the Standard Model. The answer is the same as for $N = 5$, namely 1/16. One is seeking, rather, a scheme where *most* of the matter is sterile under Standard Model charges. The existence of a large amount of dark matter in our Universe could be a key to guessing the structure of a large Grand Unified Group, and perhaps incidentally helping to solve the mystery of parity violation.

23.5 Conclusion

The advent of the CERN Large Hadron Collider will offer one possible window into extended grand unified theories, through the discovery of new forms of matter or new gauge bosons. One of the simplest such examples would be the gauge boson Z_χ coupled to the charge Q_χ mentioned above [22, 23]. It may turn out in retrospect that the role of parity and its violation in our current understanding of unified theories was just a foretaste of a much richer structure.

Acknowledgments

I thank G. Senjanovic for a helpful communication and Z. Silagadze for alerting me to several interesting interesting papers on mirror matter [27]. This work

was supported in part by the United States Department of Energy through Grant No. DE FG02 90ER40560.

References

[1] L. Pasteur, "Sur les relations qui peuvent exister entre la forme cristalline, la composition chimique, et le sens de la polarisation rotatoire," Ann. Chim. Phys. **24**, 442 (1848).

[2] T. Erber, "Reflections on Parity," Foundations of Physics 34, 1515 (2004).

[3] Roald Hoffmann, The Same and Not The Same, New York, Columbia University Press, 1995.

[4] A. Pais, Inward Bound, Clarendon Press, Oxford, 1986.

[5] P. A. M. Dirac never saw any reason for incorporating parity invariance in a field theory. See G. Farmelo, The Strangest Man: the Hidden Life of Paul Dirac, Basic Books, New York, 2009.

[6] T. D. Lee and C. N. Yang, "Question of Parity Conservation in Weak Interactions," Phys. Rev. 104, 254 (1956).

[7] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental Test Of Parity Conservation In Beta Decay," Phys. Rev. 105, 1413 (1957).

[8] R. L. Garwin, L. M. Lederman, and M. Weinrich, "Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon," Phys. Rev. 105, 1415 (1957).

[9] J. I. Friedman and V. L. Telegdi, "Nuclear Emulsion Evidence For Parity Nonconservation In The Decay Chain $\pi^+ \to \mu^+ \to e^+$," Phys. Rev. 105, 1681 (1957).

[10] Y. Yamagata, "A hypothesis for the asymmetric appearance of biomolecules on earth," J. Theor. Biol. 11, 492 (1966); D. K. Kondepudi and G. W. Nelson, "Weak Neutral Currents And The Origin Of Biomolecular Chirality," Nature 314, 438 (1985); D. K. Kondepudi, "Selection of Handedness in Prebiotic Chemical Processes," in Discovery of Weak Neutral Currents: The Weak Interaction Before and After," AIP Conference Proceeding 300, edited by A. K. Mann and D. B. Cline, New York, AIP, 1994, p. 491.

[11] L. Coffey, "The Surface Coulomb Problem: Energy Minima and Hausdorff Metrics," this volume.

[12] M. Gleiser and S. I. Walker, "The Chirality of Life: From Phase Transitions to Astrobiology," arXiv:0811.1291 [astro-ph].

[13] S. Coleman and J. Mandula, "All Possible Symmetries Of The S Matrix," Phys. Rev. 159, 1251 (1967).

[14] C. Amsler *et al.* (Particle Data Group), "Review of Particle Physics," Phys. Lett. B 667, 1 (2008).

[15] F. De Bernardis, P. Serra, A. Cooray, and A. Melchiorri, "An improved limit on the neutrino mass with CMB and redshift-dependent halo bias-mass relations from SDSS, DEEP2, and Lyman-Break Galaxies," Phys. Rev. D 78, 083535 (2008).

[16] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces," Phys. Rev. Lett. **32**, 438 (1974).

[17] H. Georgi, "The State of the Art - Gauge Theories," in Proceedings of the Division of Particles and Fields, American Physical Society, Williamsburg, VA, 1974, AIP Conf. Proc. 23, 575 (1975), edited by C. E. Carlson; H. Fritzsch and P. Minkowski, "Unified Interactions of Leptons and Hadrons," Ann. Phys. 93, 193 (1975).

[18] J. C. Pati and A. Salam, Phys. Rev. D 10, 275 (1974); Erratum ibid. D 11, 703 (1975).

[19] R. Slansky, "Group Theory for Unified Model Building," Phys. Rep. 79, 1 (1981).

[20] F. Wilczek and A. Zee, "Families from Spinors," Phys. Rev. D 25, 553 (1982).

[21] R. N. Mohapatra and J. C. Pati, "Left-Right Gauge Symmetry and an Isoconjugate Model of CP Violation," Phys. Rev. D 11, 566 (1975); R. N. Mohapatra, J. C. Pati, and A. Salam, "A Natural Left-Right Symmetry," Phys. Rev. D 11, 2558 (1975). For a more general discussion see G. Senjanovic and R. N. Mohapatra, "Exact left-right symmetry and spontaneous violation of parity," Phys. Rev. D 12, 1502 (1975).

[22] P. Langacker, R. W. Robinett, and J. L. Rosner, "New Heavy Gauge Bosons in pp and $p\bar{p}$ Collisions," Phys. Rev. D 30, 1470 (1984).

[23] P. Langacker, "The Physics of Heavy Z' Gauge Bosons," Rev. Mod. Phys. 81, 1199 (2009); "Z' Physics at the LHC," arXiv:0911.4294 [hep-ph].

[24] E. Komatsu *et al.* [WMAP Collaboration], "Five-Year Wilkinson Microwave Anisotropy Probe Observations:Cosmological Interpretation," Astrophys. J. Suppl. **180** (2009) 330.

[25] J. L. Rosner, "Dark matter in many forms," arXiv:astro-ph/0509196, presented at 2005 ALCPG & ILC Workshops, Snowmass, Colorado, August 14–27, 2005, published in the Proceedings [ECONF **C0508141**, ALCPG0106, (2005)].

[26] D. J. Gross, J. A. Harvey, E. J. Martinec, and R. Rohm, "Heterotic String Theory. 1. The Free Heterotic String," Nucl. Phys. B256, 253 (1985); "Heterotic String Theory. 2. The Interacting Heterotic String," Nucl. Phys. B267, 75 (1986).

[27] Z. K. Silagadze, "TeV scale gravity, mirror universe, and ... dinosaurs," Acta Phys. Polon. B32, 99 (2001); "Mirror objects in the solar system?," *ibid.* B33, 1325 (2002); "Mirror dark matter discovered?," arXiv:0808.2595; see also the review by L. B. Okun, "Mirror particles and mirror matter: 50 years of speculation and search," Phys. Usp. 50, 380-389,2007, arXiv:hep-ph/0606202.

24

Eigenschrift: The End of the Classical Theories of Radiation Reaction

T. Erber
Department of Physics and
Department of Applied Mathematics
Illinois Institute of Technology

24.1 Introduction

Is it proper for a Festschrift to include an Eigenschrift? There certainly are lofty precedents for including a set of "... remarks concerning the essays brought together in this cooperative volume ..." by the 'Festkind' (see, for example [1]). This format is essentially the literary equivalent of a receiving line in which polite thanks and acknowledgements are murmured — occasionally offset by the sharper tone of polemics or even veiled recriminations. Here the intent is far more mild, modest, and restricted in scope. Among several research areas, electrodynamics has been a theme of my interests for more than fifty years, and one can ask — leaving aside the bolus of publication lists — what new things have really been discovered? In particular, from the outset [2] it seemed plausible that an improved understanding of the physics of radiation reaction could be obtained by pushing to the extremes of energy and radiation rates where the palliatives of perturbation expansions or self-consistency restrictions no longer apply. Following the old adage "To see something new one has to build something new", a survey of possibilities showed that synchrotron radiation, i.e.,

magnetic bremsstrahlung emitted by high energy ($\geq$ GeV) electrons traversing megagauss fields would broach the limits of conventional experimental parameters [3]. Indeed, megagauss fields (in vacuum) have energy densities comparable to high explosives { $\mathcal{E}$ (kJ/cm^3) $\sim$ 4 H^2 (MG) }, and can exert pressures on ambient conductors sufficient to blow them apart {P (kbar) $\sim$ 40 H^2 (mG) }. [4, 5]. The means for generating such fields originated in the secret environment of weapons laboratories in the 1940s [6], and despite declassification in 1960 [7], still have a recondite character. Experimental work in this new area was initiated at IIT in 1964, and by 1966 a functioning flux compression facility had been established [8]. The technical design for a magnetic bremsstrahlung experiment was completed by 1967 [9], and detailed research proposals were submitted to various government agencies and the Stanford Linear Accelerator Center (SLAC) in 1968. The project was approved in 1969, and the experimental work carried out successfully at SLAC during the period July - November of 1970 [10].

The complete analyses of the experimental results took nearly two years since information concerning the electron beam deflections and the corresponding bremsstrahlung (synchrotron) spectral intensity had to be laboriously extracted from microscope scans of nuclear emulsions [11, 12]. In contrast to the experimental novelties, the theoretical expectations concerning the results tended to follow conventional paths. It was anticipated that both the electron trajectories as well as the radiation spectra could be accurately described by classical electrodynamics, with small corrections — marginally detectable under SLAC conditions – derived from radiation reaction theory. It was also generally presumed that quantum modifications of synchrotron radiation would be negligible under forseeable terrestrial conditions [13]. However, the experimental results prompted a drastic shift in perspective. The electron trajectories, which were plainly visible in the nuclear emulsions, had broadened from a tightly focused beam — 1.3 mm in diameter contained within a cone of full angle 3×10^{-4} radians — into a large spray that extended over nearly the entire length of the emulsions ($>$ 60 mm). An effect of this kind had actually been predicted by Sokolov and Ternov in connection with the quantum excitation of betatron oscillations [14, 15]; it is expected to occur when the emitted (synchrotron) photon momentum exceeds the momentum spread of the 'most classical' electron orbits in the magnetic field [16, 17]. Quantitatively, the threshold criterion for these dispersions is

$$\hbar/\lambda \sim \Delta p \cong m\,c\,\sqrt{H/H_{cr}} \tag{1.1a}$$

or, equivalently,

$$\left[E/(mc^2)\right]^4 \cdot [H/H_{cr}] \geq 1 \tag{1.1b}$$

and, in practical units

$$E^4\,(\mathrm{GeV}) \cdot \mathrm{H}\,(\mathrm{MG}) \geq 3 \times 10^{-6} \tag{1.1c}$$

where E denotes the total electron energy, $m\,c^2$ the rest energy, and

$$H_{cr} = \frac{m^2\,c^3}{e\,\hbar} \cong 4.414 \times 10^{13}\ \mathrm{G} \tag{1.1d}$$

is the characteristic quantum mechanical measure of magnetic field strength. The parameters of the SLAC experiments $\{\ E \sim 19$ GeV, $H \sim 1.5$ MG $\}$ imply that the threshold for quantum induced electron dispersion is surpassed by many orders of magnitude.

According to classical electrodynamics the total magnetic bremsstrahlung energy radiated by an electron traversing a path length Δ normal to a magnetic field H is given by [18]

$$I^{cl} = \frac{2\alpha}{3}\,\frac{mc^2}{\lambda_c}\,\Delta\Upsilon^2 \tag{1.2a}$$

where

$$\Upsilon = \frac{E}{m\,c^2}\,\frac{H}{H_{cr}} \cong 4.43 \times 10^{-5} \cdot E\ (\mathrm{GeV}) \cdot \mathrm{H}\ (\mathrm{MG}) \tag{1.2b}$$

and $\alpha^{-1} =$ "137", $\lambda_c = \hbar/(m\,c)$. Since this is a classical result it is easy to check that all factors of $\hbar$ cancel. In practical units.

$$I^{cl}\,(\mathrm{MeV}) = 1.263 \times 10^{-2} \cdot \Delta\,(\mathrm{mm}) \cdot [\mathrm{H}\,(\mathrm{MG}) \cdot \mathrm{E}\,(\mathrm{GeV})]^2 \tag{1.2c}$$

Clearly, for the SLAC parameters (with $\Delta \sim 5$ mm) the radiated energy per electron is large, i.e., $I^{cl} \sim 50$ MeV, whereas the dimensionless Υ parameter remains small, $\Upsilon \sim 0.0012$. Physically, Υ is a measure of the ratio of the momenta of the radiated photons to the electron momentum; as long as it remains small quantum effects in the synchrotron spectrum are expected to be negligible. Indeed, the quantum modifications of the synchrotron energy radiation can be expressed as a power series in Υ, viz. [19, 20, 14].

$$I^{QM} = I^{cl}\left(1 - 5.953\,\Upsilon + 48\Upsilon^2\right) \tag{1.3}$$

Comparing the two quantum effect criteria (1.1c) and (1.3), it is clear that there is no contradiction because two different aspects of magnetic bremsstrahlung

are involved; electron recoil and spectral characteristics. However, the stricture that the corrections in (1.3) would "... never be approached in practice ..." has been overtaken by technical developments. According to (1.2b), for energies of 1 TeV and fields of 2 MG, Υ is of the order of 0.09, and therefore (1.3) indicates a significant decrease in radiation rates, $I^{QM} \sim 0.85\, I^{cl}$. In fact, even under these 'modest' conditions, the step-by-step inclusion of higher order quantum corrections fails because the expansion (1.3) becomes numerically useless just at the point where the quantum effects in the synchrotron spectrum become noticeable. For instance, as shown by exact calculation, the expected decrease in radiation is actually $I^{QM} \sim 0.68\, I^{cl}$ [21, 22].

A basic premise of the classical theory of radiation reaction is that the charge dynamics can be described by a modification of Newton's second law

$$\frac{d\vec{p}}{dt} = \vec{F}_{ext} + \vec{F}_{RR} \tag{1.4}$$

where $\vec{F}_{RR}$ incorporates the effects of radiation reaction [2]. In the case of magnetic bremsstrahlung $\vec{F}_{ext}$ corresponds to the Lorentz force, and simple momentum transfer considerations show that [3, 23]

$$\frac{|\vec{F}_{RR}|}{|\vec{F}_{L}|} \cong \frac{2\alpha}{3}\,\frac{H}{H_{cr}}\left(\frac{E}{m\,c^2}\right)^2 \cong 4.16\times 10^{-4}\cdot E^2\,(\mathrm{GeV})\cdot \mathrm{H}\,(\mathrm{MG}) \tag{1.5}$$

For $E \sim 19$ GeV and $H \sim 1.5$ MG, this ratio is 0.23, and leads to the expectation that radiation reaction may be at the threshold of detectability in the SLAC experiments. But comparison with (1.1c) shows that any effects of this kind would be completely dominated by quantum fluctuations. The observed spatial dispersion of the scattered electrons is incompatible with the idealizations underlying (1.4). In contrast to the classical picture of deterministic charge trajectories associated with a continuous emission of radiation, the experimental situation corresponds to a stochastic smearing of the electron dynamics induced by 'betatron oscillations' and discrete (hard) photon emission. So, in analogy with the quantum corrections in (1.3), just at the point where classical radiation reaction could become significant, it actually becomes irrelevant.

Fritz Rohrlich has also been interested in radiation reaction for many years [24]. His efforts have mainly been concerned with cleansing the theory of its many blemishes — pre-acceleration, runaway solutions, and similar pathologies. As he explains in another article in this volume (Chapter 22) these efforts have finally succeeded, and led to a self-consistent theory that yields unambiguous and physically sensible solutions. In this respect both of us have taken different

routes to arrive at a point where — to paraphrase Schrödinger — "the curiosity can rest", and the classical theory of radiation reaction may be regarded as complete.

24.2 Classical Electrodynamics and Radiation Reaction

The structure of classical electrodynamics may be summarized in terms of a simple block diagram:

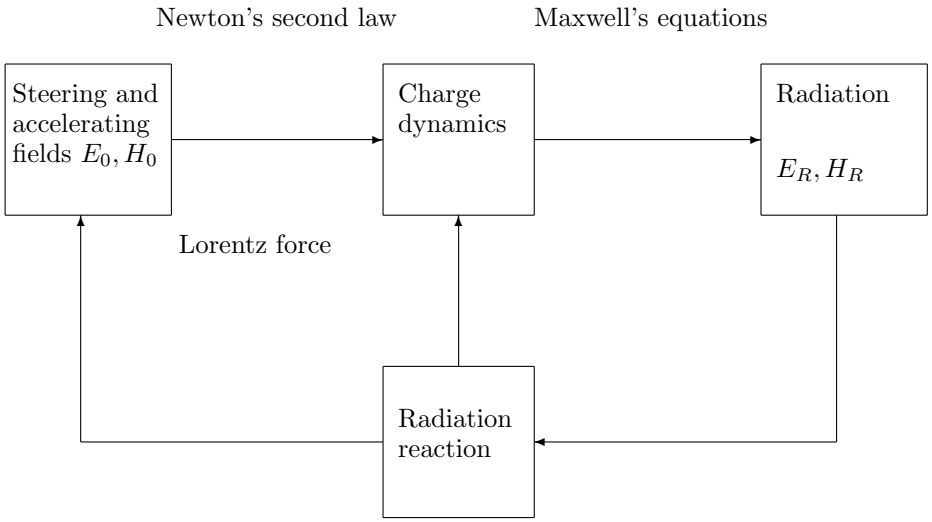


Figure 1: Classical Electrodynamics

The radiation reaction 'feedback loop' is an adjunct to the Maxwell-Lorentz theory intended to account for energy-momentum conservation. But even with this addition deep problems remain "Closed systems in which both the sources and the fields are driven through each other's motion and not by external mechanisms are most difficult to deal with in an exact manner" [25]. In particular, there are no known non-trivial exact solutions of the combined system in Fig. 1. Consider, for example, the electrodynamic two-body problem when retarded interactions are included. In this case the fields acting on each charge can be derived from the Liénard-Wiechert potentials. But these, in turn, depend

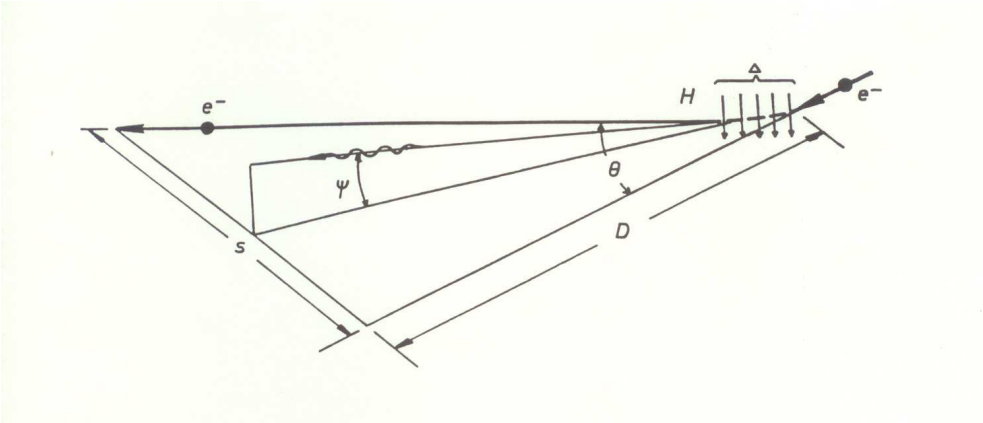


Figure 2: Kinematics of magnetic bremsstrahlung

on delay times which are expressed only implicitly in terms of the unknown charge trajectories. Mathematically the situation is described by a system of functional-differential or functional-integral equations whose solutions remain largely unknown [26, 27, 28].

The radiation reaction feedback loop in Fig 1 formally corresponds to Eq. (1.4) where the general exression for $\vec{F}_{RR}$ valid under relativistic conditions is

$$\vec{F}_{RR} = \frac{2}{3}\frac{e^2}{c^3}\gamma^2\left\{\ddot{\vec{v}} + \left(\frac{\gamma}{c}\right)^2\left(\vec{v}\cdot\ddot{\vec{v}}\right)\vec{v} + 3\left(\frac{\gamma}{c}\right)^2\left(\vec{v}\cdot\dot{\vec{v}}\right)\dot{\vec{v}} + 3\left(\frac{\gamma}{c}\right)^4\left(\vec{v}\cdot\dot{\vec{v}}\right)^2\vec{v}\right\} \tag{2.1}$$

As usual, $\gamma = 1/\sqrt{1-v^2/c^2}$, and therefore for $v \ll c$, $\vec{F}_{RR}$ reduces to the familiar form $\vec{F}_{RR} = (2/3)\,(e^2/c^3)\,\ddot{\vec{v}}$, where the over-dots denote ordinary time derivatives. In the particular case of magnetic bremsstrahlung the principal charge deflection by the magnetic field is determined by the Lorentz force, $\vec{F}_L = e\,(\vec{v}\times\vec{H}\,)$. As indicated on Fig 2 electrons traversing a path length Δ in a magnetic field $\vec{H}$ will be deflected from the forward direction through an angle θ given by

$$\theta = \frac{e\,H\,\Delta}{E}\left[1 + \tfrac{1}{2}\left(\frac{m\,c^2}{E}\right)^2\right] \tag{2.2}$$

The beam deflection s visible on an emulsion at a distance D from the field coil then is

$$s = D\,\frac{m\,c^2}{E}\,\frac{H}{H_{cr}}\,\frac{\Delta}{\lambda_c}\left[1+\tfrac{1}{2}\left(\frac{m\,c^2}{E}\right)^2\right] \tag{2.3a}$$

or, in practical units

$$s\ (\mathrm{mm}) \cong 29.98\cdot \mathrm{D}\ (\mathrm{m})\cdot \mathrm{H}\ (\mathrm{MG})\cdot \Delta\ \mathrm{mm})\,/\mathrm{E}\ (\mathrm{GeV}) \tag{2.3b}$$

Inserting representative parameters for the SLAC experiments yields the estimate

$$s\ (\mathrm{mm}) \cong 29.98\cdot 3.21\ (\mathrm{m})\cdot 1.5\ (\mathrm{MG})\cdot 5\ (\mathrm{mm})\,/19\ (\mathrm{GeV}) \cong 37.8\ (\mathrm{mm}) \tag{2.4}$$

Under these conditions the radiation reaction force (2.1) is responsible for an additional small deflection δs that increases the magnitude of the displacement shown in Fig 2. Specifically,

$$\delta s = \frac{\alpha}{3}\,D\,\left(\frac{H}{H_{cr}}\right)^3\,\left(\frac{\Delta}{\lambda_c}\right)^2 \tag{2.5a}$$

or, in practical units,

$$\delta s\ (\mathrm{mm}) \cong 1.90\times 10^{-4}\cdot \mathrm{D}\ (\mathrm{m})\cdot \mathrm{H}^3\ (\mathrm{MG})\cdot \Delta^2\ (\mathrm{mm}) \tag{2.5b}$$

For the SLAC parameters in (2.4) the corresponding radiation reaction shift is

$$\delta s\ (\mathrm{mm}) \cong 0.051\ (\mathrm{mm}) \tag{2.6}$$

Two methods were used at SLAC to check on the electron beam deflections by the pulsed megagauss fields: (i) X-ray films, sandwiched between intensifying screens, were placed 3.9 m downstream from the megagauss target; and (ii) 600 micron thick Ilford G-5 glass backed nuclear emulsions, oriented vertically with respect to the incident SLAC e^--beam, were positioned in light-tight holders (with Be windows) at a distance of 3.21 m from the coils. A 30 kG DC deflecting magnet, whose field was directed at right angles with respect to the e^- beam path and megagauss field, was located 1 m downstream from the megagauss target in order to sweep away low-energy background electrons. The low-resolution X-ray images yielded deflections of the order of $47-50$ mm [29]. If the parameters in (2.4) are rescaled for the increased distance, $D \simeq 3.9$ m, the result is shifted to 46 mm, in good agreement with the measurements.

Since the Ilford emulsions permit the identification of all individual electron tracks, it was possible to determine the absolute electron flux for several megagauss 'shots', and also to measure the detailed distribution of the electron deflections. Table 1 displays the numerical results for the number of electron tracks located in an area 8 mm × 18 mm centered on the region of maximum electron flux incident in the emulsion [11].

Table 1

s (mm)	26	28	30	32	34	36	38	40	42	44
# electrons	494	542	581	673	879	1051	991	897	647	424

s (mm)	⋯	56	58
# electrons	⋯	442	301

Clearly, the distribution peaks between 36 and 38 mm, which is in excellent agreement with the theoretical value in (2.4). The statistics are however too sparse to detect the small radiative reaction shift in (2.6). In any event, this discrimination is superfluous because the dominant scattering effect of the megagauss field is to disperse the tightly collimated incident e^--beam (1.3 mm in diameter) into a broad band that extends over a large portion of the emulsion. The entries for the two 'outliers" at 56 and 58 mm in Table 1 show that there is a significant high energy electron flux even for these large deflections. As mentioned previously, this conspicuous stochastic behavior is due to quantum recoil fluctuations, cf. (1.1c). Background effects, such as Coulomb bremsstrahlung, can be accounted for by auxiliary measurements and computer simulations [11, 30].

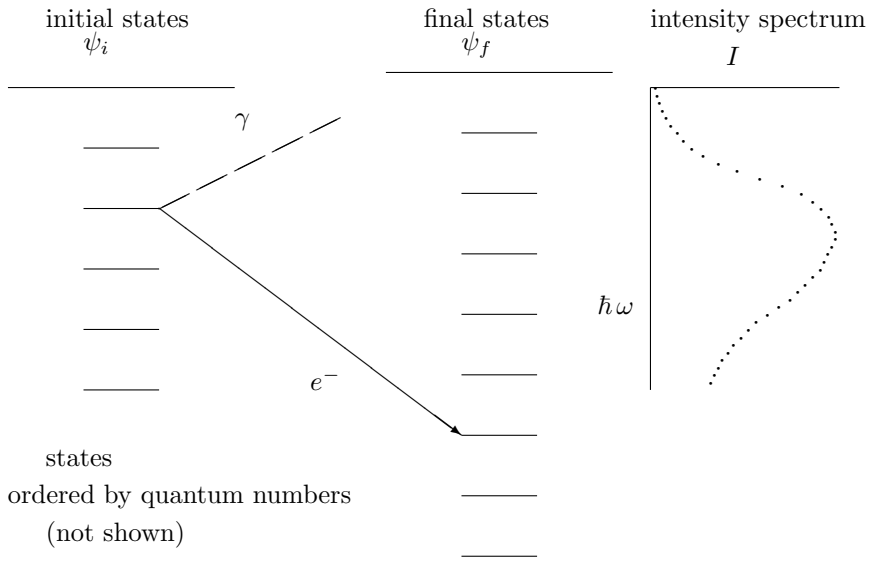
24.3 Quantum Electrodynamics

Suppose that an electron (e^-) in an initial state ψ_i, with energy E_i and momentum $\vec{p}_i$, emits a photon (γ) with energy $\hbar\omega$ and momentum $\vec{k}$, while transitioning to a final state ψ_f, with energy E_f and momentum $\vec{p}_f$. The situation is shown schematically in Fig 3:

In contrast to the classical picture in Fig. 1, energy and momentum conservation are built in from the outset. However, since the conservation laws

$$\begin{aligned} \hbar\omega &= E_i - E_f \\ \vec{k} &= \vec{p}_i - \vec{p}_f \end{aligned} \tag{3.1}$$

merely constrain the values of ω, E_f, etc., the final states are not uniquely determined. There is even more latitude in bound state situations such as synchrotron radiation where arbitrarily large momentum transfers can be absorbed by the magnetic fields: in principle this momentum is ultimately taken up by the (macroscopic) devices that sustain the fields. The stochastic elements enter in the overlap integral $< \psi_f | \exp(-i\vec{k} \cdot \vec{r}) | \psi_i >$ which, as indicated in the Figure, effectively determines the shape of the radiation spectrum.



$$I \sim \left| \int \psi_f^* \, e^{-i\vec{k}\cdot\vec{r}} \, \psi_i \right|^2$$

Figure 3: Schematic for Photon Emission

Adherents of the correspondence principle have long sought for some formal procedure that would yield a radiation reaction equation of the form (1.4) and (2.1) in a classical limit of quantum mechanics. Attempts by Fermi [31], Schwinger [32] and others, e.g. [33] were not successful. As pointed out by Rohrlich "It is deplorable that a derivation as a classical limit from relativistic quantum mechanics still does not seem to exist" [34]. In quantum mechanics the

notion of 'radiation reaction' has broadened out to mean 'radiative corrections', and this generalization includes effects going far beyond energy-momentum balance. For instance, the 'anomalous' magnet moment of the electron — in higher order corrections — includes contributions due to muons, tauons, and couplings to strongly interacting objects, none envisaged in the construction of $\vec{F}_{RR}$. Induced emission, which is the canonical quantum mechanical process of 'radiation reaction', also has no classical counterpart [35].

Suppose again that an electron with energy E traverses a distance Δ in a perpendicular magnetic field H; then the number of photons dN radiated in the energy interval between $\hbar\omega$ and $\hbar\omega + d(\hbar\omega)$ is given by the spectral distribution

$$\frac{dN}{d(\hbar\omega)} = \frac{\sqrt{3}}{2\pi}\,\frac{\alpha}{\hbar\omega}\,\frac{\Delta}{\lambda_c}\,\frac{H}{H_{cr}}\,\kappa\left[\frac{2}{3\Upsilon}\,\frac{\hbar\omega}{E}\right] \tag{3.2a}$$

where κ is the bremsstrahlung function

$$\kappa(z) = z\int_z^\infty dx\, K_{5/3}(z) \simeq \begin{cases} 2.149\, z^{1/3} & z \ll 1 \\ 1.253\, z^{1/2} e^{-z} & z \gg 1 \end{cases} \tag{3.2b}$$

and Υ is defined in (1.2b). In practical units

$$\frac{dN}{d(\hbar\omega)} = \frac{0.12}{\hbar\omega}\cdot H\ (\mathrm{MG})\cdot \Delta\ (\mathrm{mm})\cdot \kappa\left[\frac{15\cdot\hbar\omega\ (\mathrm{MeV})}{\mathrm{H}\ (\mathrm{MG})\cdot \mathrm{E}^2\ (\mathrm{GeV})}\right] \tag{3.2c}$$

The average photon energy is

$$<\hbar\omega> = \frac{4}{5\sqrt{3}}\,E\,\Upsilon \tag{3.3a}$$

or, in practical units

$$<\hbar\omega\ (\mathrm{MeV})> = 2.05\times 10^{-2}\cdot \mathrm{H}\ (\mathrm{MG})\cdot \mathrm{E}^2\ (\mathrm{GeV}) \tag{3.3b}$$

At SLAC, with $H \sim 1.5$ MG and $E \sim 19$ GeV, this average was circa 11 MeV. The corresponding expectation value for the total number of emitted photons N may be obtained by integrating over the spectrum,

$$N = \frac{5}{2\sqrt{3}}\,\alpha\frac{\Delta}{\lambda_c}\,\frac{H}{H_{cr}} \tag{3.4a}$$

or, in practical units

$$N \cong 0.6178\ \cdot \Delta\ (\mathrm{mm})\cdot \mathrm{H}\ (\mathrm{MG}) \tag{3.4b}$$

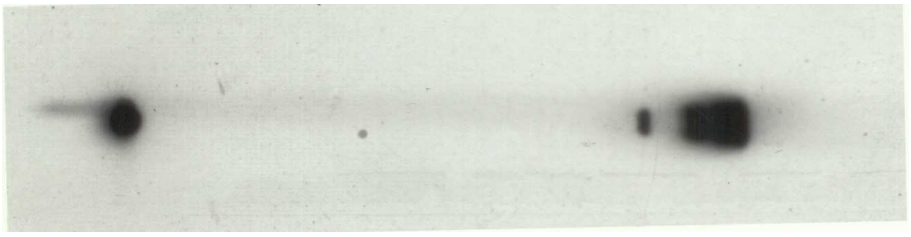


Figure 4(a)

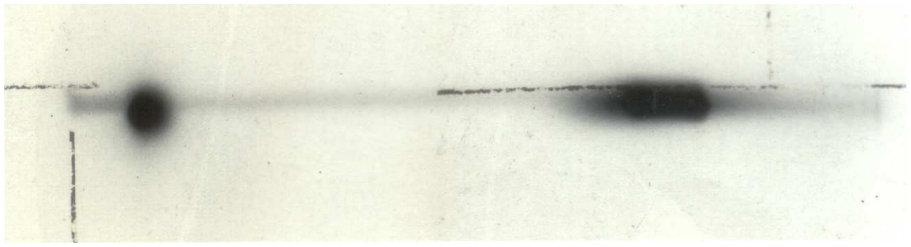


Figure 4(b)

For the SLAC experiments $N \sim 5$, which again emphasizes the importance of quantum fluctuation effects. Combining (3.3b) and (3.4b) implies a total radiative energy loss per electron of 51 MeV in agreement with the earlier estimate (1.2c) et seq.

Figures 4(a) and 4(b) are photographs of two of the Ilford emulsions used to record the electron tracks and bremsstrahlung photons at SLAC. The dark spots on the left are due to the electron tracks made by single accelerator pulses of circa 10^6 e^- passing straight through the apparatus after optical and electronic alignment. This involved some 'sharpshooting' tricks since the 1.3 mm diameter e^--beam had to pass cleanly through the 2 mm diameter holes drilled through the megagauss coils [10]. The dark streaks on the right side of the emulsions mark the pattern of electrons scattered by the pulsed megagauss fields. In first approximation, these deflections are given by (2.3b) and (2.4).

The simplified schematic of Figure 5 (not drawn to scale) connects the geometry of Figure 2 with the layout of the experiment and the features visible on the emulsions.

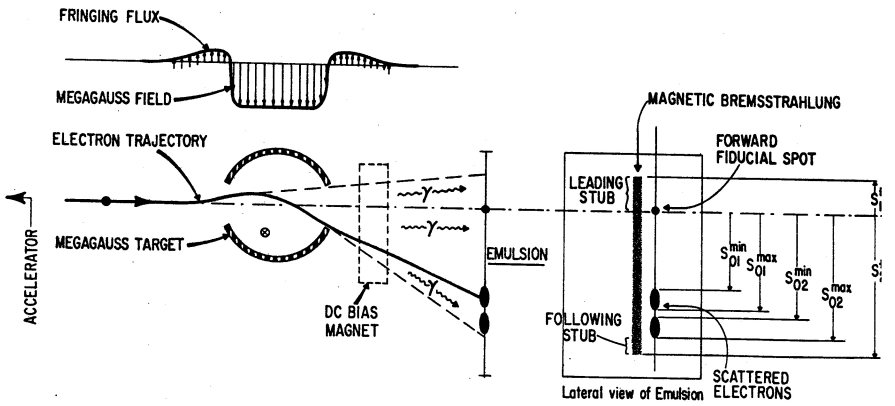


Figure 5

The faint streaks extending between the electron spots on Figs 4(a) and 4(b) are mixed images of high energy ($\sim$ 19 GeV) electron tracks and medium energy ($\leq$ 300 MeV) electron-positron pairs. Given the attenuation coefficients for $\gamma \rightarrow e^+ + e^-$ pair conversion in Ilford G5 emulsion (e.g. 0.011 $(\text{mm})^{-1}$ at 20 MeV), one can estimate that about 0.5 % of the synchrotron photons incident on the emulsions result in $e^+ - e^-$ pairs over a path length of 0.355 mm. By 'tuning' the incident e^--beam intensity to $\sim 10^6$ e^- per pulse, and adjusting the placement of emulsions (Fig 5), an optimum density of one pair event per 50 $\mu \times 50$ μ microscopic field of view can be achieved. Standard, albeit tedious, multiple scattering measurements of the $e^+ - e^-$ trajectories then permit a reconstruction of the energy spectrum of the incident γ-ray. The results of one analysis comprising about 1400 pair events are shown in Figure 6 [11]. These 'signal' pairs had a sharply peaked distribution within a band about 2 mm wide; the pairs associated with the beam halo were more diffusely spread over a strip of width 10 mm. The ratio of halo/bremsstrahlung pairs in the emulsion volumes selected for signal scanning ranged from 0.091 to 0.21. As is apparent from the graph, the observed spectrum is in good agreement with the theoretical expectations.

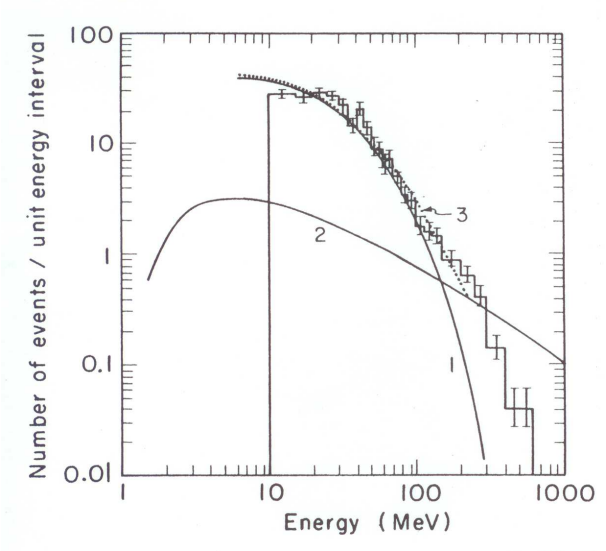


Figure 6

Magnetic bremsstrahlung (synchrotron radiation) spectrum derived from multiple scattering measurements of electron-positron pairs in Ilford G5 emulsions. Electron energy 19 GeV magnetic field strength 1.5 MG

1. Theoretical spectrum
2. Beam halo spectrum
3. Weighted composite of 1. and 2.

24.4 Aftermaths of SLAC

Analyses of the megagauss experiments at SLAC led to two new developments: (1) the discovery of synchrotron–Čerenkov radiation, and (b) a recalculation of quantum synchrotron radiation valid under extreme conditions.

24.4.1 Synchrotron Čerenkov Radiation:

The megagauss coils at SLAC had thin mylar inserts that prevented a shorting-out of the input leads. Could it be possible that this nominally trivial obstruction might affect the character of the radiation? This possibility is not totally implausible because of the sensitive dependence of synchrotron radiation on momentum transfer to the magnetic field. An elementary but basic consequence of the energy-momentum relations (3.1) is that free electrons can neither emit nor absorb photons. Čerenkov radiation is only possible because the photon momentum is shifted from its vacuum value $\hbar\omega/c$ to $n(\omega)\,\hbar\omega/c$, where the index of refraction $n(\omega)$ represents the momentum exchange with an ambient medium. Similarly, synchrotron radiation can only occur because the electron wave functions in a magnetic field are sufficiently 'stiff' so that momentum can be absorbed. A recalculation of synchrotron radiation, allowing for an index of refraction in the emitted photons, showed that the sensitive dependences on momentum transfers did indeed have unintuitive consequences [36, 16, 37]. For instance, practically all radiation from a 100 GeV electron passing through a 10 kG field is quenched below 10 keV by the mere presence of air at STP. The connection with SLAC emerges by considering 20 GeV electrons passing through lucite or mylar in the presence of a 1.5 MG field; in this case radiation below 100 keV is completely suppressed [38]. A point of fundamental interest is that because magnetic fields in vacuum are birefringent [39], 'pure' synchrotron radiation is only a first approximation to 'vacuum polarization synchrotron-Čerenkov radiation': this variant can produce exotic effects in some astrophysical situations. The existence of Airy function striations — another nonintuitive feature of synchrotron-Čerenkov radiation — was verified experimentally by McDonald [40].

24.4.2 Quantum Theory of Magnetic Bremsstrahlung:

The SLAC experiments with 19 GeV electrons demonstrated the feasibility of combining high energy accelerators with pulsed megagauss generators. Beyond this technical threshold, accelerator energies have increased towards 7 TeV, and

megagauss devices are capable of producing pulsed fields in the range $5-10$ MG. Using nominal values of 5 TeV and 5 MG, one can estimate that the Sokolov-Ternov quantum fluctuation parameter (1.1c) is 3×10^{15}, the quantum parameter Υ (1.2b) is 1, and the radiation reaction ratio (1.5) is 5×10^{4}. A fully relativistic quantum theory of synchrotron radiation applicable to these extreme conditions cannot rely on heuristic simplifications. However, a direct evaluation of the matrix elements is also impractical because they involve associated Laguerre functions (equivalent to confluent hypergeometric or Whittaker functions [22]) whose arguments (κ), index ($\hat{\alpha}$), and order ($\hat{n}$) are all very large numbers. A typical Laguerre function $\mathcal{L}_{\hat{n}}^{(\hat{\alpha})}(\kappa)$ is specified by

$$\begin{aligned}
&\kappa \sim 8.45\times10^{7}\cdot(\hbar\omega)^{2}\ (\mathrm{MeV})/\mathrm{H}\ (\mathrm{MG}) \rightarrow 4.2\times10^{12}\\
&\hat{\alpha} \sim 1.69\times10^{11}\cdot(\hbar\omega)\ (\mathrm{MeV})\cdot\mathrm{E(GeV)}\ /\mathrm{H}\ (\mathrm{MG}) \rightarrow 8.5\times10^{16}\\
&\hat{n} \sim 8.45\times10^{13}\cdot E^{2}\ (\mathrm{GeV})\ /\mathrm{H}\ (\mathrm{MG}) \rightarrow 4.2\times10^{20}
\end{aligned} \tag{4.1}$$

where $E\sim5$ TeV, $H\sim5$ MG, and the photon energy is $\hbar\omega\sim500$ MeV. A straightforward bremsstrahlung calculation planned to deal with this problem by using an asymptotic expansion due to Erdélyi to simplify the Laguerre functions [41]. Ironically, the attempt to improve the quantum theory of synchrotron radiation led to results that were palpably incorrect [42]. Extensive checks showed that Erdélyi's work contained errors, and that it was necessary to start anew.

A reliable asymptotic approximation for $\mathcal{L}_{\hat{n}}^{(\hat{\alpha})}(\kappa)$ must satisfy two conditions: (i) it must be sufficiently simple to be amenable to explicit calculations; and (ii) the difference between $\mathcal{L}_{\hat{n}}^{(\hat{\alpha})}(\kappa)$ and the asymptotic approximations must be constrained by rigorous error bounds that can also be computed explicitly. Fortunately, techniques developed by Cherry [43] and especially Olver [44] could be adapted to the construction of suitable uniform asymptotic approximations for the Laguerre functions. Figure 7 illustrates the divergence between the old Erdélyi-Swanson approximation and the new uniform asymptotic expansion for an extremely simple case where the three arguments of the Whittaker function are very small in comparison with the values of (4.1) [22]. The situation is of course far worse for the large parameters in (4.1).

After replacing the faulty Erdélyi approximations with the new asymptotic expansions, the program of recalculations originally envisaged in [42] could be carried to completion. Figure 8 gives an indication of the nature of the results by comparing a set of predictions for the total rate of magnetic bremsstrahung for unrestricted values of the quantum parameter Υ. The series expansions in

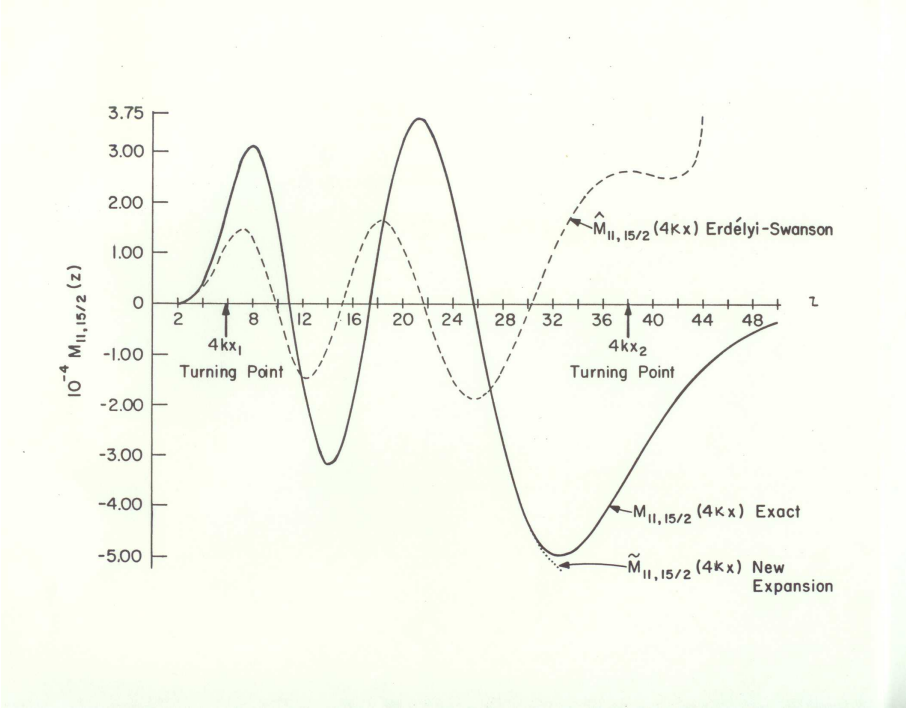


Figure 7: Comparison of the Exact values of $M_{11,\,15/2}(z)$ with the New and Erdélyi-Swanson Asymptotic Expansions

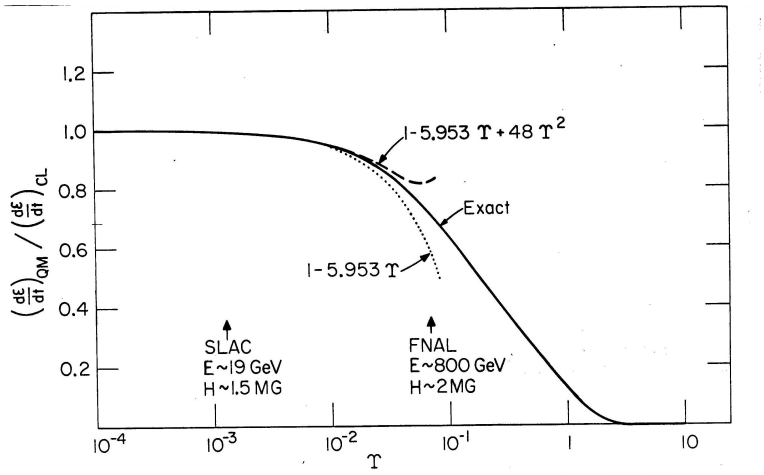


Figure 8

Υ refer to the expressions previously given in (1.3); the 'exact' curve is derived from the new results, and is mathematically accurate in the sense that the numerical error bounds are known and small. It is interesting that under extreme conditions the total radiation rate tends to decrease relative to the classical results.

Figure 9 shows that the decrease can be partially understood as a consequence of the modified shape of the quantum bremsstrahlung spectrum I_Q. In the classical case, the intensity spectrum I_c corresponds to the Fourier resolution of a continuous energy flow and is not constrained by energy conservation, i.e., the requirement that the photon energy ($\hbar\omega$) cannot exceed the radiating electron's energy. In quantum theory the radiative power is assocated with a stochastic photon flux, and the energy limit $\hbar\omega \leq E$ is automatically satisfied [45]. Finally, Figure 10 shows that under still more extreme conditions, the bremsstrahlung spectrum develops a peak at the tip $\hbar\omega \simeq E$. Unfamiliar features of this kind may occur in 'magnetars' where it is surmised that the ambient fields are of the order of 10^{15} G, even exceeding the critical field (1.1d) [46].

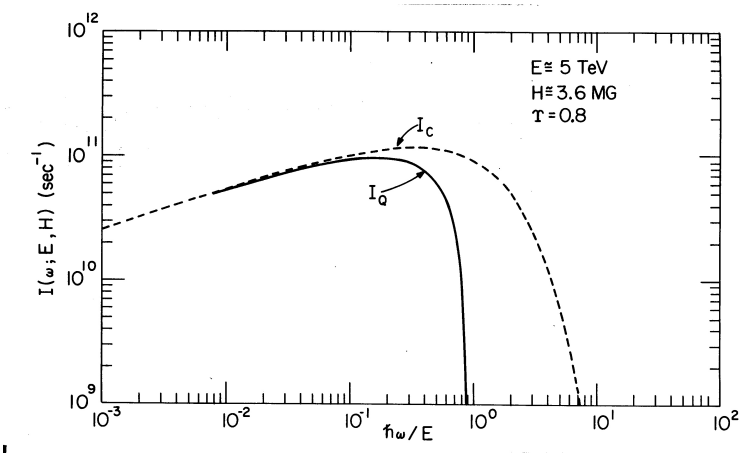


Figure 9

24.5 Endpapers

Christopher Merrill concludes his 'memories' (Chapter 4) by asking "In your opinion, what are the most important open questions in theoretical physics?" No doubt, the nature of complex systems will be one of the dominant themes of scientific research in the twenty-first century. But we have to be modest in our ambitions; Leo Kadanoff seems to have learned this lesson: "At one time, many people believed that the study of complexity could give rise to a new science. In this science, as in others, there would have been general laws, with specific situations being understandable as the inevitable working out of these laws of nature. Up to now, we have not found any such laws. ... So, even though there is apparently no science of complexity, there is much science to be learned from complex systems" [47]. Just how modest we have to be can be illustrated by a number of examples.

24.5.1 Associating scalars with arbitrary sets

Let A be a bounded subset of n-dimensional space and m an associated scalar, e.g. a 'measure'. Now impose the following three requirements: (1) If A is cut up into a finite number of disjoint sets that are reassembled to form a set B, then A and B have the same measure; (2) the measure of the union of disjoint sets is the

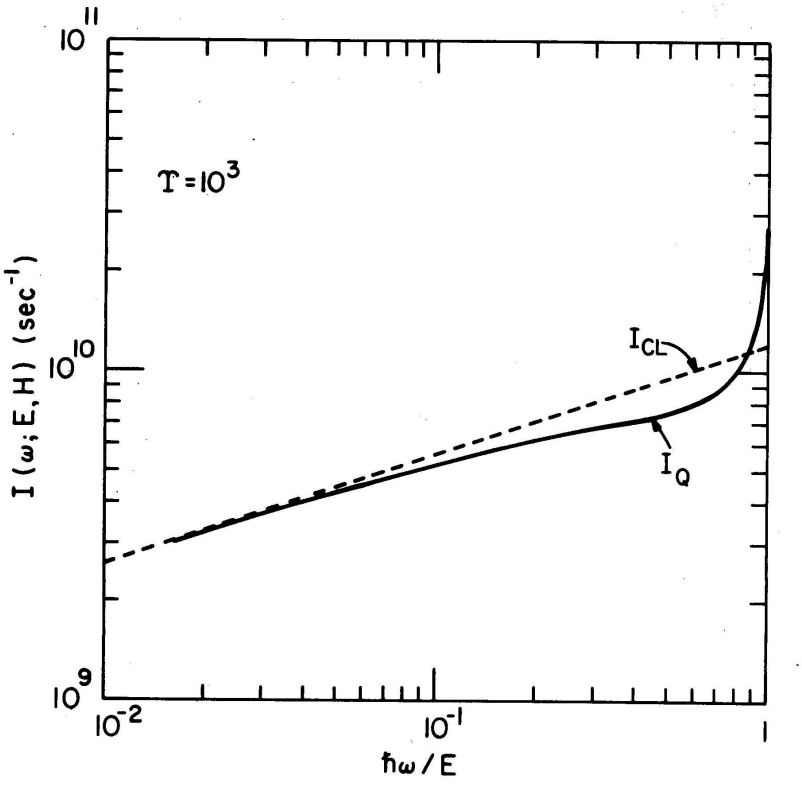

10^{11}
10^{10}
10^{9}
$\Upsilon = 10^{3}$
I_{CL}
I_Q
$I(\omega; E, H)$ (sec^{-1})
10^{-2}
10^{-1}
1
$\hbar\omega / E$

Figure 10

sum of the measures; and (3) the measure of the n-dimensional unit cube equals 1 [normalization]. As a physicist would say with somewhat less exactitude, m is a collective or extensive quantity such as volume, mass, or total energy. Now the surprising result is that these apparently self-evident criteria conceal a threshold of structural complexity. If the dimension of the sets satisfies $n \leq 2$, a well-defined m exists; otherwise not! The famous Banach-Tarski 'paradox' which shows that, in principle, a billiard ball could be decomposed and reassembled into a sphere the size of the sun, is an example of the structural complexity that can prevail in 3 dimensions [48]. The Atkinson-Johnson billiard ball conundrums (Chapter 8) may be distant relatives of these problems.

24.5.2 Consistency, Completeness, Decidability; or not

Axiom systems have a logical purity that attracts theoretical physicists who may be uncomfortable with the ambiguities of the laboratory. Clearly, axiom systems should be free of internal inconsistencies. Furthermore, they should be complete — that is, adequate to determine the truth or falsity of every statement within the compass of the theory. Finally, is the system decidable in the sense that there is some systematic procedure for determining the truth or falsity of every such statement? It was originally proved by M. Presburger that the additive theory of natural numbers (a rudimentary system in which the basic relation is addition) is consistent, complete, and decidable. However, the theory of natural numbers including both operations of addition and multiplication is undecidable [48]. The general form of this result is Gödel's famous incompleteness theorem which effectively states that no sufficiently strong system is complete; it was subsequently proved that no such system is decidable. It certainly seems that these results still "... remain remote from most areas of mathematics and irrelevant to the efforts of most workaday mathematicians. But that's just not so! Undecidable problems surround us everywhere ..." [49]. If Gödel's tenebrous musings have not yet reached physics, perhaps Seth Putterman's article (Chapter 21) is the sound of a key turning in the lock.

24.5.3 Hilbert's Thirteenth Problem

Those who believe that the book of nature is written in mathematical language ought to be aware of Hilbert's conjecture "... there are continuous functions of three variables not representable by continuous functions of two variables" [50]. Physicists will immediately think of the connection with so-called 'three-body forces' which have been mentioned in the literature on molecular interactions

and nuclear physics. If there is something inherently more complicated in three variable systems then we would have another example of a threshold of structural complexity. Mathematically, it turned out that Hilbert's expectations were not confirmed. If by 'representation' one also allowed the substitution of functions into other functions, i.e. functional composition, then it was shown by Kolmogorov that n-variable continuous functions could be represented (encoded) by combinations of functions of one variable. The physical implications of these results are not clear. As pointed by Present [51] we do not have any experimental criteria for distinguishing between 'genuine' three-body forces and non-linear superpositions of two-body forces.

24.5.4 Arrow's Impossibility Theorem

Suppose three voters, call them I, II, and III, are each given two alternatives, say A and B; how can one in a fair way aggregate their individual preferences into a collective decision? There would be little argument with the suggestion that majority rule would settle the matter with no acrimony. But what is the situation if the three voters were given the option of choosing among three alternatives, A, B, and C? This turns out to be another threshold of structural complexity because with three alternatives the intransitive set of preferences $A < B < C < A$ can occur ($A < B$ means B is preferred over A, etc.). Now majority rule is no longer the obvious 'fair' voting recipe, and other schemes need to be devised. This subject is of great social importance and has a long history. Finally, around 1950, the economist Kenneth Arrow gave a sharp axiomatic formulation of the meaning of 'fair' (no dictators, independence of irrelevant alternatives, etc.) and proved a surprising (depressing?) impossibility theorem — no voting rule satisfying the fairness axioms existed [52]. The implication for economics is that no universally valid scheme for assigning monetary value to economic transactions exists. The implications for physical processes are still open.

References

[1] Albert Einstein: Philosopher - Scientist, P. A. Schilpp, editor (Library of Living Philosophers, Evanston, IL 1949).

[2] The Classical Theories of Radiation Reaction, T. Erber, Fortschritte der Physik, 9, 343-392 (1961).

[3] High-Energy Electromagnetic Conversion Processes in Intense Magnetic Fields, T. Erber, Rev. Mod. Phys., 38, No. 4, 626-659 (1966).

[4] Megagauss Physik, H. G. Latal, Acta Phys. Austriaca 34, 65-82 (1971).

[5] Flux Compression Theories, T. Erber and H. G. Latal, Repts. Progress in Physics, 33, No. 11, 1069-1127 (1970).

[6] The Los Alamos Primer, Robert Serber (University of California Press, Berkeley, 1992).

[7] Production of Very High Magnetic Fields by Implosion, C. M. Fowler, W. B. Garn, and R. S. Caird, J. Appl. Phys. 31, 588-594 (1960).

[8] The IIT Flux Compression Facility, T. Erber, G. K. Forsberg, H.G Latal, J. A. Mazzie, and J. E. Kennedy, Les Champs Magnétiques Intenses, CNRS Paris, 335-343 (1967).

[9] Megagauss Cnare Targets for High Energy Magnetic Bremsstrahlung, T. Erber, Zeit. Angewandte Phys. 24, No. 4, 188-191 (1968).

[10] Experiments with Megagauss Targets at SLAC, F. Herlach, R. Mc Broom, T. Erber, J. J. Murray, and R. Gearhart, Proc. IEEE Trans. Nucl. Science, NS-18 No. 13, 809-814 (1971).

[11] Measurement of Magnetic Bremsstrahlung Spectrum and Pair Production Cross Sections, M. Mashkour, PhD Dissertation, Illinois Institute of Technology, December 1972.

[12] Measurement of Electron Pair Production Cross Sections, M. Mashkour, Phys. Rev. A8, No. 5, 2342-2347 (Nov. 1973).

[13] Climbing the Mountain: The Scientific Biography of Julian Schwinger, J. Mehra and K. A. Milton (Oxford U Press, Oxford, 2000).

[14] Synchrotron Radiation, A. A. Sokolov and I. M. Ternov (Pergamon Press, New York, 1968).

[15] Quantum Electrodynamics, A. A. Sokolov, I. M. Ternov, V. Ch. Zhukovskii, and A. V. Borisov, (Mir Publishers, Moscow,, 1988).

[16] Inner Bremsstrahlung Processes II, T. Erber, D. White, H. G. Latal, Acta Phys. Austriaca 45, 29-64 (1976).

[17] Quantum Modifications in Magnetic Bremsstrahlung, H. G. Latal and T. Erber, Annals Phys. (NY) **108**, No. 2, 408-442 (1977).

[18] Electromagnetic Radiation and the Mechanical Reactions Arising from It, G. A. Schott (Cambridge U. Press, Cambridge, 1912).

[19] The Quantum Correction in the Radiation by Energetic Accelerated Electrons, J. Schwinger, Proc. Natl. Acad. Sci. USA, **40** 132-136 (1954).

[20] New Approach to Quantum Correction in Synchrotron Radiation, J. Schwinger and W-y Tsai, Annals Phys. (NY) **110** 63-84 (1978).

[21] Megagauss Bremsstrahlung and Radiation Reaction, T. Erber, G. B. Baumgartner Jr., D. White, and H. G. Latal, Proc. 12^{th} Intern. Conf. High Energy Accelerators, Fermilab, 372-374 (1983).

[22] Uniform Asymptotic Approximations for the Whittaker Function $M_{\kappa,m}(z)$, G. B. Baumgartner, Jr., PhD Dissertation, Illinois Institute of Technology, December 1980.

[23] Some External Field Problems in Quantum Electrodynamics, T. Erber, Acta Phys. Austriaca Suppl. **VIII**, 323-357 (1971).

[24] Classical Charged Particles, F. Rohrlich (Addison-Wesley, Cambridge MA, 1965, 1990).

[25] Introductory Quantum Electrodynamics, E. A. Power (Elsevier, New York, 1964).

[26] A Two-Body Problem of Classical Electrodynamics: The One-dimensional Case, R. D. Driver, Annals of Phys. (NY) **21**, 122-142 (1963).

[27] Existence and Continuous Dependence of Solutions of a Neutral Functional-Differential Equation, R. D. Driver, Archive for Rat. Mech. Analysis, **19**, 149-186 (1965).

[28] Radiation Reaction in the Two-Body Problem of Classical Electrodynamics, D. P. Hsing and R. D. Driver, Tech. Rept. No. 61. Dept. Mathematics, U. Rhode Island, 1975.

[29] The Production of Megagauss Magnetic Fields for the Stanford Magnetic Bremsstrahlung Experiment, R. C. Mc Broom, MSc Dissertation, Illinois Institute of Technology, December 1971.

[30] D. White, unpublished.

[31] Sul Meccanismo dell'emissione Nella Meccanica Ondulatoria, E. Fermi, Rend. Lincei **5**, 795-800 (1927).

[32] J. Schwinger, private communication.

[33] Quantum Electrodynamics and Radiation Reaction: Nonrelativistic Atomic Frequency Shifts and Lifetimes, J. R. Ackerhalt amd J. H. Eberly, Phys. Rev. **D10**, No. 10, 3350-3375 (1974).

[34] Dynamics of a Classical Quasi-Point Charge, F. Rohrlich, Phys. Lett. **A303**, 307-310 (2002).

[35] Quantentheorie der Strahlung, A. Einstein, Phys. Zeitschrift **18**, 121-128 (1917).

[36] Čerenkov- Magnetobremsstrahlung, T. Erber, Notices Am. Math. Soc. **22**, No. 4, A472 (1975).

[37] Classical and Quantum Theory of Synergic Synchrotron-Čerenkov Radiation, J. Schwinger, W.-y. Tsai, and T. Erber, Annals of Physics (NY) **96**, No. 2, 303-332 (1976). [Reprinted in the 40^{th} Anniversary Edition of Annals of Physics, **281**, No. 1-2, 1019-1048 (2000).]

[38] Experimental Aspects of Synchrotron-Čerenkov Radiation, T. Erber, D. White, W.-y. Tsai, and H. G. Latal, Annals of Phys. (NY) **102**, No. 2, 405-447 (1976).

[39] Propagation of Photons in Homogeneous Magnetic Fields: Index of Refraction, T. Erber and W.-y. Tsai, Phys. Rev. **D12**, No. 4, 1132-1137 (1975).

[40] Observation of Interference Between Čerenkov and Synchrotron Radiation, K. D. Bonin, K. T. McDonald, D. P. Russel, and J. B. Flaz, Phys. Rev. Lett. **57**, No. 18, 2264-2267 (1986).

[41] Asymptotic Forms of Whittaker's Confluent Hypergeometric Functions, A. Erdélyi and C. A. Swanson, Mem. Am. Math. Soc. No 25, (1957).

[42] Quantum Modifications in Magnetic Bremsstrahlung, H. G. Latal and T. Erber, Annals Phys. (NY) **108**, No. 2, 408-442 (1977).

[43] Uniform Asymptotic Formulae for Functions with Transition Points, T. M. Cherry, Trans. Am. Math. Soc. **68**, 224-257 (1950).

[44] Asymptotics and Special Functions, F. W. J. Olver (Academic Press, New York, 1974).

[45] Unified Radiation Formulae for Classical and Quantum Electrodynamics, T. Erber and H. G. Latal, Eur. J. Phys. **24**, 67-79 (2003).

[46] Strongest Magnet in the Cosmos, S. Zane and R. Turolla, Physics World, **16**, No. 1, 19-20 (2003).

[47] Turbulent Heat Flow: Structures and Scaling, L. P. Kadanoff, Physics Today **54**, No. 8, 34-39 (2001).

[48] Alfred Tarski: Life and Logic, A. B. Feferman and S. Feferman (Cambridge U. Press, Cambridge, 2005).

[49] Can't Decide? Undecide, C. Goodman-Strauss, Notices Am. Math. Soc. **57**, No. 3, 343-356 (2010).

[50] The 13^{th} Problem of Hilbert, G. G. Lorentz, Proc. Symposia in Pure Math. **28**, 419-430 (1976).

[51] Non-additive Interactions, R. D. Present, Contemp. Phys. **12**, No. 6, 595-602 (1971).

[52] Social Choice and Individual Values, K. J. Arrow (Yale U. Press, New Haven, 1951).

Appendix A

Curriculum Vitae

Thomas Erber

Professor of Physics
Illinois Institute of Technology
Chicago IL 60616

Academic Record

B.S. (Physics) Massachusetts Institute of Technology	1951
M.S. (Physics) University of Chicago	1953
Ph.D. (Physics) University of Chicago	1957
Assistant Professor of Physics	1957-1962
Illinois Institute of Technology	
Associate Professor of Physics	1962-1964
Illinois Institute of Technology	
Professor of Physics	1964-
Illinois Institute of Technology	
(Emeritus as of May 2008)	
Professor of Mathematics	1986-
Illinois Institute of Technology	
Distinguished Professor	1999-
Illinois Institute of Technology	

IIT Faculty Research Fellow	1958-1959
Research Fellow, Université Libre de Bruxelles	1963-1964
Visiting Scientist Stanford Linear Accelerator Center	1970
Visiting Professor of Physics, University of Graz	1971
Professor Honoris Causa, University of Graz	1971
Visiting Professor of Physics, University of California at Los Angeles	1978
Visiting Professor of Physics, University of Grenoble	1982
Visiting Professor of Physics, University of Graz	1982
Visiting Professor of Physics, University of California at Los Angeles	1984-1985
Visiting Professor of Physics, University of California at Los Angeles	1987, 1992, 2006
Visiting Scholar, Enrico Fermi Institute, University of Chicago	1998-1999

Professional Affiliations

- Fellow of American Physical Society
- American Mathematical Society
- IEEE, Senior Member
- Magnetics Society
- Society of Sigma Xi (IIT Chapter President 1968-69)
- European Physical Society
- Österreichische Physikalische Gesellschaft
- Editorial Board, Acta Physica Austriaca
- Advisory Board, Research Corporation
- New York Academy of Sciences
- American Academy of Mechanics

- Nuclear, Plasma, and Magnetics Societies (IEEE)
  Chicago Chapter President (1996-1997)
- Fellow of the Institute of Physics (UK)
- Victor Franz Hess Gesellschaft

Publications

[1] Annals of Physics, 6, 319-340 (1959)
Coherent Compton Amplitude at High Energies

[2] Annals of Physics, 8, 435-472 (1959)
The Photoelectric Effect at High Energies

[3] American Journal of Phys. 27, 607 (1959)
Transformation of Acceleration in Special Relativity
(with R. Malhiot)

[4] Journal of Mathematics and Physics 38 331-334 (1960)
Stirling Numbers and Hypergeometric Functions

[5] Archive for Rational Mechanics and Analysis 4, 341-351 (1960)
Inequalities for Hypergeometric Functions

[6] Nature 190 25-27 (1961)
Velocity of Light in a Magnetic Field

[7] Skandinavisk Aktuarietidskrift, 27-28 (1960)
The Gamma Function Inequalities of Gurland and Gautschi

[8] Progress of Theoretical Physics 25, 714-716 (1961)
Čerenkov Self-Excitation of The Electron

[9] Fortschritte der Physik 9, 343-392 (1961)
The Classical Theories of Radiation Reaction

[10] Proceedings of the International Conference on High Magnetic Fields, H. Kohn et al. M.I.T. Press and J. Wiley, pp. 702-718, (1962)
The Index of Refraction of a Magnetic Field

[11] Discovery **24**, No. 2, 10-12 (1963)
The Nature of Magnetism

[12] Mathematics of Computation **17**, 162-169 (1963)
Tabulation fo the functions $\partial I/\partial n$, $n = \pm 1/3$
(with Alan Gordon)

[13] Bulletin Classe des Sciences, Académie Royale de Belgique, **Vol. 50**, pp. 328-338 (1964)
Poynting Vector and Energy Density in Dispersive Media

[14] Acta Physica Austriaca **19**, 17-44 (1964)
Čerenkov Radiation from Extended Charge Structures
(with Hsio-Chang Shih)

[15] Proceedings of the International Conference on the Role of Atomic Electrons in Nuclear Transformations, Polish Academy of Sciences (1965) **Vol. 3**, 520-528
Vacuum Polarization in Mu-Mesic Atoms
(with H. C. Shih)

[16] Reviews of Modern Physics **38**, 626-659 (1966)
High Energy Electromagnetic Conversion Processes in Intense Magnetic Fields

[17] Les Champs Magnètiques Intenses, CNRS Paris pp. 335-343,(1962)
The IIT Flux Compression Facility
(with G.K.Forsberg, H.G. Latal, J.A. Mazzie, J.E. Kennedy)

[18] Bulletin Classe des Sciences, Académie Royale de Belgique, **Vol. 53, No 9**, 1019-1042 (1967)
A State-Area Principle for Magnetic Condensation Processes
(with H.G. Latal)

[19] Physica **37**, 489-491 (1967)
A State-Area Principle for Magnetic Phase Transitions
(with H.G. Latal)

[20] Prospects for Simulation and Simulators of Dynamic Systems, edited by G. Shapiro and M. Rogers, Spartan Books, N.Y., pp. 203-226 (1967)
Study of a Magnetic Cooperative System
(with G.R. Marousek and G.K. Forsberg)

[21] Zeitschrift für angewandte Physik Vol. 24, No. 4, 188-191 (1968)
Megagauss Cnare Targets for High Energy Magnetic Bremsstrahlung

[22] Acta Physica Austriaca, 28, 325-335 (1968)
Electrodynamics of Extended Charge Structures
(with H.C. Shih)

[23] Frances Bitter: Selected Papers and Commentaries, MIT Press $x + 551$ pages (1959)
(edited by T. Erber and C.M. Fowler)

[24] Review of Scientific Instruments 41, 1-7 (1970)
Electromagnetically Driven Flux Compression
(with F. Herlach and D. Kachilla)

[25] Acta Physica Austriaca 30, 271-294 (1969)
Study of a Magnetic Cooperative System
(with G.R. Marousek and G.K. Forsberg)

[26] Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Mathem.-naturw. Klasse, Abteilung II, Vol 178, Nos. 1-3, 25-38 (1969)
Ein idealisierte Behandlung der magnetischen Flusskompression
(with H.G. Latal and P. Urban)

[27] Reports on Progress in Physics 33, 1069-1127 (1970)
Flux Compression Theories
(with H.G. Latal)

[28] Communications in Mathematical Physics 20, 205-219 (1971)
Probabilistic Metric Spaces and Hysteresis Systems
(with A. Sklar and B. Schweizer)

[29] Acta Physica Austriaca **32**, 224-244 (1971)

Force-Free Modes, Work-Free Modes, and Radiationless Electromagnetic Field Configurations

(with S. M. Prastein)

[30] Proc. X International Schadming Conference, Acta Physica Austriaca, Suppl. VIII, Springer Verlag, pp. 323-357, (1971)

Some External Field Problems in Quantum Electodynamics

[31] Proceedings IEEE Transactions Nuclear Science, NS-18, No. 13, pp. 809-814 (1971)

Experiments with Megagauss Targets at SLAC

(with F. Herlach, R. McBroom, R. Gearhart, J.J. Murray)

[32] Advances in Chemical Physics, edited by I. Prigogine and S. Rice, J. Wiley, New York **Vol 20**, 71-134,(1971)

The Origin of Hysteresis in Simple Magnetic Systems

(with H.G. Latal and B.N. Harmon)

[33] Annals of Physics **69**, 161-192 (1972)

A General Phenomenology of Hysteresis

(with S.A. Guralnick and H.G. Latal)

[34] Acta Physica Austriaca **34**, 313-330 (1971)

Statistical Theory of Spontaneous Magnetization in a Simple System

(with H.G. Latal)

[35] Acta Physica Austriaca **34**, 337-344 (1971)

Fourier Analysis of Flux Compression Fields

(with H.G. Latal and P. Urban)

[36] Acta Physica Austriaca **34**, 345-350 (1971)

Population Inversion in a Magnetic Condensation

(with H.G. Latal)

[37] Acta Physica Austriaca **36**, 171-197 (1972)
Analysis of Flux Compression Experiments I
(with J.E. Kennedy, H.G. Latal, S.M. Prastein)

[38] Acta Physica Austriaca **36**, 257-280 (1972)
Analysis of Flux Compression Experiments II
(with J.E. Kennedy, H.G. Latal, S.M. Prastein)

[39] Acta Physica Austriaca **36**, 316-353 (1972)
Analysis of Flux Compression Experiments III
(with J.E. Kennedy, H.G. Latal, S.M. Prastein)

[40] Communications in Mathematical Physics **29**, 311-317 (1973)
Mixing Transformation on Metric Spaces
(with B. Schweizer and A. Sklar)

[41] Magnetism and Magnetic Materials – 1972 AIP Conf. Proc. No. 10, part 2, 1710-1714 (1973)
Breaking of Dilatational Symmetry of Dipole Interactions by Octupole Forces
(with M. Duda, R. Olenick, H.G. Latal)

[42] Astrophysical Journal **184**, No. 1, part 1, 301-304 (1973)
Radiative Corrections and Soft Photon Emission in Magnetic Bremsstrahlung
(with H.N. Spector)

[43] American Journal of Physics **42**, 338-339 (1974)
A "λ-Transition" of Two Magnetic Dipoles
(with R. Olenick)

[44] Modern Developments in Thermodynamics, edited by B. Gal-Or; J. Wiley / Israel Universities Press, New York-Jerusalem, pp. 281-301, (1974)
Macroscopic Irreversibility as a Manifestation of Micro-Instabilities
(with A. Sklar)

[45] Nuclear Instruments and Methods **118**, 147-148 (1974)

Comments on "Separated High-Energy Electron Beams Using Synchrotron Radiation"

(with D. White)

[46] Physical Review **D10**, 492-499 (1974)

Photon Pair Creation in Intense Magnetic Fields

(with Wu-Yang Tsai)

[47] Physical Review **D12**, 1132-1137 (1974)

Propagation of Photons in Homogeneous Magnetic Fields: Index of Refraction

(with W-Y Tsai)

[48] Annals of Physics **96**, 303-332 (1976)

Classical and Quantum Theory of Synergic Synchrotron-Čerenkov Radiation

(with J. Schwinger and W-Y Tsai)

[49] Acta Physica Austriaca **44**, 315-336 (1976)

Inner Bremsstrahlung Processes: I

(with D. White and H.G. Latal)

[50] Acta Physica Austriaca **45**, 29-64 (1976)

Inner Bremsstrahlung Processes: II

(with D. White and H.G. Latal)

[51] Acta Physica Austriaca **45**, 245-254 (1976)

The Propagation of Photons in Homogeneous Magnetic Fields II: Dispersion Relations and Propagation Modes

(with W-Y Tsai)

[52] Physics Letters **57A**, 357-358 (1976)

Magnetic Detection of Crack Initiation and Propagation

(with J.E. Nuti and S.A. Guralnick)

[53] Annals of Physics **102**, 405-447 (1976)
Experimental Aspects of Synchrotron-Čerenkov Radiation
(with D. White, W-Y Tsai, H.G. Latal)

[54] Annals of Physics **108**, 408-442 (1977)
Quantum Modifications in Magnetic Bremsstrahlung
(with H.G. Latal)

[55] Journal of Applied Physics **49**, 2233-2240 (1978)
The Angular Distribution of Synchrotron Čerenkov Radiation
(with T.M. Rynne and G.B. Baumgartner, Jr.)

[56] Physical Review **D18**, 2152-2165 (1978)
Interference between Transition and Čerenkov Radiation
(with L.L. DeRaad, Jr. and W-Y Tsai)

[57] Applied Physics Letters **33**, 908-909 (1978)
Transition Radiation from Neutral Beams of Hydrogen Isotopes
(with L.L. DeRaad, Jr.)

[58] Journal of Computational Physics **32**, No. 2, 168-211 (1979)
The Simulation of Random Processes on Digital Computers with Čebešev Mixing Transformations
(with P. Everett and P.W. Johnson)

[59] Applied Physics Letters **35**, 752-754 (1979)
Optical Synchrotron-Čerenkov Radiation
(with T.M. Rynne)

[60] Journal of Applied Physics **52**, No. 3, 1944-1946 (1981)
Dilational Symmetry Breaking in Magnetic Cooperative Systems
(with H.G. Latal and R.P. Olenick)

[61] Acta Physica Austriaca **53**, 145-155 (1981)
Quantum Mechanics and Mixing Transformations
(with T.M. Rynne and A. Sklar)

[62] Physics Letters 85A, 61-63 (1981)
Čebyšev Mixing and Harmonic Oscillator Models
(with P. Everett and P.W. Johnson)

[63] Journal of Computational Physics 49, No. 3, 394-419 (1983)
The Simulation of Random Processes on Digital Computers: Unavoidable Order
(with T.M. Rynne, W.F. Darsow, M.J. Frank)

[64] Proc. 12th International Conference on High-Energy Accelerators, Fermilab, pp. 372-374 (1983)
Megagauss Bremsstrahlung and Radiation Reaction
(with G.B. Baumgartner, Jr., D. White, and H.G. Latal)

[65] Proc. 12th International Conference on High-Energy Accelerators, Fermilab, pp. 375-375 (1983)
Synchrotron-Čerenkov Radiation
(with G.B. Baumgartner, Jr., M.J. Lundsten, T.M. Rynne)

[66] Journal of Structural Engineering 110, No. 9, 2103-2119 (1984)
Plastic Collapse, Shakedown, and Hysteresis
(with S.A. Guralnick and S. Singh)

[67] Proc. 17th International Conference on Low Temperature Physics, edited by V. Eckern, A. Schmid, W. Weber, H. Wuhl, North Holland, Vol. 2, 935-936, (1984)
SQUID Detection of Barkhausen Emission and Onset of Hysteresis in Iron
(with H. Weinstock and M. Nisenoff)

[68] Physical Review B31, 1535-1553 (1985)
Threshold of Barkhausen Emission and Onset of Hysteresis in Iron
(with H. Weinstock and M. Nisenoff)

[69] Nature 318, No. 6041, 41-43 (1985)
Randomness in Quantum Mechanics: Nature's Ultimate Cryptogram?
(with S. Putterman)

[70] Philosophical Magazine **B52**, No. 5, 963-969 (1985)
Memory Dependence of Barkhausen Emission
(with J.L. Porteseil and P. Molho)

[71] Journal of Structural Engineering **112**, No. 12, 2610-2627 (1986)
Plastic Collapse, Shakedown, and Hysteresis of Multistory Steel Structures
(with S.A. Guralnick, J. Stefanis, O. Soudan)

[72] Journal of Structural Engineering **114**, No. 1, 31-49 (1988)
Energy Method for Incremental Collapse Analysis of Framed Structures
(with S.A. Guralnick, J. Stefanis, O. Soudan)

[73] Annals of Physics **181**, 25-33 (1988)
Hysteresis and Incremental Collapse: The Iterative Evolution of a Complex System
(with S.A. Guralnick)

[74] Journal of Applied Physics **63**, 3952-3955 (1988)
Training of Barkhausen Emission in Nickel and Iron
(with H. Weinstock)

[75] Annals of Physics **190**, 254-309 (1989)
Resonance Fluorescence and Quantum Jumps in Single Atoms: Testing the Randomness of Quantum Mechanics
(with P. Hammerling, G. M. Hockney, M. Porrati, S. Putterman)

[76] Physics Letters **A141**, 43-47 (1989)
Quantum Jumps in a Single Atom: Prolonged Darkness in the Fluorescence of a Resonantly Driven Cascade
(with S. Putterman)

[77] Journal of Applied Physics **68**, No. 3, 1370-1372 (1990)
Onset of Hysteresis Measured by Scanning Tunneling Microscopy
(with H.A. McGreer, E.R. Nowak, J-C Wan, H. Weinstock)

[78] Computers and Structures **37**, No. 6, 883-892 (1990)
A Computational Strategy to Determine the Incremental Collapse Envelope for Framed Structures
(with S.A. Guralnick, J. He, O. Soudan)

[79] Journal of Structural Engineering **117**, No. 6, 1815-1833 (1991)
Incremental Collapse of Structures with Constant Plus Cyclically Varying Loads
(with S.A. Guralnick, O. Soudan, J. He)

[80] Physica **A117**, Nos. 1-3, 394-400 (1991)
The Iterative Evolution of Complex Systems
(with D. Gavelek)

[81] Continuum Mechanics and Thermodynamics **3**, 293-309 (1991)
Thermodynamics of Plastic Hinges with Damage
(with B. Bernstein and S.A. Guralnick)

[82] Journal of Physics A **24**, No. 24, L1369-L1377 (1991)
Equilibrium Configurations of N Equal Charges on a Sphere
(with G.M. Hockney)

[83] Journal of Computational Physics **101**, 25-30 (1992)
Shadowing and Iterative Interpolation for Čhebyšev Mixing Transformations
(with D. Gavelek)

[84] Annals of Physics **224**, 157-192 (1993)
Hysteresis and Fatigue
(with S.A. Guralnick and S.C. Michels)

[85] Chaos, Solitons, and Fractals **3**, No. 3, (269-277) (1993)
Sneaky Plasticity and Mesoscopic Fractals
(with B. Bernstein and M. Karamolengos)

[86] European Journal of Physics **15**, 111-118 (1994)
Modern Calculus Notation Applied to Physics

[87] Annals of New York Academy of Sciences **755**, 748-757 (1995)

Testing the Randomness of Quantum Mechanics: Nature's Ultimate Cryptogram?

[88] Advances in Chemical Physics (edited by I. Prigogine and S. A. Rice), J. Wiley, New York, **98**, 495-594 (1997)

Complex Systems: Equilibrium Configurations of N Equal Charges on a Sphere ($2 \leq N \leq 112$)

(with G.M. Hockney)

[89] Journal of Physics D: Applied Physics **30**, 2818-2836 (1997)

Piezomagnetism and Fatigue I

(with S.A. Guralnick, R.D. Desai, W. Kwok)

[90] Journal of Physics A: Mathematical and General **32**, No. 12, 2263-2284 (1999)

Chiral Hausdorff Metrics and Structural Spectroscopy in a Complex System

(with L. Coffey and J. Drapala)

[91] Journal of Physics A: Mathematical and General **32**, No. 43, 7581-7602 (1999)

Reversibility, Irreversibility: Restorability, Non-Restorability

(with B. Bernstein)

[92] The Engineering Science of Structures, edited by J. Shen, Whitehall, pp. 61-68 (2000)

From Reversibility to Irreversibility: A Canonical Series of Processes

48 Annals of Physics **281**, 1019-1048 (2000)

Classical and Quantum Theory of Synergic Synchrotron Čerenkov Radiation

(with J. Schwinger and W-Y Tsai)

Reprinting of (48) in the Fortieth Anniversity of the Annals of Physics

[93] European Journal of Physics **22**, No. 5, 491-499 (2001)

Hooke's Law and Fatigue Limits in Micromechanics

[94] European Journal of Physics **24**, 67-79 (2003)
Unified Radiation Formulæ for Classical and Quantum Electrodynamics
(with H.G. Latal)

[95] Foundations of Physics **34**, No. 10, 1515-1540 (2004)
Reflections on Parity

[96] Journal of Physics D: Applied Physics **41**, 115006 (11pp) (2008)
Piezomagnetism and Fatigue II
(with S.A. Guralnick and S. Bao)

It is widely asserted that the great physicists who grasped the full unity of physics are all dead, having been replaced in this age of specialization by scientists who have a deep understanding only for issues of rather limited scope.

Therefore, it is refreshing to consider **Tom Erber,** who has sought to find unity by working in many areas—basic and applied, experimental and theoretical, practical and fanciful. He has been a faculty member at Illinois Institute of Technology for more than four decades.

As a reflection of the breadth of Tom's interests in physics, this Festschrift in his honor contains articles by former students, colleagues at IIT, collaborators, and friends, which deal with a disparate variety of topics. The final article, by Tom Erber himself, is entitled ``Eigenschrift: The End of the Classical Theories of Radiation Reaction''. This article represents a survey of work done on this topic over his career, as well as prospects for future research.

Tom definitely sees the **big picture** in science, and many who have known him or have worked with him have been infused with his zest for exploration.

ILLINOIS INSTITUTE OF TECHNOLOGY

IIT Press

3300 S. Federal St., Room 301
Chicago, Illinois 60616