# International Journal of Fusion Energy

Vol. 3, No. 2 April 1985

Published Quarterly by the Fusion Energy Foundation

International Journal of Fusion Energy (ISSN: 0146-4981) is published quarterly by the Fusion Energy Foundation, P.O. Box 17149, Washington, D.C. 20041-0149. Tel. (703) 689-2490.

Subscription price: $80 per volume (four issues) domestic; $100 per volume foreign. Payment must be in U.S. dollars.

Address all correspondence to IJFE, Fusion Energy Foundation, P.O. Box 17149, Washington, D.C. 20041-0149.

Postmaster: Send address changes to IJFE, Fusion Energy Foundation, P.O. Box 17149, Washington, D.C. 20041-0149.





# International Journal of Fusion Energy

**Vol. 3, No. 2** **April 1985**

# Editorial Policy

The International Journal of Fusion Energy (IJFE) is an independent scientific journal published quarterly by the Fusion Energy Foundation. The IJFE is dedicated to the promotion of fundamental advance in science, with special emphasis on the following areas:

1. The physics of plasmas at high energy densities, and research bearing on the scientific and technological mastery of nuclear fusion processes.

2. Coherent, directed forms of electromagnetic action, including laser and particle beams and superconductivity.

3. The physics of living processes, with applications to fundamental problems of biology and medicine.

In addition to research articles and state-of-the-art reviews, the IJFE welcomes short, informal communications addressing questions of interest to researchers and others in the cited areas. Contributions in other fields will be accepted on the basis of extraordinary scientific interest or manifest relevance to the three specific fields covered by the journal.

IJFE will also run abstracts of relevant, recent published and unpublished work as a regular service to the reader. The editors will be grateful for references to significant new work, not previously covered in the abstracts and other departments of the journal.

# International Journal of Fusion Energy

**Vol. 3, No. 2** **April 1985**

## *Editor's Note*

# Were Gauss and Riemann Right About Quantum Electrodynamics?

by Jonathan Tennenbaum

This issue of the IJFE contains two novel contributions to the branch of physics commonly known as special relativity theory. The first, by Friedwardt Winterberg, presents a fundamental modification of the theory, based on the notion of an actually existent "quantum ether" characterized by a critical length and finite spatial extension of particles. Winterberg's theory would imply experimentally observable deviations from the classical Lorentz transform at very high energies.

The second paper, by Masayuki Note, presents an extension of special relativity to an "acceleration field," implying in particular that the mass of a planet depends upon its position in the gravitational field of the Sun. On this basis Note is able to derive in beautifully simple fashion the two results generally cited as "proofs" of general relativity: the precession of Mercury and the deflection of light near the Sun.

The papers of Winterberg and Note are published here not only for their intrinsic interest, but also in the hope that they may inspire and assist researchers in what the journal editors consider to be a long-overdue revision of the axiomatic assumptions of modern physics.

The summary remarks below are devoted to elucidating this point. We take as illustration the most exquisitely elaborated portion of modern physics, quantum electrodynamics.

For some time now, quantum electrodynamics has been reputed to be "the most accurate theory ever created by Man"—despite its inclusion of paradoxes and contradictions (the so-called infinite self-energy of the electron, in particular) that raise some question whether quantum electrodynamics constitutes a "theory" at all. So far, it seems, clever computational methods have made it possible to circumvent the infinities and to generate numbers in remarkably close agreement with experiment.

Recently, however, it has been reported that experiments at Lawrence Livermore National Laboratory on the radiative properties of multiply ionized atoms yield results totally at variance with relativistic quantum theory. Although the experiments in question were undertaken in connection with the development of high-power X-ray lasers and remain classified, published results by Charles Rhodes and his group at the University of Illinois point in the same direction.[1] It now appears likely that the "new physical principles" announced by Edward Teller in reference to classified research into directed energy weapons imply a major revision, if not a total revamping, of quantum electrodynamics.

This situation is pregnant with historical irony. For, as Uwe Parpart-Henke pointed out some years ago, the inconsistencies in present quantum electrodynamics are chiefly located at exactly those points where the architects of modern quantum physics departed methodologically from the electrodynamics research program pursued by the "Göttingen School" of Karl Gauss, Wilhelm Weber, and Bernhard Riemann more than a century ago.[2] Although neither Gauss nor Riemann is available today for comment, we can still say, "they told us so" when the holy edifice of quantum electrodynamics begins coming apart at the seams.

### The Existence of Singularities

These brief remarks are not the place to attempt a comprehensive summary of the Gauss-Riemann program.[3] The point of immediate relevance here—the crucial distinctions between Maxwell's electrodynamics and the Gauss-Riemann theory—has already been addressed in a number of documents in this and the preceding issue of the IJFE. We refer in particular to the papers on "retarded potential" by Riemann and his Italian collaborators Enrico Betti and Eugenio Beltrami.[4]

Maxwell's continuum field theory denied the existence of *singularities*. It therefore already contained the root fallacy that cropped up later in the "wave-particle paradox" of quantum mechanics and in Einstein's unsuccessful attempts to construct a general relativity theory able to account for the electron and other particles. Philosophically, the Maxwell approach involves a denial of the possibility of *real change* within the universe. Change is admitted in Maxwellian physics only in the form of continuous variation of scalar parameters (for example, field strengths) within a fixed topology.

This crucial axiomatic feature—or, we assert, *fallacy*—in Maxwell's physics, was inherited in full by Einstein in his work on general relativity. Einstein recognized that any correct electrodynamics must, unlike the Maxwell equations, be nonlinear. He sought some general principle through which to introduce the nonlinearity into his field equations in a unique and "natural" way. He hoped thereby to obtain a mathematical schema subsuming, in principle, the entirety of physical phenomena. At the same time, however, he insisted that there be *no singularities* in the theory: "If one had the field-equation of the total field, one would be compelled to demand that the particles themselves would everywhere be describable as singularity-free solutions of the completed field equations. Only then would the general theory of relativity be a complete theory."[5]

Einstein died in 1955 without having solved this problem to his satisfaction. Apparently, he had not considered the possibility that the problem, as posed, might be literally *impossible*—that the requirement of completeness on the one hand, and freedom from singularities on the other, might be mutually incompatible (at least in our universe). We are reminded of Kurt Gödel's famous theorem about the incompleteness of formal mathematical systems, from whose devastating implications the modern formalists of "Big Bang" cosmology and "Grand Unification" are able to escape only by pleading "statistical indeterminacy." In fact, Gödel elucidated the philosophical implications of Einstein's approach most explicitly in 1949, by demonstrating the existence of solutions to Einstein's field equations—solutions known as "rotating universes"—in which travel backward in time would, in principle, be possible.[6] These rotating universes would admit of no consistent definition of past and future, no objective criterion of passage of time. Speaking of Einstein's principle of the relativity of simultaneity, Gödel observes: "It seems that one obtains an unequivocal proof for the view of those philosophers who, like Parmenides, Kant, and the modern idealists, deny the objectivity of change and consider change as an illusion or an appearance due to our special mode of perception."

## The Göttingen School

Gauss and Riemann were explicitly anti-Kantian, as Gauss's attacks on Kant's *a priori* categories of space and time and Riemann's writings on the principles of natural science, as well as his famous "Hypotheses Upon Which Geometry Is Based," most clearly document.[7] Göttingen was in fact a major center of opposition to Kant's philosophy, part of a current of opposition going back to Friedrich Schiller and, ultimately, to Gottfried Wilhelm Leibniz. In Riemann's view, not only the human mind (science), but the universe as a whole is a process of negentropic evolution, in which the progression to higher states of negentropy involves real change and an objective ordering of time.

The electrodynamics of the Göttingen School was profoundly shaped by this view of the universe. Gauss, Weber, and Riemann based their approach on the concept of *generation of singularities mediating changes in the topology and metrical characteristics of the potential field* (that is, the manifold of the physical process). It is sometimes remembered that the Göttingen School anticipated Lorentz's electron theory many decades earlier, in Wilhelm Weber's theory of "atomic electricity." But Riemann's outline of a physics based on the geometry of the successive generation of species of singularities goes far beyond anything Lorentz and his contemporaries could have imagined. In the intervening period, the philosophical bias underlying Maxwell's electrodynamics had virtually saturated the community of physicists. As a result, the "Göttingen School" program remains still to this day to be elaborated. We still lack even the beginnings of a truly scientific electrodynamics, one which would explicate the necessary existence and characteristics of the singularities designated as "photons" and "electrons."

Fortunately, although the Gauss-Riemann program as a whole remained virtually forgotten until now, certain aspects of the Göttingen thrust did take hold in the subsequent development of physics. The historical line leading in particular to Winterberg's investigations, which are published here, is worth examining in this context.

In his early thermodynamics work leading to the discovery of "wave mechanics," Erwin Schrödinger remarked that the properties of atomic particles, implied by the thermodynamics of gases, hardly resemble those of the "hard-ball" mass-points prescribed by Newtonian mechanics. Instead, Schrödinger writes, "One is reminded of the behavior of 'waves of finite amplitude.'"[8] He was referring to the 1859 paper in which Bernhard Riemann predicted the generation of singularities—"Verdichtungsstösse," now known as shock waves—in any plane acoustical wave of sufficiently large amplitude. The shock is generated when the acoustical process, the wave, *drives itself* to velocities higher than the local velocity of sound. This paper, of course, was the basis of the later development of supersonic aerodynamics. Schrödinger proposed that Riemann's "waves of finite amplitude" be taken as a paradigm for the generation of particles. But, he wrote, "the universal radiation as whose 'signals' or perhaps singularities corpuscles are to be defined, is something essentially more complicated than, say, the wave-radiation of the Maxwell theory. . . ."

Unfortunately, Schrödinger seems not to have posed *to himself the question: What physical limiting condition*, comparable to the velocity of sound for acoustical shock waves, might determine the generation of the particle-singularities? Schrödinger's subsequent attempts to develop a nonlinear wave equation to account for the stability (that is, self-focusing) of his particle "wave packets" faced a virtual inquisition at the hands of Niels Bohr. The demoralized Schrödinger was not able to bring his ideas to fruition, and later efforts by de Broglie and others (de Broglie's "double solution") have remained relatively sterile.

## Hydrodynamic Processes

*The line of investigation initiated by Riemann* persisted, however, in the investigations of Werner Heisenberg on a conjectured "fundamental wavelength" governing the generation of particles in cosmic ray showers. In 1938, Heisenberg proposed that phase changes must occur whenever the wavelength of an interaction, in the sense of de Broglie-Schrödinger, becomes smaller than a certain critical limit.[9] He stated also that the value of the fundamental wavelength must determine "hereditarily" the rest masses of the fundamental particles. Contrary to the rigid algebraicism typical of the later development of quantum electrodynamics, Heisenberg conceived of particle interactions as *hydrodynamic processes*. When a cosmic ray collides with a nucleus, the electromagnetic "bow wave" "detaches" from the decelerated particle and results in the formation of multiple electromagnetic shock waves manifested by the shower of secondary particles scattering from the scene of the collision. Heisenberg's thinking thus paralleled that of the Göttingen hydrodynamicists Ludwig Prandtl and Adolf Busemann, who were developing the theory of supersonic flight around the same time. The Göttingen hydrodynamics school, established by Felix Klein in 1904, was perhaps the main channel of influence of Riemann's mathematical methods into the 20th century.

*Heisenberg's line of inquiry* was carried forward by his student Erich Bagge,[10] and also forms the point of departure for Winterberg's investigations reported here. It is most suggestive to compare Winterberg's construction of a nonlinear Lorentz transformation with Bostick's "hydroelectromagnetic" electron model and Bagge's purely electrodynamical theory of beta decay, based on phase transitions from negative to positive energies. These contributions all point toward a new, nonlinear electrodynamics in which the particles, rather than being static entities existing side by side schizophrenically with a lifeless field, are instead singularities of real phase changes.

Masayuki Note's derivation of the chief "hard evidence" heretofore adduced in support of the general theory of relativity, from a natural and straightforward modification of special relativity, suggests that the needed nonlinear electrodynamics might turn out to be much simpler than Einstein suspected. However, we submit, it must be sought in quite the opposite direction to that pursued by Einstein: Rather than seek a singularity-free, complete mathematical model, develop instead a *universal, geometrical characterization of the process of singularity-formation*.

In doing so, we should recognize that the mere variation of scalar parameters is not an adequate characterization of action, or change, in our universe. Real change always involves the addition of a topological singularity, which then modifies the manifold of action in the manner indicated paradigmatically by Riemann's application of "Dirichlet's Principle" to the theory of functions of a complex variable.[11] In the typical case of a negentropic process, like that in living organisms, the addition of each new singularity mediates an increase in the rate of singularity-generation, as well as shifts in the harmonic characteristics, the spectra, of the process. If we look, for example, at the process of photosynthesis from the standpoint that the chlorophyll molecule is an electromagnetic configuration, and indeed that the growth of a plant is a typical electrodynamic process, then we obtain valuable insight into the nature of the new, nonlinear electrodynamics that it is our task now to develop.[12]

## References


1. A report on Rhodes's work appears in this issue, p. 56.
2. See Uwe Parpart, "The Concept of the Transfinite," *The Campaigner*, Jan.-Feb. 1976, p. 3.
3. The Fusion Energy Foundation plans to make available in translation the relevant original documents of the Göttingen School and to develop their implications for present-day research through appropriate articles in this journal.
4. See "On Electrodynamics" by Enrico Betti, IJFE **3 (1):** 89; "A Contribution to Electrodynamics" by Bernhard Riemann, IJFE **3 (1):** 91. The translation of Eugenio Beltrami's critique of Maxwell's theory will be published in a future issue.
5. See autobiographical notes in *Albert Einstein: Philosopher-Scientist* (Cambridge: Cambridge University Press, 1949), p. 81.
6. See Kurt Gödel's essay, "A Remark About The Relationship Between Relativity Theory and Idealistic Philosophy" in *Albert Einstein: Philosopher-Scientist*, pp. 557-562.
7. See Gauss's "Second Memoir on Biquadratic Residues," *Collected Works*, and "Fragmente philosophischen Inhalts" and "Erkenntnistheoretisches" in Riemann's *Collected Works*. English translations are in preparation for the IJFE.
8. *Physik. Zeitschrift* **27:** 95 (1926). See also "The Theoretical Impasse in Inertial Confinement Fusion" by Uwe Parpart, *Fusion*, November 1979, p. 31.
9. *Ann. Physik* **32:** 20 (1938).
10. E. Bagge, *Atomkernenergie* **25:** 251 (1975) and **26:** 70 (1975).
11. See Riemann's *Schwere, Electricität, und Magnetismus* (Hannover, 1876). For a general epistemological discussion of Dirichlet's Principle, see Lyndon H. LaRouche, Jr. *So, You Wish to Learn All About Economics?* (New York: New Benjamin Franklin House, 1984).
12. Articles on this subject will appear in the next issue of the journal.

# Nonlinear Relativity and the Quantum Ether

**F. Winterberg**

*Desert Research Institute*
*University of Nevada System*
*Reno, Nevada 89506*

*Abstract*—The orthodox approach to relativistic quantum field theory axiomatically postulates Lorentz invariance, followed by the axioms of quantum mechanics. Departing from this orthodox approach we have designed a heuristic procedure by which the Lorentz transformations follow from the interactions characterized by the quantum mechanical commutation rules, rather than being postulated a priori. The most general operator representation of the quantum mechanical commutation rules that involve a length constant, have in the past been rejected on the ground that they violate Lorentz invariance. We therefore have applied our heuristic principle to derive the corresponding transformation equations that take into account the length constant appearing in this most general representation. We find a nonlinear generalization of the Lorentz transformations, departing from special relativity at very high energies and establishing the observable existence of a substratum (ether). The principle that the velocity of light is the same in all reference system still holds, but the theory gives a finite zero point vacuum energy. Furthermore, a non-Lorentz invariant behavior of cross sections at very high energy is predicted. In the limiting case where the length constant is set equal to zero, the zero point energy diverges and special relativity is recovered.

The theory satisfies the philosophical principle by Leibniz, that the space-time structure should be determined from the interactions instead of being postulated a priori.

*"Einstein simply postulates what we have deduced . . . yet, I think, something may also be claimed in favor of the form in which I have presented the theory. I cannot but regard the ether, which can be the seat of an electromagnetic field with its energy and vibrations, as endowed with a certain degree of substantiality, however different it may be from all ordinary matter. In this line of thought, it seems natural not to assume at starting that it can never make any difference whether a body moves through the ether or not, and to measure distances and lengths of time by means of rods and clocks having a fixed position relative to the ether."*

—H. A. Lorentz

## Introduction

One of the most sacrosanct concepts of modern physics is the principle of relativity. It was motivated by the inability to observe an absolute motion through the hypothetical ether. Poincaré (1905), who first coined the term *principle of relativity*, still believed in the existence of an ether. It was Einstein (1905) who denied that an ether even existed and with it a distinguished absolute reference system in which the ether is at rest. Influenced by the positivism of Mach, Einstein concluded that what is not observable does not exist. However, it was soon found that through the quantum-mechanical zero-point field fluctuations, an ether of some sort reenters physics. It cannot be stressed strongly enough that this quantum ether is not an artifact of quantum theory, but makes itself felt through the Lamb shift and more directly through the Casimir effect. If special relativity is correct, this ether must have a divergent $\omega^3$ spectrum, the only one invariant under a Lorentz transformation, with the Doppler and aberration effects canceling each other out. Relativity simply requires that this quantum ether must appear equal in all reference systems, something that apparently only an infinite energy ether is able to do.

The real problem with the infinite quantum ether is that it would lead to infinite gravitational forces. Since nothing can be infinite, it has been frequently stated that general relativity will ultimately provide a natural cutoff at the Planck length of $\sim 10^{-33}$ cm. However, even though the mass density of the quantum-mechanical zero-point fluctuations becomes finite there, the value of the mass density turns out to be enormous, more precisely $\sim 10^{95}$ g/cm$^3$, large enough to put the entire mass of the known universe in a cube with a side length less than 1 fermi.

Another profound problem resulting from the same principle of relativity is the requirement for elementary particles to be mathematical points (Landau and Lifshitz 1975). This requirement is the reason that in relativistic quantum field theories the masses of elementary particles turn out to be infinite. In face of these infinities, it is a widespread myth that the theory of relativity is without exception in extremely good agreement with the empirical evidence. The infinite vacuum energy and the infinite masses of elementary particles it predicts are at gross variance with observation. But in spite of these infinities, relativity must be at least a very good approximation to reality. It is not just the success of relativistic mechanics, without which the design of high-energy particle accelerators would be impossible, or the well-founded energy-mass relation, supported by the empirical evidence in nuclear physics, but rather the extremely well-established agreement between theory and experiment in quantum electrodynamics. This theory, for example, correctly predicts the anomalous magnetic moment of the electron by seven significant places. Another example for the success of the theory is the Lamb shift, which is also explained to a very high degree of accuracy. In praising this success, it is however often overlooked, that both the anomalous magnetic moment and the Lamb shift are *relative* quantities. In their calculation the infinities are circumvented by the mathematical renormalization technique, with the infinities, like the self-energy of the electron, which is an *absolute* quantity, remaining infinite. These infinities in relativistic quantum field theories are the strongest reason that the theory must ultimately be wrong. Since these infinities express themselves through divergencies at very high energies, it would follow that relativity must break down at very high energies.

It is extremely unlikely that quantum theory, the other basic ingredient of relativistic quantum field theories, is the culprit for the unphysical infinities, because nothing like that happens in nonrelativistic quantum field theories, which, if applied to solid-state physics problems, always give finite results. Unlike the space-time of special relativity, a solid has a distinguished reference system at rest with the solid, and a natural cutoff provided through the lattice constant.

To remove the unphysical infinities from relativistic quantum field theories, it was advocated a long time ago by Heisenberg (1938) that a universal length should be incorporated into the fundamental laws of physics. We believe Heisenberg failed because he did not question relativity. Attempts made by Heisenberg and others to introduce a fundamental length into relativistic field theories led to a violation of time sequence causality. Time sequence causality means that in all reference systems, an effect must follow a cause. This is why elementary particles must have the structure of mathematical points and why the concept of a fundamental length is incompatible with special relativity.

As with the zero-point energy, some researchers have maintained that here too general relativity provides a natural cutoff, and that the size of elementary particles is the Planck length. However, this would mean that an electron, for example, would have an internal structure of about 20 orders of magnitude, from $10^{-13}$ cm down to $10^{-33}$ cm. This is the same ratio as 100 light years to 1 cm. According to the Exclusion Principle, it is hard to believe that two electrons with such a huge internal structure should then be completely indistinguishable.

In relativistic quantum field theories, *absolute* quantities turn out to be infinite, and *relative* quantities turn out to be finite. It therefore appears that *absolute* quantities can be computed only with an *absolute* theory. It is the property of *relative* versus *absolute* quantities that mirrors itself in *relative* versus *absolute* theories.

Apart from the unphysical infinities, there is another indication that relativity may be only an approximation to reality. The extremely uniform cosmic microwave background radiation defines a distinguished reference system, contradicting the spirit of relativity. The microwave radiation itself might be seen as the manifestation of a small temperature of the quantum ether in which this radiation is at rest.

A correct theory, however, not only should lead both to a finite vacuum energy and finite masses of elementary particles, but also must retain special relativity as a good approximation valid for sufficiently low energies to reproduce the extremely well established results of quantum electrodynamics, in addition to all the other confirmed results of special relativity.

## Relativity According to Einstein and Poincaré

In a search for possible generalizations of special relativity, a comparison of the mathematically equivalent, but otherwise different relativity theories developed by Einstein and Poincaré can be helpful. Even though mathematically equivalent, the theories are derived from different sets of axioms. In pure mathematics, a comparison of equivalent theories based on a different sets of axioms is quite useless. What is true in mathematics is far from true in physics. As Feynman (1965) observes:

Suppose you have two theories, A and B, which look completely different psychologically, with different ideas in them and so on, but that all the consequences that are computed from each are exactly the same, and both agree with experiment. The two theories, although they sound different at the beginning, have all consequences the same, which is usually easy to prove mathematically by showing that the logic from A and B will always give corresponding consequences. Suppose we have two such theories, how are we going to decide which one is right? There is no way by science, because they both agree with experiment to the same extent. So two theories, although they may have deeply different ideas behind them, may be mathematically identical, and then there is no scientific way to distinguish them.

However, for psychological reasons, in order to guess new theories, these two things may be very far from equivalent, because one gives a man different ideas from the other. By putting the theory in a certain kind of framework you get an idea of what to change. There will be something, for instance, in theory A that talks about something, and you will say, "I'll change that idea in here." But to find out what the corresponding thing is that you are going to change in B may be very complicated—it may not be a simple idea at all. In other words, although they are identical before they are changed, there are certain ways of changing one which look natural which will not look natural in the other.

In the spirit of Feynman's remark we therefore compare the theory of relativity as it was presented by Einstein, with the alternative axiomatic formulation presented by Poincaré. We first present Einstein's axiomatic formulation, followed by Poincaré's alternative formulation.

*A. Axioms by Einstein*

(1) The velocity of light *is* isotropic and *is* equal to *c in all* inertial reference systems.

(2) The principle of relativity (≡ the transformation equations are reciprocal).

*B. Axioms by Poincaré*

(1) All rods in absolute motion through the ether suffer a *real* (Fitzgerald-Lorentz) contraction.

(2) The velocity of light is isotropic and equal to $c$ in a distinguished reference system at rest with the ether but it also *appears* to be constant and equal to $c$ in all other inertial reference systems.

We now list the consequences derived from these different sets of axioms.

*Consequences of Einstein's Axioms*

(1) All clocks in relative motion *appear* to suffer a retardation.

(2) All rods moving in relative motion *appear* to suffer a contraction.

*Consequences of Poincaré's Axioms*

(1) All clocks in absolute motion through the ether suffer a *real* retardation.

(2) The coordinate transformations in between any two inertial frames moving relative to the ether are reciprocal (≡ principle of relativity).

*We note the following important differences.*

(1) Whereas *Einstein postulates* relativity as an *axiom, Poincaré derives* it as a consequence of the rod contraction for rods in absolute motion through the ether, and the constant light velocity postulate.

(2) Whereas in Einstein's theory the rod contraction and clock retardation always have only the character of an *appearance,* caused by relative motion, they are in Poincaré's theory the consequence of a *real* rod contraction and real clock retardation for rods and clocks in *absolute motion* through the ether.

(3) Whereas in Einstein's theory, the velocity of light *is* always isotropic and equal to $c$, in Poincaré's theory it only *appears* to be that way.

The reason for the difference in the wording *real* or *appear* has its root in each different viewpoint. For Poincaré, the velocity of light only *appears* to be constant and equal to $c$, because his measuring rods and clocks suffer a *real* change, whereas in Einstein's reasoning *exactly* the opposite is true.

To measure the velocity of light and therefore to define what is meant by constant light velocity requires the synchronization of clocks. There are basically two different ways in which clocks can be synchronized. In one method the synchronization is done by reflected light signals, and in the other method by slow clock transport. Einstein, in particular, proposed the synchronization by reflected light signals taking the arithmetic average of the time a signal needs for its round trip. This synchronization procedure, however, really makes sense only if the one-way velocity of light is already *assumed* to be equal to $c$. As Builder (1958) has pointed out, any experiment designed to measure the one-way velocity using clocks synchronized by Einstein's procedure is, in reality, always a two-way measurement because, through the clock synchronization procedure using reflected light signals, one always measures in reality only the average to-and-fro velocity. Einstein's one-way constant light velocity postulate and clock synchronization procedure are therefore a tautology.

It has been argued by Pais (1982) that Poincaré never really understood the theory of relativity, because he always insisted that the rod contraction for a rod in absolute motion through the ether must be a real physical effect, rather than an appearance as it was claimed by Einstein. We cannot agree at all with this opinion. Poincaré must have recognized that from a logical point of view, Einstein's alternative postulate, that light should propagate isotropically with the same velocity in all inertial reference systems (which is the constant one-way light velocity axiom), does make even less sense than the rod-contraction axiom. Lorentz (1909), who later succeeded in giving a plausible physical explanation for the reality of the contraction effect, shared Poincaré's view until his death.

Einstein rejected the ether hypothesis on the ground that the ether could not be experimentally verified. However, with his average time clock synchronization procedure he replaced the ether hypothesis with the isotropic one-way light-velocity hypothesis, which, like the ether hypothesis, is incapable of experimental verification. In Poincaré's view, one can, for the reason given, measure *only* the average to-and-fro velocity of light. This average is always $c$ and is the reason that the velocity *appears* to be constant and equal to $c$.

In spite of its logical shortcomings, Einstein's clock synchronization method, taking the arithmetic time average of the out and return trips for a reflected light signal, nevertheless remains very convenient.

A transparent physical exposition for Poincaré's point of view has been given by Prokhovnik (1967). He showed that the clock retardation effect can be understood as resulting from the combined action of the anisotropic light propagation in a frame in absolute motion through the ether and the rod-contraction effect. For his explanation he uses the concept of a light clock, which consists of a rod with two mirrors attached to its ends in between which a light signal can be reflected back and forth. Ordinary clocks are not built that way, but if all interactions in a solid body are electromagnetic and communicated by zero-rest-mass photons, real clocks should behave in exactly the same way as light clocks.

With the concept of a light clock and its retardation as a consequence of the rod-contraction effect, it is possible to reduce Poincaré's second axiom to the isotropic light propagation in the preferred reference system at rest with the ether. The rod-contraction effect and its resulting clock-retardation effect then explain why the velocity of light also appears the same in all other inertial reference systems. The isotropic light propagation in an ether frame of reference makes physical sense if the electromagnetic radiation is propagated as a wave. The rod-contraction effect therefore remained as the only unexplained part in an attempt to derive the theory of relativity from plausible physical concepts.

## The Theory of Lorentz and the Origin of the Infinities

In a crucial step going beyond Poincaré and Einstein, Lorentz tried to give a physical explanation of the theory of relativity (Prokhovnik 1967). According to Poincaré and Prokhovnik, the theory of relativity can be derived from the rod-contraction axiom alone. Therefore, all that had to be done was to give a physical explanation for the rod-contraction effect. It is of great importance that Lorentz actually succeeded in deriving the contraction effect for one very special case, thereby giving relativity a plausible physical explanation.

Possibly the best exposition of Lorentz's idea has been given by Bohm (1965):

> Lorentz assumed that the electrical forces were in essence states of stress and strain in the ether. From Maxwell's equation (assumed to hold in the reference frame in which the ether was at rest) it was possible to calculate the electromagnetic field surrounding a charged particle. For a particle at rest in the ether, it followed that this field was derived from a potential $\phi$, which was a spherically symmetric function of the distance $r$ from the charge, that is, $\phi = q/r$ (where $q$ is the charge of the particle). When a similar calculation was done for a charge moving with a velocity $v$ through the ether, it was found that the force field was no longer spherically symmetric. Rather its symmetry became that of an ellipse of revolution, having unchanged diameters in the directions perpendicular to the velocity, but shortened in the direction of motion in the ratio $\sqrt{1 - v^2/c^2}$. This shortening is evidently an effect of the movement of the electron through the ether.
>
> Because the electrical potential due to all the atoms of the crystal is just the sum of the potentials due to each particle out of which it is constituted, it follows that the whole pattern of equipotentials is contracted in the direction of motion and left unaltered in a perpendicular direction, in just the same way as happens with the field of a single electron. Now the equilibrium positions of the atoms are at points of minimum potential (where the net force on them cancels out). It follows then that when the pattern of equipotentials is contracted in the direction of motion, there will be a corresponding contraction of the whole bar, in the same direction, so that it will be shortened in the ratio $\sqrt{1 - v^2/c^2}$. As a result, a measuring rod of length $l_0$ at rest will, when moving with a velocity $v$ along the direction of its length, have the dimension
>
> $$l = l_0\sqrt{1 - v^2/c^2} \qquad (1)$$

But if the bar is perpendicular to the direction of motion, its length will of course not be altered.

Of particular importance is that the derivation of the length contraction (1) by Lorentz is valid only for *pointlike* charges, because only then is the electrostatic potential $\phi$ for all values of $r$ given by $\phi = q/r$. If, instead, the charge is spread out over a distance $r_0$, the potential for distances $r < r_0$ would be quite different, and the contraction could not be more exactly described by Eq. (1).

The physical reason for the infinities now becomes clear. Lorentz had shown that only point charges, and hence pointlike electrons, lead to the contraction factor given by Eq. (1) and therefore to relativity. However, the self-energy of point charges already diverges in classical physics, and the principle of relativity itself—which according to Lorentz can be derived only under the singular assumption of point charges—is therefore the reason for the infinities.

The derivation of the theory of relativity by Lorentz also shows that the relativistically invariant nonlinear theory proposed by Mie, or the related theory by Born and Infeld, are quite unsatisfactory (Sommerfeld 1952). Both theories lead to extended charges and hence a nondivergent force law. They therefore avoid the unphysical infinities. However, as Lorentz had convincingly shown, the theory of relativity can only be derived from the singular inverse square law. From Lorentz's point of view, the theories by Mie and Born-Infeld are inconsistent with the theory of relativity, because they both have a force law from which the theory of relativity cannot be derived.

A further problem for these theories arises from quantum field theory. In its language, extended particles have internal fields, which if quantized again lead to pointlike particles representing these internal fields and of which the extended particles are composed. This property of the theory appears to be very fundamental. It is why there is no known general prescription for the quantization of a nonlinear theory.

Rather than introducing a relativistically invariant nonlinear theory from which the infinities are removed, it is in the spirit of Lorentz much more logical to search for a generalization of special relativity, derived from a nonsingular interaction law, in the same way as special relativity was derived by Lorentz from the singular inverse square law.

We now understand the significance of Feynman's remark, because only in the framework of Poincaré's axiomatic formulation can we make a reasonable guess as to how the theory of relativity may be generalized to avoid the infinities. All that has to be done is to generalize the contraction formula by assuming extended rather than pointlike charges, and to compute the contraction effect in a way similar to that used by Lorentz for point charges. Starting from Einstein's axiomatic formulation it would be much more difficult to see how the theory could be generalized. It would require a generalization of the principle of relativity, and there is no clear prescription for how this could possibly be done.

It would be unsatisfactory simply to generalize the theory of Lorentz by making a calculation for extended charges, because it would give different results for different elementary particles having different charge distributions. What is needed, instead, is a universal principle, completely independent of the properties of elementary particles. It is most interesting that quantum mechanics, through the generalized representations of the commutation relations first proposed by Bagge (1962), provides for such a universal principle. These generalized representations permit the unambiguous introduction of a universal length constant by which the inverse square force law, valid for all fundamental interactions communicated by zero-rest-mass particles, can be regularized in a unique way. In the past, these generalized representations have been criticized on the ground that they are not Lorentz-invariant. We turn the argument around, assuming instead that the generalized representations of the commutation relations are correct and that the Lorentz transformations must be modified. The price to be paid is to abandon relativity, though it would still be a very good approximation for distances that are large compared to the fundamental length. Only at distances comparable to the fundamental length $r_0$, that is, at energies $E \sim \hbar c/r_0$, would relativity break down. Introducing a fundamental length that regularizes the force law necessarily eliminates the infinite self-energies of elementary particles.

As we shall see below, the so-defined generalized transformation equations replacing the Lorentz transformations not only turn out to be nonlinear, but they contain, in addition to the relative velocities, also absolute velocities against a substratum. In the proposed nonlinear theory, special relativity is a *singular* limit, and the singularities in relativistic quantum field theories are a reflection of this singular limit.

## Nonlinear Generalization of Special Relativity

To fully comprehend the approach taken, and to understand how the proposed nonlinear generalization of special relativity is derived, we first show how Einstein and Poincaré, using quite different lines of reasoning, arrived at the same transformation equations.

According to *Einstein*, the first axiom, that the velocity of light is constant in all inertial reference systems, leads to the equation interrelating two reference systems $I_A$ and $I_B$:

$$r_A^2 - c^2t_A^2 = \kappa(r_B^2 - c^2t_B^2), \quad (2)$$

where $\kappa$ is a constant and where

$$\begin{aligned} r_A^2 &= x_A^2 + y_A^2 + z_A^2 \\ r_B^2 &= x_B^2 + y_B^2 + z_B^2. \end{aligned} \quad (3)$$

Applying the second axiom, which is the principle of relativity, requires also that

$$r_B^2 - c^2t_B^2 = \kappa(r_A^2 - c^2t_A^2). \quad (4)$$

But this is only possible for $\kappa = \pm 1$. The negative sign is excluded from the identity when the $I_A$ and $I_B$ reference systems are the same. One therefore has

$$r_A^2 - c^2t_A^2 = r_B^2 - c^2t_B^2, \quad (5)$$

and one immediately verifies that Eq. (5) satisfies the Lorentz transformations for a uniform motion with the velocity $v$ along the $x$ axis:

$$\begin{aligned} x_A &= \gamma(x_B + vt_B) \\ y_A &= y_B \\ z_A &= z_B \\ t_A &= \gamma(t_B + vx_B/c^2) \end{aligned} \quad (6)$$

where

$$\gamma \equiv (1 - v^2/c^2)^{-1/2} \quad (7)$$

because

$$\begin{aligned} & r_A^2 - c^2t_A^2 \\ &= \gamma^2(x_B + vt_B)^2 + y_B^2 + z_B^2 - c^2\gamma^2(t_B + vx_B/c^2)^2 \\ &= r_B^2 - c^2t_B^2. \end{aligned} \quad (8)$$

According to *Poincaré*, the second axiom, which is that the light velocity appears to be constant and equal to $c$, not only in a frame $I_s$ at rest with the ether, but also in any other $I_A$ frame in relative motion against $I_s$, leads to

$$r_A^2 - c^2t_A^2 = \kappa\,(r_s^2 - c^2t_s^2), \quad (9)$$

where $r_s, t_s$ are measured in the ether rest frame $I_s$. As before, $\kappa$ is a constant to be determined. From Eq. (9) one obtains the transformation formulas

$$\begin{aligned} x_A &= \sqrt{\kappa}\gamma_A(x_s + u_At_s) \\ y_A &= \sqrt{\kappa}y_s \\ z_A &= \sqrt{\kappa}z_s \\ t_A &= \sqrt{\kappa}\gamma_A(t_s + u_Ax_s/c^2) \end{aligned} \quad (10)$$

where

$$\gamma_A \equiv (1 - u_A^2/c^2)^{-1/2}. \quad (11)$$

In Eqs. (10) and (11) $u_A$ is the *absolute* velocity of the $I_A$ system against the $I_s$ ether rest frame system.

From Poincaré's first axiom, that a rod in absolute motion through the ether is contracted by the factor $\sqrt{1 - u_A^2/c^2}$, it follows that $\kappa = 1$. Therefore, we can write

$$r_A^2 - c^2t_A^2 = r_s^2 - c^2t_s^2, \quad (12)$$

but also

$$r_B^2 - c^2t_B^2 = r_s^2 - c^2t_s^2, \quad (13)$$

and therefore

$$r_A^2 - c^2t_A^2 = r_B^2 - c^2t_B^2. \quad (14)$$

As in Einstein's approach, Eq. (14) then leads to the Lorentz transformations Eq. (6).

It was pointed out by Poincaré, in the case $\sqrt{\kappa} = 1$, the resulting transformations form a group. In his view, the appearance of this group is a direct consequence of the rod-contraction axiom.

We have already said that Poincaré's axioms offer a natural generalization of special relativity simply by admitting other rod-contraction laws. Under these circumstances, the resulting transformations no longer form a group, and the space–time symmetry of special relativity is broken. However, if the rod contraction only differs slightly from the Fitzgerald-Lorentz formula, the space–time symmetry of special relativity would remain a very good approximation.

Following Heisenberg's idea that a fundamental length constant should enter the basic laws of physics, we would like to formulate the generalized contraction formula in such a manner that it can incorporate a length. We therefore require that a rod in absolute motion through the ether, with the velocity $u_A$, shall be contracted by the amount $\sqrt{1 - u_A^2/c^2}F_A$. The factor $F_A$, which determines the departure from the Fitzgerald-Lorentz value, shall thereby depend on the absolute velocity $u_A = \beta_Ac$, but also on the ratio $r_A/r_0$, where $r_0$ is the universal length constant.

From Eq. (10) we now obtain:

$$\kappa = 1/F_A^2, \quad (15)$$

and hence for Eq. (9)

$$r_s^2 - c^2t_s^2 = (r_A^2 - c^2t_A^2)F_A^2. \quad (16)$$

By a similar argument applied to a rod moving through the ether with the absolute velocity $u_B$ we find

$$r_s^2 - c^2t_s^2 = (r_B^2 - c^2t_B^2)F_B^2. \quad (17)$$

We therefore have

$$(r_A^2 - c^2t_A^2)F_A^2 = (r_B^2 - c^2t_B^2)F_B^2. \quad (18)$$

Equation (18) is satisfied by the following transformation equations:

$$\begin{aligned} x_A &= (\gamma F_B/F_A)[x_B + vt_B] \\ y_A &= (F_B/F_A)y_B \\ z_A &= (F_B/F_A)z_B \\ t_A &= (\gamma F_B/F_A)[t_B + vx_B/c^2] \end{aligned} \quad (19)$$

where

$$\gamma \equiv (1 - v^2/c^2)^{-1/2}. \quad (20)$$

The same transformations were obtained previously (Winterberg 1984), following a much more arduous path along lines used originally by Lorentz to derive the Lorentz transformations.

The nonlinearity of the transformations (19) is caused by the dependence of $F_A$ and $F_B$ on $r_A/r_0$ and $r_B/r_0$, where $r_A = \sqrt{x_A^2 + y_A^2 + z_A^2}$ and $r_B = \sqrt{x_B^2 + y_B^2 + z_B^2}$. It must be emphasized that because of their nonlinearity, the coordinate values of Eq. (18) mean differences, with the measuring apparatus located at the origin of the respective coordinate system. It is therefore better to express the transformations in the inhomogeneous form, valid for an arbitrary position of the measuring apparatus, and where the symbol $\Delta$ signifies coordinate differences:

$$\begin{aligned} \Delta x_A &= (\gamma F_B/F_A)[\Delta x_B + v\Delta t_B] \\ \Delta y_A &= (F_B/F_A)\Delta y_B \\ \Delta z_A &= (F_B/F_A)\Delta z_B \\ \Delta t_A &= (\gamma F_B/F_A)[\Delta t_B + v\Delta x_B/c^2] \end{aligned} \quad (21)$$

We immediately see that the transformation equations (21) satisfy the relativistic velocity addition theorem. Putting $v_A = \Delta x_A/\Delta t_A$ and $v_B = \Delta x_B/\Delta t_B$ we find that

$$v_A = \frac{v_B + v}{1 + v_B v/c^2} \quad (22)$$

The validity of the relativistic addition theorem of velocities has the important consequence that $m = m_0\gamma$, where $m_0$ is the rest mass. From there on the expressions for momentum and energy are obtained, which turn out to be the same as in special relativity. If the theory could not reproduce this important result of special relativity, it would have to be abandoned right there.

From Eq. (16) it follows that $F_s = 1$. For an $I_A$ system moving with the velocity $u_A$ against the $I_s$ system we therefore have

$$\begin{aligned} \Delta x_s &= \gamma_A F_A(\Delta x_A + u_A\Delta t_A) \\ \Delta t_s &= \gamma_A F_A(\Delta t_A + u_A\Delta x_A/c^2) \\ \gamma_A &\equiv (1 - u_A^2/c^2)^{-1/2} \end{aligned} \quad (23)$$

with $v_s = \Delta x_s/\Delta t_s$, $v_A = \Delta x_A/\Delta t_A$, we obtain from Eq. (23)

$$v_s = \frac{v_A + u_A}{1 + v_A u_A/c^2} \quad (24)$$

In a similar way we obtain for a $I_B$ system moving with the velocity $u_B$ against $I_s$:

$$v_s = \frac{v_B + u_B}{1 + v_B u_B/c} \quad (25)$$

Eliminating $v_s$ from Eqs. (24) and (25) we obtain Eq. (22), if we put

$$v = \frac{u_B - u_A}{1 - u_B u_A/c^2} \quad (26)$$

Equation (26) shows that the absolute velocities $u_A$, $u_B$, appearing in $F_A$ and $F_B$ are not independent, but are rather related to each other through the relativistic velocity addition theorem for the relative velocity $v$.

Unlike the linear Lorentz transformations, the nonlinear transformations (21) also lead to a transverse contraction effect, expressed in the transformation formulas for $\Delta y$ and $\Delta z$. The general expressions for rod contraction and time dilation are given by

$$\begin{aligned} \Delta x_B &= \Delta x_A(F_A/F_B)\sqrt{1 - v^2/c^2} \\ \Delta y_B &= \Delta y_A(F_A/F_B) \\ \Delta z_B &= \Delta z_A(F_A/F_B) \end{aligned} \quad (27)$$

$$\Delta t_A = \frac{(F_B/F_A)\Delta t_B}{\sqrt{1 - v^2/c^2}} \quad (28)$$

From the transformation equations (21) we find that

$$(\Delta r_A^2 - c^2\Delta t_A^2)F_A^2 = (\Delta r_B^2 - c^2\Delta t_B^2)F_B^2. \qquad (29)$$

It therefore follows that the theory has the invariant

$$\Delta s^2 = \Delta s_0^2 F^2, \qquad (30)$$

where

$$\Delta s_0^2 = \Delta r^2 - c^2\Delta t^2 \qquad (31)$$

is the Minkowskian space–time difference element. Because in the expression (30) for the space–time interval, the Minkowski space–time interval is multiplied by $F$, we call $F$ the gauge function of the nonlinear theory.

The metric expressed by Eq. (30), although non-Euclidean, is conform invariant. This property has two important consequences. First, because wave equations for zero-rest-mass particles are conform-invariant, those particles are not deflected by the metric (30), and, second, the linear superposition principle is preserved.

There is, however, an important difference between the metric expressed by Eq. (30) and a Riemannian metric. In the latter the line element is expressed in differential form, and not through finite differences, as it is in the case of Eq. (30). This means that a metric of the form (30) cannot in general be expressed in a Riemannian form. Therefore, for such a metric the Pythagorean theorem in general does not hold in the small. That such a situation may actually arise was already recognized by Riemann himself, who in his Habilitation Paper (1854), states that in the small an even more complicated situation may arise if the presumed form of a line element through the square root of a quadratic differential does not take place.

### The Structure of the Gauge Function $F$

We demand that apart from its dependence on the absolute velocity $u = \beta c$, the gauge function $F$ shall also depend on the fundamental length $r_0$. For dimensional reasons, $F$ can then only be a function of $\beta$ and $r/r_0$. Furthermore, since $F_s = 1$, we can write

$$F(0, r/r_0) \equiv F_s = 1. \qquad (32)$$

To derive an expression for the function $F(\beta, r/r_0)$, we consider a rod of length $r$ moving with the velocity $u$ against the ether rest-frame system. According to Eq. (27) if seen from the $I_s$ system, its length $r'$ is

$$r' = r\sqrt{1 - \beta^2}F. \qquad (33)$$

For electric point charges, Lorentz found that $F = 1$. For his derivation he used electrostatic inverse-square-law forces acting between pointlike charges. As stated above, this idea suffers from generality. A derivation universally valid, not only for electrostatic interactions but for all other interactions as well, is suggested by quantum mechanics, where all interactions are communicated by probability waves. For pointlike particles in a frame at rest with the ether, these probability waves have a spatial dependence of outgoing spherical waves:

$$\psi = A\exp[i(kr - \omega t)]/r \qquad A = \text{const.} \qquad (34)$$

In the ether rest frame and at a distance $r$ away from the particles, the flux of the probability waves communicating a force to another particle is equal to

$$c\psi^*\psi = cA^2/r^2, \qquad (35)$$

where we have assumed that the probability waves are moving through the ether with the velocity $c$. Now, if two point particles are moving one behind the other with the velocity $u$ through the ether, the particle behind the first particle receives a flux of probability waves that is increased by the Doppler effect; that is, by the factor $(c + u)/c$. Likewise, the particle positioned ahead receives a smaller flux, diminished by the factor $(c - u)/c$. (See Figure 1.) The overall mutual interaction effect is thereby changed by the factor ($\beta \equiv u/c$):

$$(1 + \beta) \times (1 - \beta) = 1 - \beta^2. \qquad (36)$$

Therefore, if both particles were originally at rest in the ether frame and separated by the distance $r$, then after being set into motion through the ether with the velocity $u$, they would be subject to the same magnitude of interaction at a smaller distance $r'$ if

$$(1 - \beta^2)/r'^2 = 1/r^2. \qquad (37)$$

From Eq. (33) it follows that

$$r' = r\sqrt{1 - \beta^2}, \qquad (38)$$

which is the Fitzgerald–Lorentz contraction.

As we had seen above, special relativity can be understood as a consequence of the Fitzgerald–Lorentz rod-contraction effect in conjunction with the isotropic light propagation in a preferred ether rest frame. As the derivation of Eq. (38) makes clear, we have therefore succeeded in deriving special relativity from quantum-mechanical principles. The only ad hoc hypothesis we had to make in addition, was the assumption that the probability waves propagate isotropically with the velocity of light relative to an ether rest frame.

To see how the inverse square law (35) derived from quantum-mechanical principles can possibly be generalized, we consider the fundamental quantum-mechanical commutation relations (where $t$ must be a "$c$ number"):

$$
\begin{aligned}
pq - qp &= \hbar/i, \\
Et - tE &= -\hbar/i.
\end{aligned}
\tag{39}
$$

These commutation relations admit the Lorentz-invariant operator representation

$$
\begin{aligned}
p &\to (\hbar/i)(\partial/\partial q) \qquad & q &\to q, \\
E &\to -(\hbar/i)(\partial/\partial t) \qquad & t &\to t.
\end{aligned}
\tag{40}
$$

It is this representation that leads to wave equations where the probability amplitude is described by spherical waves of the form (34), and that communicates the interaction between pointlike particles.

The operator representation (40) of the commutation relations (39) is not the only one possible. It can also be satisfied by the generalized operator representation proposed by Bagge (1962):

$$
\begin{aligned}
p &\to (\hbar/i)(\partial/\partial q) \qquad & q &\to q + iq_0, \\
E &\to -(\hbar/i)(\partial/\partial t) \qquad & t &\to t + it_0,
\end{aligned}
\tag{41}
$$

where $q_0$ and $t_0$ are real constants.* (It would, of course, also be possible to put $q \to q + q_0$, $t \to t + t_0$, but this would only mean an uninteresting shift in the origin of the coordinate system.)

Furthermore, to reduce the two constants $q_0$, $t_0$, to just one universal length constant, we put

$$
c = q_0/t_0. \tag{42}
$$

For the nonlinear generalization of special relativity, the unknown function $F(\beta, r/r_0)$ has to be determined. We now show how this unknown function can be derived from the generalized representation of the fundamental commutation relations.

Making the substitution $q \to q + iq_0$, $t \to t + it_0$, and putting $q \equiv r$, $q_0 \equiv r_0$, transforms Eq. (34) into

$$
\psi = A\{\exp[i(kr - \omega t)]/(r + ir_0)\}. \tag{43}
$$

* Following Heisenberg's idea regarding a universal length constant the representation (41) was discussed in the late 1930s within the inner circle of the Heisenberg Institute. The late H. Euler (known for his work with Heisenberg on the nonlinear electrodynamics) in particular was intrigued by this generalization, because it opened a way for the *universal* introduction of a length constant into the laws of physics, but he did not find an answer to how these generalized representations of the commutation relations could be reconciled with relativity. (For this historical note, the author is indebted to E. Bagge.)

Equation (43) changes the flux of the probability waves from the inverse square law given by Eq. (35) to

$$
c\psi^*\psi = cA^2/(r^2 + r_0^2). \tag{44}
$$

If set into motion, the interaction is therefore again only changed by the factor $(1 - \beta^2)$, and instead of Eq. (37) we now have

$$
(1 - \beta^2)/(r'^2 + r_0^2) = 1/(r^2 + r_0^2). \tag{45}
$$

From Eq. (45) it then follows that

$$
r' = r\sqrt{1 - \beta^2}\left[1 - \left(\frac{\beta\gamma r_0}{r}\right)^2\right]^{\frac{1}{2}} \tag{46}
$$

which shows that

$$
F(\beta, r/r_0) = \left[1 - \left(\frac{\beta\gamma r_0}{r}\right)^2\right]^{\frac{1}{2}}. \tag{47}
$$

The nonsingular expression for the flux of the probability density shows how the generalized commutation relations regularize the inverse square interaction law. According to Eq. (43) the probability waves can be understood to emerge from a pointlike particle positioned at $r = -ir_0$. Equation (44), on the other hand, shows that in real space this corresponds to an extended particle. Because of this extension in real space there can be no infinite self-energies.

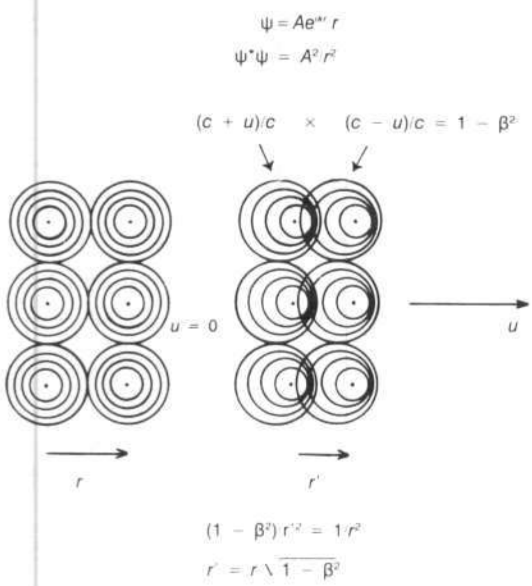



**Figure 1.** Quantum-mechanical explanation of Lorentz contraction. The interaction between pointlike particles leads to Lorentz contraction and therefore in the set of Poincaré's axiom to special relativity theory.

We can now complete the nonlinear transformation Eq. (21). We find

$$\begin{aligned} \Delta x_A &= (\gamma F_B/F_A)\,[\Delta x_B + v\Delta t_B] \\ \Delta y_A &= (F_B/F_A)\Delta y_B \\ \Delta z_A &= (F_B/F_A)\Delta z_B \\ \Delta t_A &= (\gamma F_B/F_A)\,[\Delta t_B + v\Delta x_B/c^2] \end{aligned} \tag{48}$$

$$\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$$

$$v = \frac{u_B - u_A}{1 - u_B u_A/c^2}$$

$$F_A = \left[1 - \left(\frac{\beta_A\gamma_A r_0}{\Delta r_A}\right)^2\right]^{\frac{1}{2}},\ \Delta r_A = \sqrt{\Delta x_A^2 + \Delta y_A^2 + \Delta z_A^2}$$

$$F_B = \left[1 - \left(\frac{\beta_B\gamma_B r_0}{\Delta r_B}\right)^2\right]^{\frac{1}{2}},\ \Delta r_B = \sqrt{\Delta y_B^2 + \Delta y_B^2 + \Delta z_B^2}.$$

The lateral radial contraction for an absolute motion with the velocity $u = \beta c$ against the substratum is finally given by

$$r_\perp = rF, \qquad F \equiv [1 - (\beta\gamma r_0/r)^2]^{1/2}. \tag{49}$$

If we introduce the relativistic specific impulse $u_0 = u\gamma$, furthermore, a specific velocity $c_0 = c(r/r_0)$, the longitudinal and lateral rod-contraction factors are given by

$$\begin{aligned} l_\parallel/l_0 &= \sqrt{1 - u^2/c^2} \times \sqrt{1 - u_0^2/c_0^2}, \\ l_\perp/l_0 &= \sqrt{1 - u_0^2/c_0^2}, \end{aligned} \tag{50}$$

where $l_\parallel$ is the length parallel to the direction of motion and $l_\perp$ is the length perpendicular to the direction of motion. These expressions show better than anything else why the generalized rod-contraction formulas obtained make sense.

It may superficially seem that in writing Eq. (36) we might have introduced an inconsistency by not using the relativistic formula for the Doppler effect. The contradiction is resolved if it is recognized that through the rod-contraction the measure of time is changed by the time-dilation equation. If the Doppler effect actually measured is corrected accordingly, the relativistic Doppler-effect formula is obtained. This means that the Doppler-effect shift $\nu/\nu_0 = 1 \pm \beta$ must (in the case of special relativity) be multiplied by the time-dilation factor $(1 - \beta^2)^{-1/2}$, with the result that $\nu/\nu_0 = \sqrt{(1 \pm \beta)/(1 \mp \beta)}$, which is the special relativistic formula for the Doppler effect. (A similar correction must be made in the nonlinear generalization of special relativity.) The Doppler-effect formula used in Eq. (36) is the *absolute* Doppler effect, which could only be observed if the measuring apparatus would remain uncontracted.

## The Non-Euclidean Space–Time Structure

The $1/r^2$ interaction law, valid for pointlike particles, is a fundamental expression for the Euclidean metric of the position space. Likewise, a force law of the form $1/(r^2 + r_0^2)$ is an expression for a non-Euclidean metric. It finds its expression through the finite difference line element (30) together with the expression of the gauge function given by Eq. (47). As pointed out above, it therefore represents a non-Riemannian metric, for which in general the line element cannot be expressed as a quadratic differential, even though this is still possible in certain cases.

To obtain a line element for this non-Euclidean, non-Riemannian metric we write the invariant (30), using the expression for $F$ given by Eq. (47):

$$\begin{aligned} s^2 &= (\mathbf{r}^2 - c^2t^2)F^2 \\ &= \mathbf{r}^2 - (\beta\gamma r_0)^2 - [1 - (\beta\gamma r_0/\mathbf{r})^2]c^2t^2. \end{aligned} \tag{51}$$

The space part of Eq. (51) is given by

$$\sigma^2 = \mathbf{r}^2 - (\beta\gamma r_0)^2. \tag{52}$$

If we call $\underline{\rho}^2 = \mathbf{r}^2 - (\beta\gamma r_0)^2$, then

$$\sigma^2 = \underline{\rho}^2. \tag{53}$$

Introducing spherical coordinates, we have

$$d\sigma^2 = d\rho^2 + \rho^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{54}$$

Along the radius vector one has

$$\rho = \sqrt{r^2 - (\beta\gamma r_0)^2}. \tag{55}$$

For a coordinate system in absolute motion against the substratum with the velocity $u$, $\beta = u/c = \text{const}$, one has

$$d\rho = \frac{dr}{\sqrt{1 - (\beta\gamma r_0/r)^2}} \tag{56}$$

and therefore

$$ds^2 = d\sigma^2 - (1 - (\beta\gamma r_0/r)^2)c^2dt^2$$

$$= \frac{dr^2}{1 - (\beta\gamma r_0/r)^2} + (1 - (\beta\gamma r_0/r)^2)r^2\,(d\theta^2 + \sin^2\theta d\phi^2) - (1 - (\beta\gamma r_0/r)^2)c^2dt^2\,. \qquad (57)$$

The line element (57) can be seen as the expression of a curved space–time, which, however, in contrast to general relativity, also depends upon the absolute velocity $u = \beta c$ of the reference system. The curved space–time metric is therefore different in reference systems having different absolute velocities.

With the metric tensor given by Eq. (57), it is then possible to write any wave equation, like the Klein–Gordon or Dirac equation, in a covariant form valid for a reference system in absolute motion against the substratum.

In the weak field approximation of general relativity, which is also valid here, the time part of the metric tensor is related to a gravitational potential $\phi$ by

$$[1 - (\beta\gamma r_0/r)^2]c^2dt^2 = (1 + 2\phi/c^2)c^2dt^2, \qquad (58)$$

hence

$$\phi = -(c^2/2)(\beta\gamma r_0/r)^2. \qquad (59)$$

Unlike the Newtonian gravitational potential this is a short-range velocity-dependent potential. Whereas Newtonian gravity is caused by the gravitational coupling constant, the short-range velocity-dependent gravitational potential given by Eq. (59) has its cause in the fundamental length constant $r_0$.

We remark that for zero-rest-mass particles the metric (57) is not the only one possible. This can be most easily seen from Eq. (51), because for zero-rest-mass particles $\mathbf{r}^2 - c^2t^2 = 0$, and hence there also

$$ds^2 = d\mathbf{r}^2 - c^2dt^2 = 0. \qquad (60)$$

Zero-rest-mass particles therefore have two branches of solutions along which they can propagate. One branch corresponds to an always-free motion. It is determined by the line element (60). The other branch is determined by Eq. (57), putting $ds^2 = 0$. It can even lead to trapped particle trajectories if $r \lesssim \beta\gamma r_0$, very much as it happens to photons trapped in a black hole.

## Consequences for High-Energy Physics

Departures from Lorentz invariance can be expected at very high energies where the interaction distance becomes comparable to $r_0$. The empirical velocity of $\simeq 390$ km/sec of the Earth frame against the substratum (assuming it is at rest with the microwave background) can be neglected at these energies, and the laboratory system approximated by an $I_s$ system. According to Eq. (49) (putting $\beta \simeq 1$) non-Lorentz effects become predominant if

$$r' = r/\gamma \leqslant r_0. \qquad (61)$$

Combining this with Heisenberg's uncertainty principle, we see that this happens for energies

$$E \gtrsim \hbar c/r_0. \qquad (62)$$

At teraelectron volt particle energies, soon to be reached, effects of this kind could be observed if $r_0 \simeq 10^{-17}$ cm. However, depending on the measurement accuracy, violation of Lorentz invariance may be observable even if $r$ is still much larger than $r_0$.

A direct way in which violation of Lorentz invariance could be observed is through the comparison of cross sections for interactions with the same center of mass energy, but with a different energy of the colliding particles relative to the $I_s$ system. Experimentally, this could be done with colliding beams of different velocities $u_A = \beta_A c$ and $u_B = \beta_B c$, which have approximately the same center of mass energy if

$$\gamma_0 = 2\sqrt{\gamma_A\gamma_B}, \qquad (63)$$

where $\gamma_0 \simeq E_{cm}/m_0c^2$, and $E_{cm}$ is the center-of-mass energy for two colliding particles of rest mass $m_0$.

For the cross sections of the colliding particles we must take into account the lateral contraction effect given by Eq. (50). The total geometric cross section is given by

$$\sigma = \pi(r_A + r_B)^2, \qquad (64)$$

where $r_A = rF_A$ and $r_B = rF_B$, and where we have assumed that two identical particles of radius $r$ collide. We further assume that $r >> r_0$ and can therefore put $r \simeq r_A$ and $r \simeq r_B$ into the expressions for $F_A$ and $F_B$. At high energies one can also put $\beta_A \simeq \beta_B \simeq 1$. Expanding up to second order in $r_0/r$, we find

$$\sigma \simeq \sigma_0[1 - \tfrac{1}{2}(r_0/r)^2(\gamma_A{}^2 + \gamma_B{}^2)], \qquad (65)$$

where $\sigma_0 = 4\pi r^2$. In combination with Eq. (63), we find that

$$\sigma \simeq \sigma_0[1 - (\gamma_0^2/8)(r_0/r)^2(x^2 + x^{-2})], \qquad (66)$$

where $\gamma_A = x\gamma_0/2$, $\gamma_B = (1/x)(\gamma_0/2)$. For $x = 1$ both beams have the same energy. Let us compare the value of $\sigma$ at $x = 1$ with some other value of $x$ by taking the difference

$$\Delta\sigma = \sigma(1) - \sigma(x) = \pi r_0^2\gamma_0^2[\tfrac{1}{2}(x^2 + x^{-2}) - 1]. \qquad (67)$$

If the maximum $\gamma$ value for each beam is $\gamma_{max}$, then $\gamma_0 = 2\gamma_{max}$, and hence

$$\Delta\sigma = 4\pi r_0^2\gamma_{max}^2[\tfrac{1}{2}(x^2 + x^{-2}) - 1]. \qquad (68)$$

If, for example, $x = \tfrac{1}{2}$ we find

$$\Delta\sigma = (9\pi/2)r_0^2\gamma_{max}^2. \qquad (69)$$

For $r_0 \simeq 10^{-17}$ cm and $\gamma_{max} \simeq 10^3$, one would find that $\Delta\sigma \approx 1.5 \times 10^{-27}$ cm$^2$. Of course, we do not know how small $r_0$ is, and for this reason $\Delta\sigma$ can be much smaller. For energies of ~100 GeV, which are presently available, and for the same $r_0$, one would have $\Delta\sigma \simeq 1.5 \times 10^{-29}$ cm$^2$.

### Compensation of the Zero-Point Energy

The total zero-point energy of a field in vacuum is obtained by the summation of the zero-point energy for all modes. For a harmonic oscillator of frequency $\omega$ the zero-point energy is $\tfrac{1}{2}\hbar\omega$. The volume in frequency space is $(2\pi c)^3$. If there are two directions of polarization the spectrum of the density $\varepsilon$ of the zero-point energy is given by

$$d\varepsilon/d\omega = \hbar\omega^3/2\pi^2c^3. \qquad (70)$$

In the context of our theory, this result is valid only in the distinguished reference system at rest with the substratum where $F = F_s = 1$. Integrating Eq. (70) up to the frequency $\omega$, one obtains the total energy density:

$$\epsilon = \frac{\hbar\omega^4}{8\pi^2c^3}. \qquad (71)$$

It has the mass density given by

$$\rho = \hbar\omega^4/8\pi^2c^5. \qquad (72)$$

Putting $\omega \simeq \pi c/r_0$, where $r_0$ is a cut-off length below which the zero-point energy vanishes by some unknown physical mechanism, one finds from Eq. (72) that

$$\rho \simeq \hbar/r_0^4c. \qquad (73)$$

Present thinking is that the cutoff for the zero-point energy results from ordinary gravity. It requires putting $r_0 \simeq (G\hbar/c^3)^{1/2} \simeq 10^{-33}$ cm. Inserting this number into Eq. (73) one finds that $\rho \simeq 10^{95}$ g/cm$^3$. This density is large enough to put the mass of the entire universe in a cube with a side length less than 1 fermi. This result is clearly nonsense, and it shows better than anything else that the present theory—and that means special relativity—must be wrong.

Experiments so far suggest no measurable deviation from Lorentz invariance down to $10^{-15}$ cm. The zero-point mass density should be therefore at least $10^{26}$ g/cm$^3$. Even this much smaller mass density appears still far too large, and if the principle of equivalence also holds for the zero-point energy, light should thus be appreciably deflected by this mass. A mechanism by which the gravitational effect of this large mass density can be compensated is therefore called for. It appears to be a very fortunate accident that our theory provides for just that mechanism through an attractive force, expressed by the velocity-dependent, short-range gravitational potential, Eq. (59).

A gravitational potential of the kind expressed by Eq. (59) exists only in the presence of a particle with nonvanishing rest mass. To make the compensation possible therefore requires the existence of a heavy massive particle. We then assume that two such heavy particles, each possessing the mass $m_0$, can be produced in pairs, with the energy supplied by the zero-point energy of the vacuum. For the compensation mechanism to work, it is only necessary that the two particles are continuously created and destroyed out of the vacuum. The particles thus produced form a two-body system. Each of the two particles moves with an absolute velocity around a common center of inertia. This absolute velocity $u = \beta c$ can be estimated from Heisenberg's uncertainty principle. With $m_0/2$ the reduced mass of the two-body system, Heisenberg's relation gives

$$\beta\gamma m_0 c \simeq 2\hbar/r. \qquad (74)$$

In this equation $r$ is the distance of separation in between both particles. A coordinate system at rest with one of these particles, moving with the absolute velocity $u = \beta c$ against the substratum, becomes, according to Eq. (59), the source of a negative gravitational-like potential.

To create two particles with the energy $2m_0c^2$ out of the vacuum, this energy must be balanced by the negative potential energy. One thus has

$$2m_0c^2 = -m_0\phi = (m_0c^2/2)(\beta\gamma r_0/r)^2. \qquad (75)$$

Furthermore, the rest mass $m_0$ must be related to $r_0$.

The only way this can be done is by putting

$$r_0 = \hbar m_0 c. \quad (76)$$

From Eqs. (74) through (76) we find

$$\beta\gamma = 2 \quad \text{and} \quad r = r_0. \quad (77)$$

If $\varepsilon_{tot}$ is the sum of the zero-point energy and its negative potential energy, one has

$$2\varepsilon_{tot} = 2\varepsilon + \rho\phi. \quad (78)$$

Inserting the values for $\varepsilon$ and $\rho$ given by Eq. (71) and (72) one finds

$$\varepsilon_{tot} = \varepsilon[1 - \tfrac{1}{4}(\beta\gamma r_0/r)^2] = 0. \quad (79)$$

The physical interpretation of this result is that the non-Euclidean structure in the small leads to a gravitational-like short-range force, resulting in a pressure that compensates the zero-point energy.

To this result we would like to add the remark that, instead of deriving the gauge function from quantum-mechanical postulates, we could have also derived it from the requirement that the zero-point energy should be compensated by a pressure.

### The Structure of the Substratum

According to our basic assumption, all interactions between elementary particles are communicated by probability waves propagating with the velocity of light through a substratum. This raises the question about the nature of this substratum. In 19th-century physics the substratum, called ether, was thought to be endowed with mechanical properties by which the observed nature of electromagnetic waves could be explained. In the quantum-mechanical view of the substratum, no such simple mechanical model is possible. Instead, the ether must be defined by a field equation. The postulate that the waves associated with this field have the character of probability waves moving with the velocity of light restricts the field equation of the substratum to a Dirac equation for spin-½, zero-rest-mass particles. For zero-rest-mass particles, the Dirac equation degenerates into the two-component equations first studied by Weyl and used for the description of neutrinos. The Weyl equation describes two zero-rest-mass, spin-½ particles with different helicity. This property means that the substratum is composed of two spin-½ fields. Still both equations can be combined into one 4-component Dirac equation given by

$$\gamma^\mu(\partial\psi/\partial x^\mu) = 0, \quad (80)$$

where

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\delta^{\mu\nu} \quad (81)$$

are the Dirac matrices. Equation (80) describes wave solutions in flat space for which the metric expressed by Eq. (60) is valid.

However, the metric of the curved space–time, Eq. (57), with $ds^2 = 0$, is also valid for zero-rest-mass particles. The Dirac equation in this metric is instead given by

$$\gamma^\mu(\partial\psi/\partial x^\mu) + \gamma^\mu\Gamma_\mu\psi = 0, \quad (82)$$

where

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \quad (83)$$

with $g^{\mu\nu}$ given by the line element (57). Note that the Dirac matrices in curved space–time determined by Eq. (83) have, in general, nonconstant elements, depending on the coordinates.

In Eq. (82) the ordinary differentiation is replaced by the covariant differentiation, where

$$\Gamma_\mu = \tfrac{1}{8}[\gamma^\nu(\partial\gamma_\mu/\partial x^\nu) - (\partial\gamma_\mu/\partial x^\nu)\gamma^\nu - \Gamma_{\mu\nu}{}^\rho(\gamma^\nu\gamma_\rho - \gamma_\rho\gamma^\nu)] \quad (84)$$

are the Fock–Ivanenko coefficients and $\Gamma_{\mu\nu}{}^\rho$ the Christoffel symbols of the second kind.

Comparing Eq. (82) with the Dirac equation for a massive particle

$$\gamma^\mu(\partial\psi/\partial x^\mu) + i(m/\hbar)\psi = 0, \quad (85)$$

we see that through the interaction with the short-range gravitational-like field, described by the line element (57), the zero-mass, spin-½ particles can acquire an effective mass given by

$$m = -i\hbar\gamma^\mu\Gamma_\mu. \quad (86)$$

This, of course, requires that the particles are trapped by the curved space–time within a small volume. The trapping itself can be instead approximately described by the gravitational like potential, Eq. (59). The short-range potential itself is set up by the two massive particles of mass $m_0$ created out of the vacuum. With the value $\beta\gamma = 2$ given by Eq. (77), we have for the potential (59):

$$\phi = -2c^2(r_0/r)^2. \quad (87)$$

With $mc^2 < m\phi$, it follows that the zero-rest-mass, spin-½ particles can be trapped if

$$r < \sqrt{2} r_0. \quad (88)$$

The mechanism thus described provides for the formation of massive particles as bound states built from zero-rest-mass particles, very much like the mass of a black hole can be enlarged by the entrapment of photons. In its spirit, the mechanism is reminiscent of Heisenberg's nonlinear spinor theory. However, in contrast to Heisenberg's nonlinear spinor equation, there is no cubic interaction term. The interaction is rather caused by the non-Euclidean structure in the small, and is determined by the Fock–Ivanenko coefficients, which in turn are determined by the metric tensor of the line element (57).

## Conclusion

It was originally claimed by Einstein that the theory of relativity could do away with the ether. This nice idea, however, did not last very long because it was later found that through the quantum-mechanical vacuum fluctuations, a zero-point-energy ether of infinite energy is introduced having an $\omega^3$ spectrum, which is the only Lorentz invariant spectrum where the Doppler and aberration effects cancel each other out under a Lorentz transformation. Furthermore, the requirement of causality led to the postulate that the elementary particles must be pointlike. Pointlike particles already lead in classical physics to infinite self-energies, in obvious contradiction to reality. Since no theory of reality can have an infinite vacuum energy, or much less pointlike particles (with infinite self-energies), present thinking is that gravity must ultimately provide a natural cutoff at the Planck length of $\sim 10^{-33}$ cm. This idea, however, leads to the absurd vacuum mass density of $10^{95}$ g/cm$^3$, corresponding to a density that would enable putting the mass of the entire universe in a cube of side length less than 1 fermi. Furthermore, with a classical electron radius of $\sim 10^{-13}$ cm, this would mean an overall structure of the electron of 20 orders of magnitude (like the ratio of $\sim 100$ light years to $\sim 1$ cm), again an absurd conclusion. Most recently, SU5 gauge theories, predicting a cut-off length at $\sim 10^{-30}$ cm, have been found wrong by the absence of the proton decay they predict. It is therefore more likely to think that the cut-off distance is much larger than the grand unification length of SU5 at $\sim 10^{-30}$ cm or the Planck length. A much larger cut-off length had also been demanded by Heisenberg a long time ago, but Heisenberg failed in introducing such a length into the laws of physics because he did not question relativity.

Generalizing the set of axioms used by Poincaré to *derive* special relativity, we have in combination with the fundamental laws of quantum mechanics succeeded in finding its natural nonlinear generalization. The proposed theory is an absolute theory. It not only leads to a solution of the infinite vacuum energy problem, but also gives elementary particles a finite mass. Only an "absolute" theory is expected to predict *absolute* values, like the self-energies. Special relativity that is a "relative" theory can only predict *relative* quantities, like the Lamb shift. It completely fails to determine absolute values, where it leads to nonsense. Our theory has both elements in it combined. It contains both relative and absolute velocities, but also special relativity as a singular limit. The existence of a preferred reference system is also supported by the highly isotropic cosmic microwave radiation, which can be understood as a small temperature of the quantum ether. No such isotropy has been observed with regard to the distribution of the galaxies.

As the singular, linear limit of a more general nonlinear theory, special relativity is certainly simpler than the nonlinear theory we propose, but that does not mean that special relativity must be closer to reality. The introduction of a fundamental length into the proposed nonlinear theory breaks the space–time symmetry of special relativity. This should be of no surprise. Our theory is an attempt to derive relativity from quantum-theoretical principles, but these principles, in the absence of a fundamental length, even deny the space–time symmetry of special relativity. In its most general representation, quantum mechanics can do away with position space, unlike time, which always remains a parameter.

The dogmatic adherence to the axioms of the special theory of relativity is not without historical precedent. In the Middle Ages, ecclesiastical doctrine dictated that the planetary orbits must be circles, a dogma even accepted by Copernicus. Circular orbits are mathematically simpler than elliptic orbits, but elliptic orbits and not circles ultimately turned out to be true. Likewise, pointlike particles are certainly simpler than extended particles. The circular orbits in the theory of Ptolemy therefore remind us of the pointlike structure of elementary particles in the theory of relativity. The Ptolemaic theory was nevertheless quite successful, because circles are a fairly good approximation for the elliptic planetary orbits in the solar system. We may similarly expect that special relativity, even though an extremely good approximation for low energies, could ultimately turn out to be false for very high energies. The infinities of quantum field theory, which are clearly nonsense, are the best argument against the immutable truth of special relativity.

## References

Bagge, E. (1962) *Naturwissenschaften* **23:** 1. *Atomkernenergie* **7:** 427.
Bohm, D. (1965) *The Special Theory of Relativity* (New York: W.A. Benjamin).
Builder, G. (1958) *Aust. J. Phys.* **11:** 457.
Einstein, A. (1905) *Ann. Phys.* **17:** 891.

Feynman, R. (1965) *The Character of Physical Law* (Boston: MIT Press), p. 168.

Heisenberg, W. (1938) *Ann. Phys.* **32:** 20.

Landau, L.D., and E.M. Lifshitz (1975) *The Classical Theory of Fields* (New York: Pergamon), pp. 43ff.

Lorentz, H.A. (1909) *The Theory of Electrons* (Leipzig: B.G. Teubner).

Pais, A. (1982) *"Subtle is the Lord", The Science and Life of Albert Einstein* (Oxford: Clarendon), p. 168.

Prokhovnik, S.J. (1967) *The Logic of Special Relativity* (Cambridge, England: Cambridge Univ. Press).

Poincaré, H. (1905) *C.R. Acad. Sci. Paris* **140:** 1504.

Sommerfeld, A. (1952) *Electrodynamics* (New York: Academic Press), pp. 301ff.

Winterberg, F. (1984) *Atomkernenergie/Kerntechnik* **44:** 88; **44:** 238.

# The Special Theory of Relativity in a Gravitational Field

**by Masayuki Note**

*Department of Physics*
*University of Missouri-Rolla*
*Rolla, Missouri 65401*

***Abstract***—A theory is presented that generalizes the special theory of relativity to an acceleration field. The theory is based on the reinterpretation of linear motion as an acceleration field and is applied to a gravitational field. It is shown that the theory describes the precession of Mercury's perihelion and the deflection of light near the Sun.

## Introduction

The general theory of relativity, which A. Einstein proposed in 1915, has fascinated physicists for many reasons. For example:

(1) The theory is able to explain the precessional motion of Mercury's perihelion and gives 43 seconds of precession per century, which is in good agreement with the observations.

(2) The theory predicts a deflection of light by the Sun, giving a deviation of 1.72 seconds of arc from a straight line. This is also observed.

(3) The theory exhibits mathematical elegance and beauty.

After the appearance of the theory of general relativity, however, some physicists made other attempts to describe the above observed effects. Some of them were merely variations of a description of gravity by a symmetric tensor field in flat space-time (Misner et al. 1973). These efforts are characterized by a value of 4/3 of the observed value for the perihelion precession, and any attempt to rectify this leads to a generalization of the space-time continuum to that of general relativity. It has been shown that within a four-dimensional framework only the nonlinear terms unique to the general theory of relativity could lead to the observed value for the perihelion precession (Duff 1974).

Based on the special theory of relativity, E. Bagge assumed that the relativistic increase of a moving mass should influence the gravitational force (Bagge 1981). Thus he modified the Newtonian potential to:

$$V = -\frac{GM}{r}\frac{m}{\sqrt{1-\beta^2}} \tag{1}$$

where $m$ = mass of Mercury, $M$ = mass of the Sun, $r$ = Mercury's distance from the Sun, $\beta = v/c$: Mercury's velocity in units of the velocity of light, and $G$ = gravitational constant.

He concluded that the potential given by Eq. (1) gives the motion of Mercury's perihelion as 42.087 seconds of arc per century, which is in good agreement with observations. But the theory does not consider the relativistic effect of the acceleration field, which was first studied by Einstein (Einstein 1911).

On the other hand, Einstein's general theory has difficulty with the so-called rotating-disk problem. This problem, first discussed by P. Ehrenfest in 1909 (Ehrenfest 1909), has caused disagreement among physicists because it seems that the theory violates a requirement of the theory of special relativity (Hill 1946; Rosen 1942; Groen 1975).

A good general theory of relativity must not only explain the two astronomical effects, but also give a relativistic interpretation of an acceleration field that does not conflict with Einstein's theory of special relativity. Such being the case, this paper first shows it is possible to generalize the special theory of relativity to an acceleration field. In this generalization it becomes clear that the mass in an acceleration field (gravitational field) changes as a function of position rather than velocity as in Bagge's theory. The proposed formulation, when it is applied to the orbit of Mercury and the deflection of light by the Sun, yields results that are in agreement with observations. Black holes are also considered.

# Theory

### Principle of Equivalence and the Gravitational Field

It was first recognized by Einstein that the red shift can be explained by using the principle of equivalence and the special theory of relativity (1911). Consider *two identically constructed clocks that are placed at* rest, a distance $h$ apart along the lines of force in a uniform gravitational field of acceleration $g$, as in Figure 1. In accordance with the principle of equivalence,

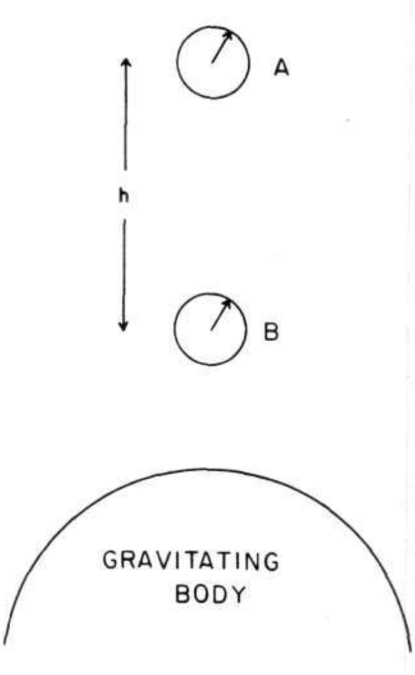



**Figure 1.** Two identically constructed clocks, $A$ and $B$, in a gravitational field.

any comparison of the periods of these clocks can be made as well in a gravitation-free region, in which they are accelerated upward with the acceleration $g$, as in Figure 2. To compare clocks $A$ and $B$, we consider a third, identically constructed clock, $C$, which is at rest. Since $C$ is at rest in a gravitation-free region, it makes a suitable standard for comparing $A$ and $B$ with each other.

Suppose that $A$ and $B$ have upward speeds of $v_a$ and $v_b$, respectively, when they pass $C$. If the period of $C$ is $T$, then, according to the theory of relativity, the periods of $A$ and $B$ that are seen by an observer on $C$ are,

$$T_a = T\left(1 - \frac{v_a^2}{c^2}\right)^{-\frac{1}{2}} \approx T\left(1 + \frac{v_a^2}{2c^2}\right) \tag{2}$$

$$T_b = T\left(1 - \frac{v_b^2}{c^2}\right)^{-\frac{1}{2}} \approx T\left(1 + \frac{v_b^2}{2c^2}\right) \tag{3}$$

respectively. Here we have assumed that the speed of light, $c$, is much greater than $v_a$ and $v_b$.

From Eq. (2) and Eq. (3), we obtain

$$\begin{aligned} T_b &\approx T_a\left\{1 + \frac{(v_b^2 - v_a^2)}{2c^2}\right\} \\ &= T_a\left(1 + \frac{gh}{c^2}\right) \end{aligned} \tag{4}$$

Thus we have found the red shift resulting from a uniform or nearly uniform acceleration field.

The above approach suggests a possible application of the special theory of relativity to a gravitational

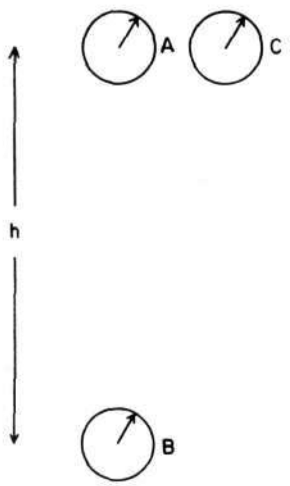



**Figure 2.** Accelerated clocks $A$ and $B$.

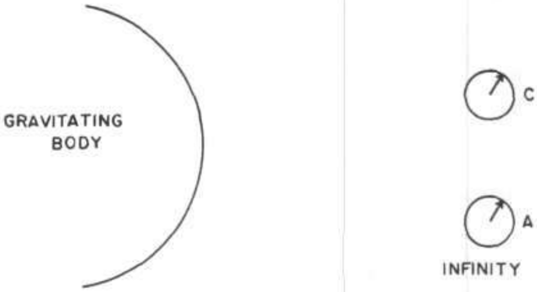



**Figure 3.** Clocks $A$ and $C$ infinitely far from the gravitating body.

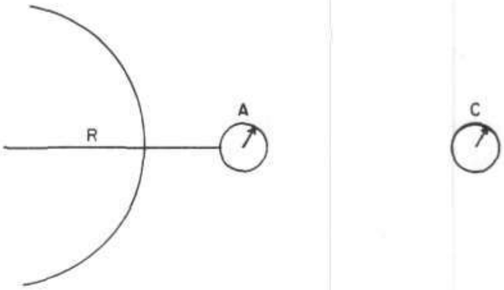



**Figure 4.** $A$ moves toward the gravitating body.


**Figure 5.** $A$ and $C$ in gravitation-free space.

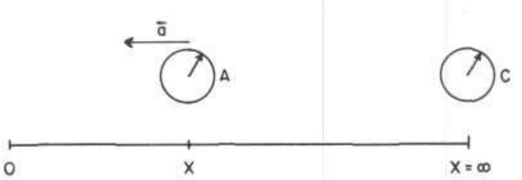



**Figure 6.** $A$ is accelerated. $C$ remains at infinity.

field. In fact, in the same paper in which Eq. (4) was derived, Einstein attempted to calculate the deflection of light. However, he obtained half the correct value with his approach. The same kind of comparison can be made for measuring mass and length. This implies that relativistic mass and length must also be taken into account in a gravitational field theory.

## Generalization of the Special Theory of Relativity to an Acceleration Field

In this section we consider a nonuniform gravitational field. Consider two identically constructed clocks, $C$ and $A$, which are placed in a gravitational field. If $C$ is placed infinitely far from the gravitating body, there is no gravitational effect on $C$; that is, $C$ is in a gravitation-free space. Therefore, $C$ makes a suitable standard for measuring time. We want to compare the period $T_a$ of $A$, with the period $T$ of $C$. But we cannot use Eq. (4) for this comparison, because our gravitational field is not uniform in this case. For this reason, we must find a new method to compare these two clocks on the basis of the principle of equivalence.

Suppose there are two identically constructed clocks, $A$ and $C$, infinitely far from a spherically symmetric gravitating body (Figure 3). If $A$ moves with an infinitesimal velocity toward the gravitating body along a radial line, it experiences an acceleration field. If the velocity is kept constant, the acceleration is a function of $r$, the distance from the center of the body, such that $a = a(r)$. At infinity, we assume $a(\infty) = 0$. (See Figure 4.)

Now, consider $A$ and $C$ in a gravitation-free space. Suppose $A$ and $C$ are placed infinitely far from the origin of a one-dimensional coordinate system (Figure 5). We also suppose $A$ and $C$ are at rest with respect to the coordinate system. If $A$ is accelerated toward the origin along the $x$-axis, the velocity of $A$ is given by, at $x = R$ (Figure 6)

$$v(R) = \sqrt{-2\phi(R)}. \tag{5}$$

The potential $\phi$ is defined here as

$$\begin{aligned} \phi(R) &= -\int_{\infty}^{R} a(x)dx \\ &= -\int_{\infty}^{R} \frac{dV}{dx}\frac{dX}{dt}\,dx \\ &= -\frac{1}{2}\,v(R)^2, \end{aligned} \tag{6}$$

where $a(x)$ is the acceleration of $A$ at $x$. Thus the factor

$$k = \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}, \tag{7}$$

which plays an important role in the special theory of relativity, becomes

$$k = \left(1 + \frac{2\phi}{c^2}\right)^{\frac{1}{2}} \tag{8}$$

For the clock $A$ in the gravitational field considered before, we can define $\phi$ as

$$\phi = -\int_{\infty}^{R} a(r)dr. \qquad (9)$$

Note that the clock in the gravitational field moved with a constant velocity, and the clock in the gravitation-free space moved with a varying velocity. However, both clocks have experienced a force (acceleration) field. Using the equivalence principle, the relativistic factor $k$ for a gravitational field can be defined:

$$k(r) = \left(1 + \frac{2\phi}{c^2}\right)^{\frac{1}{2}}. \qquad (10)$$

Thus

$$T = T_a \left(1 + \frac{2\phi}{c^2}\right)^{\frac{1}{2}} \qquad (11)$$

where $T$ is the period of $C$, which is infinitely far from the gravitating body, the $T_a$ is the period of $A$, which is a distance $r$ from the gravitating body.

A result similar to Eq. (10) has been derived by considering rotational motion (Pauli 1958; Adler et al. 1975). Let us take a reference system $K$, which rotates relative to the Galilean system $K_0$ with angular velocity $\omega$. A clock $A$, say, at rest in $K$ will then be slowed down. The time dilation is given by

$$\begin{aligned} T &= T_a \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}} \\ &= T_a \left(1 - \frac{\omega^2 r^2}{c^2}\right)^{\frac{1}{2}} \end{aligned} \qquad (12)$$

where $T$ is the period of clock $C$, which is at rest in the Galilean system $K_0$, and $r$ is the distance of $A$ from the rotating axis. But in $K$ a gravitational field (field of the centrifugal force) exists with the potential given by

$$\phi = -\frac{1}{2}\omega^2 r^2. \qquad (13)$$

Thus, the same result as before is obtained.

As an application of the above result, Eq. (10), consider clocks $A$ and $B$, which are placed at distances $r_a$ and $r_b$ from the gravitating body, respectively. Then

$$T = T_a \left(1 + \frac{2\phi(r_a)}{c^2}\right)^{\frac{1}{2}} \qquad (14)$$

$$T = T_b \left(1 + \frac{2\phi(r_b)}{c^2}\right)^{\frac{1}{2}}, \qquad (15)$$

where $T$ is the period of $C$, which is at infinity. Thus,

$$\begin{aligned} \frac{T_b}{T_a} &= \left[\frac{1 + \frac{2\phi(r_a)}{c^2}}{1 + \frac{2\phi(r_b)}{c^2}}\right]^{\frac{1}{2}} \\ &\approx 1 - \frac{1}{c^2}[\phi(r_b) - \phi(r_a)]. \end{aligned} \qquad (16)$$

If the gravitational field is uniform, it follows that

$$T_b = T_A \left(1 + \frac{gh}{c^2}\right) \qquad (17)$$

where $g$ is the acceleration, and $h = r_a - r_b$. This agrees with Eq. (4).

From now on, we shall consider the three-dimensional conservative acceleration field, that is, the path independent line integral

$$\phi(\mathbf{r}) = -\int^{\mathbf{r}} \mathbf{a} \cdot d\mathbf{r}. \qquad (18)$$

Here, $\mathbf{r}$ = position vector, $\mathbf{a}$ = acceleration, and $\phi(\mathbf{r})$ = potential at $\mathbf{r} = (x,y,z)$.

We naturally define a relativistic factor in the three-dimensional space as

$$k(\mathbf{r}) = \left(1 + \frac{2\phi(\mathbf{r})}{c^2}\right)^{\frac{1}{2}}. \qquad (19)$$

If a particle has a mass $m_0$ at infinity ($\phi = 0$), the mass depends on its position and it increases as it is placed in a space with negative potential $\phi(r)$:

$$m = m_0 \left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}}. \qquad (20)$$

Similarly, if the same nonsimultaneous events are seen by two observers, one in a space with $\phi = 0$, and another in a space with potential $\phi(r)$, and the time interval of the events measured in the $\phi = 0$ space is $\Delta t$, then the time interval $\Delta t_0$ measured in the $\phi$ space is related to $\Delta t$ by

$$\Delta t = \Delta t_0 \left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}}. \qquad (21)$$

It is obvious from our thought experiment that a rod that is laid along the field line appears shortened. Thus, an infinitesimal length $\Delta l_0$ in a space with $\phi = 0$ appears shortened if it is placed in a space with a potential $\phi(r)$ and it is parallel to the lines of force.

Thus, one has

$$\Delta l = \Delta l_0 \left(1 + \frac{2\phi}{c^2}\right)^{\frac{1}{2}} \tag{22}$$

It is obvious that $\Delta l = \Delta l_0$ if the rod is laid perpendicular to the field lines.

Finally, we shall find the total energy of a mass $m_0$. From Eq. (20), we have

$$\begin{aligned} E(\mathbf{r}) &= mc^2 \\ &= m_0c^2 \left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}} \end{aligned} \tag{23}$$

where $E(\mathbf{r})$ is the energy contained in the mass at $\mathbf{r} = (x,y,z)$. If the change in the energy content of $m_0$ is due to the field, the potential energy $V$ of the mass becomes

$$V = m_0c^2 - m_0c^2 \left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}} \tag{24}$$

Note that we call $V$ "potential energy" and $\phi$ "potential."

### Determination of $\phi$ for a Gravitational Field

The potential $\phi$ must be determined to apply the theory to a gravitational field. Naturally to lowest order we know that

$$\phi = -\frac{GM}{r} \tag{25}$$

Substituting Eq. (25) for $\phi$ in Eq. (24), however, gives a precession of Mercury's perihelion of only 28 seconds of arc per century. This is much smaller than the observed value of 43 seconds.*

Now, instead of taking the potential energy $V$ as

$$V = -\frac{GM\,m_0}{r}, \tag{26}$$

we assume that the first approximation of $V$ is given by

$$V = -\frac{GM\,m_0}{r}\left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}}. \tag{27}$$

In Eq. (27), we have used the relativistic mass. This assumption is reasonable, because if a mass is placed in the space, its mass increases as was shown in Eq. (20). Thus, we get for small $\phi/c^2$,

$$\begin{aligned} V &= m_0c^2 - m_0c^2 \left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}} \\ &= m_0\phi \left(1 - \frac{\phi}{c^2} + \frac{3}{2}\frac{\phi^2}{c^4} - \cdots\right) \end{aligned} \tag{24'}$$

This should be equal to the first approximation of $V$; namely,

$$\begin{aligned} V &= -\frac{GMm_0}{r}\left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}} \\ &= -\frac{GMm_0}{r}\left(1 - \frac{\phi}{c^2} + \frac{3}{2}\frac{\phi^2}{c^4} - \cdots\right) \end{aligned} \tag{27'}$$

for small $\phi/c^2$. We can estimate $\phi$ by equating the lowest-order terms in $\phi$ from the two expressions. Thus, we take

$$\phi = -\frac{GM}{r}\left(1 - \frac{\phi}{c^2}\right). \tag{28}$$

It follows that

$$\phi = -\frac{GM}{r} - \frac{1}{c^2}\left(\frac{GM}{r}\right)^2. \tag{29}$$

Finally, we obtain for Eq. (24)

$$\begin{aligned} V = {} & m_0c^2 \\ & - m_0c^2\left[1 + \frac{2}{c^2}\left\{-\frac{GM}{r} - \frac{1}{c^2}\left(\frac{GM}{r}\right)^2\right\}\right]^{-\frac{1}{2}} \end{aligned} \tag{30}$$

This gives a reasonable approximation for $V$ if $\phi/c^2$ is small.

Note that the function $\phi$ cannot be determined *uniquely within the proposed theoretical framework.* The first approximation for the potential energy as given in Eq. (27) is just a good guess as to the form of the potential energy. Further investigation is needed.

* The calculation is shown below.

# Applications

## Bending of a Light Ray in a Gravitational Field

The key to this problem is to realize that each point in a gravitational field corresponds to an inertial frame with a velocity $v = \sqrt{-2\phi}$. This "correspondence principle" implies that the velocity of light is constant at each point of the field if the velocity is measured locally.

In a gravitational field, let $c$ be the velocity of light locally measured, and let $c_0$ be the velocity of that same light as measured by an observer at infinity. An observer $B$ measuring the local velocity of light in the field finds that the light pulse travels $ds$ units of length during $dt = ds/c$ units of time. But an observer $A$ at infinity "knows" that $B$'s unit of time is longer than $A$'s. Thus the time interval that $A$ must use in plotting the path of the light pulse is

$$dt_0 = dt\left(1 + \frac{2\phi}{c^2}\right)^{-\frac{1}{2}}. \tag{31}$$

We can express $ds$ in terms of radial and tangential components, $dr$ and $dz$, as in Figure 7. Then $A$ "knows"

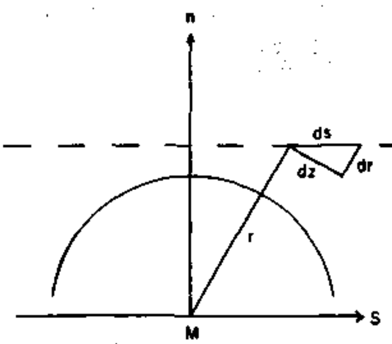



**Figure 7.** Coordinate system for $A$.

that $B$'s unit of radial distance is shorter than $A$'s, so $A$ must use the radial interval

$$dr_0 = dr\left(1 + \frac{2\phi}{c^2}\right)^{\frac{1}{2}}. \tag{32}$$

Since $dz_0 = dz$, we get

$$ds_0 = (dr_0^2 + dz_0^2)^{\frac{1}{2}}$$

$$= \left\{dr^2\left(1 + \frac{2\phi}{c^2}\right) + dz^2\right\}^{\frac{1}{2}}.$$

Using $ds^2 = dr^2 + dz^2$ yields

$$= ds\left\{1 + \frac{2\phi}{c^2}\left(\frac{dr}{ds}\right)^2\right\}^{\frac{1}{2}}$$
$$\approx ds\left(1 + \frac{\phi}{c^2}\frac{s^2}{r^2}\right). \tag{33}$$

We have made use of the relation $dr/ds = s/r$.

From Eq. (31) and Eq. (33) we get

$$C_0 = \frac{ds_0}{dt_0}$$

$$= \frac{ds}{dt}\left\{\frac{1 + \frac{\phi}{c^2}\left(\frac{s}{r}\right)^2}{1 - \frac{\phi}{c^2}}\right\}.$$

Recall that $\phi/c^2$ is small. Then this becomes

$$C_0 \approx C\left(1 + \frac{\phi}{c}\frac{s^2}{r^2} + \frac{\phi}{c^2}\right). \tag{34}$$

Thus, the velocity of light is a function of the potential (Schiff 1960). Note that the velocity of light $c_0$ in Eq. (34) is the value measured by an observer at infinity where the space is gravitation-free. Einstein was the first physicist who derived an equation for the velocity of light (but incorrectly), as a function of potential. His result was

$$C_0 = C\left(1 + \frac{\phi}{c^2}\right). \tag{35}$$

We now easily infer, by means of Huygens's principle, that a light ray propagated across a gravitational field undergoes deflection. If we calculate the angle positively when the ray is bent toward the side of decreasing $n$, the angle of deflection per unit of path of the light is (Einstein 1911)

$$\Delta\varphi = \frac{1}{c}\frac{\partial c_0}{\partial n}. \qquad (36)$$

Let $M$ be the mass of the heavenly body, and $D$ the distance of the ray from the center of the body. Then the deflection is given by

$$\varphi = \frac{1}{c}\int_{-\infty}^{\infty} \frac{\partial c_0}{\partial n}\, ds$$

$$\approx \frac{GM}{c^2}\int_{-\infty}^{\infty} \left[\frac{3S^2 n}{r^5} + \frac{n}{r^3}\right]_{n\,=\,D} dS \qquad (37)$$

$$= \frac{4GM}{c^2 D}$$

A ray of light going past the Sun would accordingly undergo a deflection of 1.7 seconds of arc, which is in good agreement with the observations.

If Eq. (35) is used for this calculation, only half the correct value is obtained.

## The Orbital Precession of Mercury

It will be shown that the orbit of a planet, with the potential energy given by Eq. (30), undergoes a perihelion precession in the direction of motion. In the case of Mercury, the result is 43 seconds of precession per century, which is in agreement with the observations.

In polar coordinates, the Lagrangian of a planet of mass $m_0$ has the form:

$$\mathcal{L} = m_0c^2 - m_0c^2\sqrt{1-\beta^2} - V(r) \qquad (38)$$

where $V$ is the potential energy, and $\beta^2 = (\dot{r}^2 + r^2\dot{\theta}^2)/c^2$. The dot represents differentation with respect to time. Substituting Eq. (30) for $V$ in Eq. (38), we have

$$\mathcal{L} = -m_0c^2\sqrt{1-\beta^2} + \frac{m_0c^2}{\sqrt{1+\alpha}}, \qquad (39)$$

where $\alpha = 2\phi/c^2$. Note that the kinetic energy term is totally independent of the potential term; that is, the potential term is based purely on the relativistic effect due to the acceleration field, and the kinetic term is due to the motion. It will be clear that the above Lagrangian gives a consistent energy equation.

Lagrange's equations thus become

$$\frac{d}{dt}\left(\frac{m_0\, r^2\dot{\theta}}{\sqrt{1-\beta^2}}\right) = 0 \qquad (40)$$

and

$$\frac{d}{dt}\left(\frac{m_0\,\dot{r}}{\sqrt{1-\beta^2}}\right) - \left(\frac{m_0 r\dot{\theta}}{\sqrt{1-\beta^2}} + \frac{\partial}{\partial r}\frac{m_0c^2}{\sqrt{1+\alpha}}\right) = 0. \qquad (41)$$

From Eq. (40) we obtain the law of conservation of angular momentum

$$\frac{m_0 r^2\dot{\theta}}{\sqrt{1-\beta^2}} = L = \text{constant}. \qquad (42)$$

Multiplying Eq. (40) by $\theta$ and Eq. (41) by $\dot{r}$ and adding, we find

$$\frac{m_0c^2}{\sqrt{1-\beta^2}} - \frac{m_0c^2}{\sqrt{1+\alpha}} = E = \text{constant}. \qquad (43)$$

This is the law of conservation of energy. The kinetic energy is totally separated from the potential term.

Introducing a new variable, $u = 1/r$, simple calculations yield

$$\frac{d\theta}{du} = \pm LC\left[E^2 + \frac{2E\,m_0c^2}{\sqrt{1+\alpha}} + \frac{m_0^2c^4}{1+\alpha} - m_0^2c^4 - L^2c^2u^2\right]^{-\frac{1}{2}}. \qquad (44)$$

The above equation cannot be integrated by elementary means.

Now consider the following series for small $GMu/c^2$:

$$\frac{1}{1+\alpha} = 1 + \frac{2GMu}{c^2} + \frac{6G^2M^2u^2}{c^4} + \left\{\text{terms higher than}\left(\frac{GMu}{c^2}\right)^2\right\} \qquad (45)$$

$$\frac{1}{\sqrt{1+\alpha}} = 1 + \frac{GMu}{c^2} + \frac{5G^2M^2u^2}{2c^4} + \left\{\text{terms higher than}\left(\frac{GMu}{c^2}\right)^2\right\}, \qquad (46)$$

where $M$ is the mass of the Sun. Taking the first three terms of the series, we have

$$\frac{d\theta}{du} = \pm LC\left[E^2 + 2Em_0c^2 + (2Ek + 2km_0c^2)u - \left(L^2C^2 - 6k - \frac{5Ek^2}{2m_0c^2}\right)u^2\right]^{-\frac{1}{2}} \quad (47)$$

where $k = GMm$. Thus the solution for $\theta$ is of the form:

$$\theta = P \arccos\frac{D_1 u + D_2}{D_3} + \theta_0 \quad (48)$$

where $P$, $D_1$, $D_2$, $D_3$, and $\theta_0$ are constants.

For some $E$, the orbit becomes elliptical, and the orbital ellipse undergoes a slow rotation by an amount $\Delta\theta = 2\pi P - 2\pi$:

$$\Delta\theta = 2\pi\left[1 - \frac{\left(6 - \frac{5E}{2m_0c^2}\right)k^2}{L^2C^2}\right]^{-\frac{1}{2}} - 2\pi. \quad (49)$$

Since $k^2 << L^2c^2$ and $|5E/2mc^2| << 6$, we have

$$\Delta\theta \approx 6\pi\left(\frac{k}{LC}\right)^2. \quad (50)$$

Let $v$ be the average speed of the planet, and $R$ be the average distance from the Sun to the planet. Then

$$\frac{k}{LC} \approx \frac{GM}{vRc}. \quad (51)$$

In the case of Mercury, $\Delta\theta = 43.02$ seconds of arc per century.

### Black Holes

We had

$$V = m_0c^2 - m_0c^2\left[1 + \frac{2}{c^2}\left\{-\frac{GM}{r} - \frac{1}{c^2}\left(\frac{GM}{r}\right)^2\right\}\right]^{-\frac{1}{2}}. \quad (52)$$

Since $V$ is an observable quantity, it cannot be an imaginary number. Thus, we demand that

$$1 + \frac{2}{c^2}\left\{-\frac{GM}{r} - \frac{1}{c^2}\left(\frac{GM}{r}\right)^2\right\} \geq 0$$

$$r^2 - \frac{2GM}{c^2}r - \frac{2G^2M^2}{c^4} \geq 0. \quad (53)$$

To solve this for the smallest $r$, set

$$r_{\min} = \frac{2GM}{c^2}(1 + \delta) \quad (54)$$

where $2GM/c^2$ is the Schwarzschild radius. Then we get

$$\delta^2 + \delta - \frac{1}{2} = 0. \quad (55)$$

Solving the above equation for $\delta$,

$$\delta = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}. \quad (56)$$

Since $r_{\min}$ cannot be negative, we have to take the upper sign for $\delta$. Therefore

$$r_{\min} = (1.366)\frac{2GM}{c^2}, \quad (57)$$

which is 36.6 percent larger than the Schwarzschild radius.

The above calculation for the radius of a black hole is an approximation, because we have assumed a point mass $M$ that is inside the black hole. In fact, we should consider the black hole as a system in which all the mass contributes to the energy of the field. Further investigation of this problem is needed.

# Conclusion

It has been shown that the three crucial tests of general relativity can be derived from the equivalence principle and special relativity without reference to the geodesic equation or the field equations. This attempt is not totally new for the red shift and the deflection of light (Schiff 1960). However, the influence of the field on the mass was considered for the first time in this paper (Parthasarathy).

The general theory of relativity does not agree with the theory presented here in some calculations, and the mathematical and physical concepts of the theories are different, indeed.

Einstein's theory of general relativity is based on Riemannian four-dimensional space-time geometry. In his theory, gravity is explained as a metric phenomenon. For a spherically symmetric gravitating body, the solution of the field equation is known as the Schwarzschild solution. The Schwarzschild line element in polar coordinates $r, \theta, \psi$ is given by

$$ds^2 = \left(1 - \frac{2GM}{c^2r}\right)(dx^0)^2 - \frac{dr^2}{1 - \frac{2GM}{c^2r}} - r^2\,(d\theta^2 + \sin^2\theta\, d\psi^2), \quad (58)$$

where $ds$ is the line element and $dx^0 = cdt$.

To find the motion of a particle in a gravitational field, one must minimize the line element by solving the variational problem:

$$\delta \int ds = 0. \quad (59)$$

Solving the above equation, Einstein successfully derived Eq. (50). Further, he assumed the trajectory of a light ray in the field is a null-geodesic line; that is, a zero line element obtained by setting $ds = 0$, and again, Eq. (37) was derived.

On the other hand, the proposed theory shows that the relativistic effects due to a gravitational field can be understood only by using the principle of special relativity and the principle of equivalence, both of which have been verified by experiments.

## References


R. Adler, et al. 1975. *Introduction to General Relativity*. New York: McGraw-Hill, p. 137.
E. Bagge. 1981. *Atomkernenergie Kerntechnik* **39:** 223.
M. Duff. 1974. *Gen. Rel. Grav.* **5:** 441.
P. Ehrenfest. 1909. *Phys. Zeitschr.* **23:** 918.
A. Einstein. 1911. *Ann. der Physik.* **35:** 898.
O. Groen. 1975. *Am. J. Phys.* **43:** 869.
E. Hill. 1946. *Phys. Rev.* **69:** 488.
C. Misner, K. Thorne, and J. Wheeler. 1973. *Gravitation*. San Francisco: W. H. Freeman and Co., pp. 180–181.
R. Parthasarathy. Private communication.
W. Pauli. 1958. *Theory of Relativity*. New York: Dover, p. 151.
N. Rosen. 1942. *Phys. Rev.* **71:** 54.
L. Schiff. 1960. *Am. J. Phys.* **28:** 340.



The author's present address is 37-4 2-chome Nakadai, Itabashi-ku, Tokyo 174, Japan.


# Acknowledgements


I would like to acknowledge the support and suggestions of my advisor, Dr. Jerry L. Peacher, as well as Dr. Jack L. Rivers, Dr. Erich R. Bagge of West Germany, Dr. E.C. Sudarshan of the University of Texas-Austin, and Dr. R. Parthasarathy of the Institute of Mathematical Sciences, Madras, India.

# Theory and Applications of the Nonexistence of Simple Toroidal Plasma Equilibrium[1]

by Harold Grad

*Courant Institute of Mathematical Sciences*
*New York University*
*251 Mercer Street*
*New York, NY 10012*

***Abstract***—The basis of the theory of electrically conducting plasma in pressure balance with a magnetic field is a mathematically *composite* system [curl $\mathbf{B} \times \mathbf{B} = \nabla p$, div $\mathbf{B} = 0$] with both real and imaginary characteristics. In 1967, it was pointed out that in a hypothetical equilibrium of nested toroidal pressure surfaces, the presence of real characteristics leads to contradiction except in certain specific symmetric cases. The real characteristics represent the propagation along magnetic field lines of a number of physical properties. Thus the discovered lack of pressure balance is not a special consequence of one idealized model but carries over broadly to a variety of physical refinements and generalizations. Indeed, the mathematical result has profound physical consequences which have gradually entered the field in the context of rational rotation number resonances and multiple helicities, island formation, and turbulence.

## Introduction

A not uncommon phenomenon in science is the rapid transition, without apparent cause, from one state in which a concept is almost universally disbelieved and rejected to another state in which it is so transparent as to require no comment and exhibit no visible history. Sometimes there may be a transitional period during which both states coexist simultaneously (and schizoidally) in the scientific community. It is my view that the subject of this paper is either in the course of such a transition or close to it (but in which state it lies, preponderantly, I am not willing to guess).

Much of the material discussed here has been published in scattered form, some in physics journals, some in mathematics journals, some in not easily accessed conference proceedings and reports. The point of summarizing and unifying this material here is that, from its inception in Grad 1967b, this has been one of the most misunderstood topics in plasma physics, although it has achieved a position of great significance in helping to understand plasma confinement, both qualitatively and quantitatively. The main physical and mathematical consequence is that a toroidal sheared plasma cannot sit still.

One reason for the misunderstanding is that within part of the physics community, there is a taboo on the term "nonexistence." This requires the necessary manifestations of nonexistence to resurface in the form of more acceptable terminology (preserving mathematical divergences, singularities, resonances, islands, and so on) but with the logical implication of lack of existence arising from dense rational surfaces excised and simply not faced. It is frequently true that the statement by a mathematician that the solution to a physical problem does not exist is (a) irrelevant or (b) easily remedied by slightly altering the mathematical formulation. In the present case, however, the cited nonexistence of solutions (1) has far-reaching *physical* consequences and (2) is *not* remedied by altering the physical model (It is inherent to the physical phenomenon, and is mathematically structurally stable).

Another reason (and one that is psychologically very appealing) for rejecting the entire body of nonexistence theory is that it requires a total reformulation of the meaning of stability (Classically one requires an equilibrium or steady state as a starting point in order to observe the effects of a perturbation). A number of ways of avoiding this dilemma have emerged, but all these are more difficult than a start from a true steady state (Grad 1970, 1969).

From the physical viewpoint, the microscopic (*guiding center*) model offers the simplest interpretation. Mathematically, the macroscopic (scalar pressure MHD) model is the most natural. We shall attempt to maintain the connections between the two points of view while shuttling back and forth between them. As originally predicted (Grad 1967b), physical consequences range from loss of *microequilibrium* (a term that never took hold) with consequent fluctuations or "turbulence," to much larger scale motions such as the plasma turning itself inside out or moving to the wall. For these phenomena, entirely different from those in a particle accelerator with field errors, the relevant fields are usually produced by plasma currents, and the essential, nonremovable time dependence is that of the plasma, even when the field exhibits no pathology at all.

It is historically illuminating to view this theory as a direct continuation of the early discovery by Lyman Spitzer that equilibria did not exist in toroidal configurations that were too simple. The resolution, crudely put, was to correct the different (and unconfined) drift paths of electrons and ions by suitably twisting the global magnetic configuration to cancel the drifts. The further discovery by the author was that the stellarator solution to this problem was only valid in some overall average sense, and *no* field manipulation could correct this imbalance of particle drifts *locally* (Grad 1967b).

The original investigation in 1967 and following work (Grad 1969, 1970) concentrated on a *positive* approach, not to destroy but to salvage the large body of equilibrium and stability literature by appropriate *interpretation* of approximations to nonexistent equilibria. First and foremost, these studies point out that the phenomenon of nonexistence is frequently invisible because of the very nature of the most commonly used expansions and approximations; the studies then proceed to explore numerical and analytical means for studying "quasi"-equilibria, which, to some approximation, can take the place of traditional exact time-independent solutions of appropriately chosen model equations. One important approach, analytic and numerical as well as conceptual, is to *control the magnitude of the visible pathology*, from invisibility (to allow "equilibria" to be found) to very high visibility (for diagnostic purposes). An analogy is with an asymptotic series for which there is an optimum number of terms. Analytical tools must be chosen so they are not too precise; the same is true of numerical meshes.

## Characteristics: Qualitative Expectations

The mathematical source of all the phenomena under discussion is the presence in a toroidal domain of (degenerate) real characteristics that do not impinge upon the boundary; these characteristics, therefore, have no direct way of exchanging information with the outside world.

To begin, take the simplest example that starts with a harmonic field,

$$\text{curl } \mathbf{B} = 0, \quad \text{div } \mathbf{B} = 0. \tag{1}$$

The solution is not only smooth, it is analytic; and with homogeneous boundary condition, $B_n = 0$, it is uniquely determined in a torus by a single period, the transverse flux, $\int \mathbf{B} \cdot dS$, or axial *mmf*, $\int \mathbf{B} \cdot dx$. With a reasonably smooth boundary, there is no pathology no matter how contorted the domain may be.

The question of flux surfaces introduces a real characteristic in the third order *composite* system,

$$\text{curl } \mathbf{B} = 0, \quad \text{div } \mathbf{B} = 0, \quad \mathbf{B} \cdot \nabla\psi = 0. \tag{2}$$

Although the system in Eq. (2) is nonlinear and coupled, its form allows **B** to be found first, then $\psi$ as the solution of a first order linear equation that has the magnetic lines as characteristics. The problem is to assign a value of $\psi$ globally to "each" magnetic line. The pathology of the domain of dependence of $\psi$ for a general toroidal domain is long known; it is the relatively "good" behavior on sets of positive measure that is the more recent contribution of KAM theory. We need only recall that $\psi$ (insofar as it can be defined) takes constant values on surfaces of an infinite variety of different topologies, also (primarily from numerical evidence) on an infinite number of distinct ergodic regions, and more.

The equations governing the static pressure balance between an isotropic conducting fluid and a magnetic field are similar to Eq. (2) in structure, except for loss of the simplifying sequential separation of **B** and $\psi$ in Eq. (2).

$$\text{curl } \mathbf{B} \times \mathbf{B} = \nabla p, \quad \text{div } \mathbf{B} = 0. \tag{3}$$

The characteristics become visible upon introducing the current potential, $\zeta$, introduced into the plasma physics context by the author in 1954,

$$\mathbf{J} \equiv \text{curl } \mathbf{B} = \nabla\zeta \times \nabla p. \tag{4}$$

The system Eq. (3) is formally identical to

$$\begin{aligned} &\text{curl } \mathbf{B} = \nabla\zeta \times \nabla p, \quad && \mathbf{B} \cdot p = 0 \\ &\text{div } \mathbf{B} = 0, && \mathbf{B} \cdot \nabla\zeta = 1. \end{aligned} \tag{5}$$

It is convenient to consider the left pair as an elliptic system for **B**, given $\zeta$ and $p$, and the right pair as (degenerate) hyperbolic for $p$ and $\zeta$, given **B**. In an open domain (Figure 1) the obvious iteration between **B** and $(p, \zeta)$ suggested by this separation of Eq. (5) can be shown to converge in simple cases (Schechter

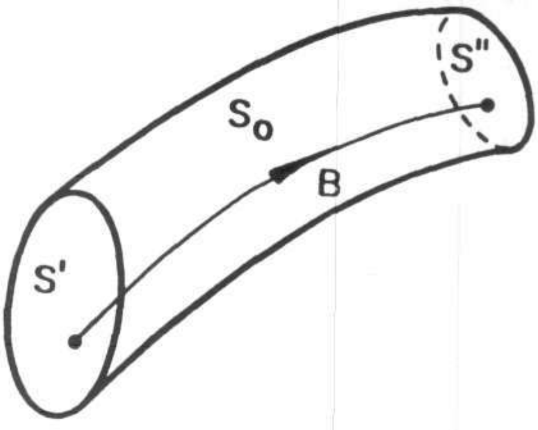



**Figure 1.** Tubular domain

1959; Bineau 1972; and Lortz 1970); for example, in two dimensions the iteration reduces to $\Delta\psi_{n+1} = f(\psi_n)$. Many qualitative and quantitative implications with regard to well-posed problems suggested by observing characteristics have emerged. A help in assessing qualitative properties of solutions is that within a $p$-surface the vector field **B** is a *surface harmonic;* it satisfies a two-dimensional elliptic equation on each surface and, in a toroidal domain, **B** is uniquely determined by two periods (Grad and Rubin 1958). Globally, the solution to Eq. (5) in a toroidal domain is therefore uniquely determined by the assignment of two *profiles;* for example, $\psi_1(p), \psi_2(p)$, the two fluxes given as functions of $p$ (Grad and Rubin 1958).

In a toroidal domain with boundary condition $\mathbf{B}_n = 0$, both $p$ and $\zeta$ are carried along magnetic lines that constitute the domain of dependence of $(p, \zeta)$ given **B**. Because of the essential nonlinear coupling between **B** and $\zeta$, no immediate rigorous conclusion can be drawn. Even so, the mathematical structure leads strongly to the presumption of pathological behavior of toroidal solutions, if they exist at all. Note that any pathology of $p$ and $\zeta$ is carried over to **J**, and therefore to **B**, in contrast to Eq. (2), where **B** is smooth despite the pathology of $\psi$.

There are exactly four known types of symmetry in which it is immediately evident that the pathological conclusion does not follow, and in each of these four symmetries existence theorems have been proved. In three cases (two dimensions, axial symmetry, helical symmetry), a flux function $\psi$ exists by virtue of the essential two-dimensionality, and the real characteristics integrate out [with two "arbitrary integration constants" $f_1(\psi)$, $f_2(\psi)$] to yield a second order elliptic equation for $\psi$. The theory in these cases, if not complete (the elliptic equation is nonlinear), is at least very familiar. The fourth case, symmetry under reflection in a plane, implies closed magnetic lines, and also admits certain existence results (Lortz 1970). A fifth exception is with anisotropic pressure, again with a classical existence theorem. No additional exceptions have arisen since 1967, when it was conjectured that toroidal existence (certainly existence of smooth solutions) with simple nested $p$-surfaces admits only these five exceptions [the anisotropic example reduces to the harmonic field Eq. (1) when it is isotropic].

It is virtually impossible to eliminate the possibility (which also has no practical interest) that for some isolated special domain and choice of constraints a solution may exist (with no others in its neighborhood). The proper formulation of the nonexistence statement is that, other than the stated symmetric exceptions, there are no *families* of solutions depending smoothly on a parameter. Three specific examples of such nonexistence theorems are given below, and the possibility of weak solutions will also be discussed. Here we continue with the qualitative implications of real characteristics in a torus for existence of toroidal solutions.

The pathology associated with flux surfaces in Eq. (2) is associated with the real characteristic $\mathbf{B} \cdot \nabla\psi = 0$. Similarly we expect pathological behavior associated with both pressure surfaces, $\mathbf{B} \cdot \nabla p = 0$, and current potential, $\mathbf{B} \cdot \nabla\zeta = 1$, in Eq. (5). More specifically, taking **B** as given, there is a difficulty in computing $\zeta$ from $\mathbf{B} \cdot \nabla\zeta = 1$. We recall that $\zeta$ is a current potential $\int \mathbf{J} \cdot d\mathbf{S} = \int d\zeta dp$. In an open-ended (nontoroidal) system (Figure 1), a nonconducting boundary condition $J_n = 0$ at the ends implies that $\int_{S'}^{S''} d\zeta = f(p)$. In a toroidal system with closed lines, $\oint d\zeta = f(p)$ also. Qualitatively, the freedom of closed magnetic lines to move relative to one another allows them to readjust so that $\oint dl/B = \text{const}$ may correspond to $p = \text{const}$. On the other hand, in a general toroidal system with *shear* and nested magnetic surfaces (containing surfaces with irrational rotation number which have ergodic field lines, and interleaved dense surfaces with rational rotation number which have closed lines), the dense set of closed lines is more rigid and less susceptible to readjustment. The condition in a sheared equilibrium, that $\oint d\zeta = \oint dl/B = \text{const}$ for all closed lines on a rational surface is evident when there is appropriate symmetry, and is clearly unlikely otherwise. This condition will be seen to be related to particle orbits (see concluding section) through its equivalence to the condition that **J** be single-valued across a cut to prevent buildup of electric charge

$$[\mathbf{J}] = 0 = \nabla[\zeta] \times \nabla p = \nabla(\textstyle\int dl/B) \times \nabla p$$
$$\text{or } \textstyle\int dl/B = f(p).$$

It may seem mathematically illogical to discuss separately two mechanisms for nonexistence when one is enough to eliminate all useful discussion. We shall see, however, that there is a constructive approach emphasizing approximate resolution of the physical problem in which case one of the difficulties may dominate practically or might be taken care of separately from the other.

The limiting case of a *force-free field*, $p = 0$, or $p \to 0$, is of particular interest since many astrophysical applications fall into this category (as do tokamaks, to a good approximation). Taking $p = 0$,

$$\text{curl } \mathbf{B} \times \mathbf{B} = 0, \ \text{div } \mathbf{B} = 0 \tag{6}$$

can be seen to be a composite system with a single real characteristic. The alternative formulation

$$\begin{gathered}\text{curl } \mathbf{B} = \lambda \mathbf{B}, \quad \mathbf{B} \cdot \nabla \lambda = 0 \\ \text{div } \mathbf{B} = 0,\end{gathered} \tag{7}$$

reveals the single real characteristic and the pathology in a torus as being related to existence of flux surfaces. The second condition, that $\int dl/B$ be constant on rational surfaces does not appear; it affects only the component of current perpendicular to **B**, which does not enter.

On the other hand, to consider the *limit* $p \to 0$, introduce into Eq. (5) the current potential $\xi$ given by

$$\mathbf{J} = \nabla \xi \times \nabla \psi \tag{8}$$

to obtain

$$\begin{aligned}\text{curl } \mathbf{B} &= \nabla \xi \times \nabla \psi, & \mathbf{B} \cdot \nabla \psi &= 0 \\ \text{div } \mathbf{B} &= 0, & \mathbf{B} \cdot \nabla \xi &= p'(\psi)\end{aligned} \tag{9}$$

The multivalued potential $\xi$ can be expressed in terms of the potential $\zeta$ of Eq. (5). (See the next section, "An Ergodic Lemma.")

$$\xi = \zeta p'(\psi) + \lambda(\psi)\theta \tag{10}$$

where $\theta$, a magnetic field potential ($\theta$ is a solution of the homogeneous equation $\mathbf{B} \cdot \nabla \theta = 0$) satisfies

$$\mathbf{B} = \nabla \theta \times \nabla \psi \tag{11}$$

We see now that a force-free field that is the *limit of scalar pressure equilibria* must satisfy the additional constraint involving $\int dl/B$; this special case of force free fields has two real characteristics.

More generally, for limiting cases of very weak *anisotropic* pressure equilibria, entirely different constraints are imposed on the force-free field limit. The limit is therefore nonuniform and is not uniquely determined without a knowledge of the properties of the plasma pressure which is no longer there![2]

The special case of a force-free field in which $\lambda$ is postulated to be a constant has received much attention (Chandrasekhar 1956; Taylor 1974). The legitimate reason to do so is that the equation is purely elliptic and linear and therefore allows easy solution. The various theoretical rationalizations and experimental justifications are all without merit.

## An Ergodic Lemma and Its Applications[3]

Consider the real function with period normalized to 1

$$f(\theta) = \sum_r a_r \exp(2\pi i r\theta), \ a_{-r} = \bar{a}_r, \ 0<\theta<1 \tag{12}$$

and mean value zero

$$\int_0^1 f(\theta)d\theta = 0, \ a_0 = 0.$$

It is a classical result that, if $f$ is continuous,

$$\lim \frac{1}{n} \sum_{r=0}^{n-1} f(\theta + r\alpha) = 0 \tag{13}$$

for all irrational values of $\alpha$. It is evident that the limit is not necessarily zero for rational $\alpha$ (the translations of $f$ are periodic). A refinement of this ergodic result is to ask for what values of $\alpha$ the partial sums

$$f^{(n)}(\theta;\alpha) = \sum_{r=0}^{n-1} f(\theta + r\alpha) \tag{14}$$

remain uniformly bounded for large $n$. The answer is contained in the trigonometric identity

$$f^{(n)}(\theta;\alpha) = \sum_r a_r \exp(2\pi i r\theta) \frac{\exp(2\pi i r n\alpha) - 1}{\exp(2\pi i r\alpha) - 1} \tag{15}$$

which can be rearranged as

$$f^{(n)}(\theta) = g(\theta + n\alpha) - g(\theta) \tag{16}$$

provided that

$$g(\theta) = \sum_r a_r \frac{\exp(2\pi i r\theta)}{\exp(2\pi i r\alpha) - 1} \tag{17}$$

converges. The most striking feature is that the sum $f^{(n)}$ depends only on the difference in the values of $g$ at the endpoints of the sum.

If $f$ is sufficiently smooth then $g$ will also be smooth (to a lesser degree) on a set of $\alpha$ of positive measure that includes somewhat less than the irrationals; specifically, at all points of a Cantor set that excludes intervals covering the rationals. There are many explicit theorems. As an example, if, for large $r$,

$$r^3|a_r| \to 0 \tag{18}$$

and for some $\varepsilon>0$ and all $(m,n)$

$$|\alpha - m/n| > \varepsilon/n^3 \tag{19}$$

then Eq. (17) will converge to a continuous function g(θ). The proof is elementary, and follows from the fact that the denominators in Eq. (17) can be small only for large $r$ at which $a_r$ is small. If $\varepsilon$ is small, the "good" set consisting of $\alpha$ for which g(θ) exists approaches measure one.

A much more sophisticated use of this type of argument gives the proof of KAM-type results (Moser 1962). But the elementary example quoted here is sufficiently sophisticated to exhibit part of the essential pathology of KAM theory, in particular the need to distinguish "approximate" rationals from "approximate" irrationals (defined by the specific covering of the rationals).

In the more general case in which the mean value of $f$ is not zero, $\int f d\theta = a_0$, the result is trivially modified (for simplicity, take $\alpha < 1$),

$$f^{(n)}(\theta) = na_0 + g(\theta + n\alpha) - g(\theta) \qquad (20)$$

The next step is to apply this ergodic refinement of Eq. (13) to solve the magnetic differential equation

$$\mathbf{B} \cdot \nabla u = v \qquad (21)$$

for a torus (that is, on a toroidal surface or in a domain covered by nested tori). The direction field **B** and inhomogeneous term $v$ are given. Take a cut $C_1$ the short way around the torus $S$ (Figure 2). Let $\alpha$ be the rotation

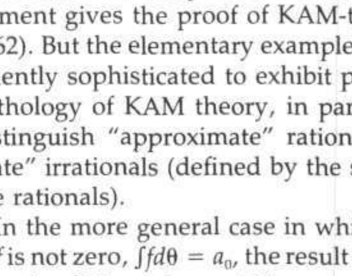



**Figure 2.** Integration along a field line

number of **B** on $S$. Let $A$ be the segment of a magnetic line originating at the point $\theta$ on $C_1$ and proceeding once around until it intersects $C_1$ at $\theta + \alpha$ (recall that $\mathbf{B} = \nabla\theta \times \nabla\psi$ and $\theta$ is the magnetic coordinate conjugate to $\psi$). Write the increment of $u$ along $A$ from $\theta$ to $\theta + \alpha$ as

$$\Delta u = \int_A v ds/B = f(\theta). \qquad (22)$$

If

$$\int_S v d\theta ds/B = 0 \qquad (23)$$

then

$$\int_0^1 f(\theta) d\theta = 0 \qquad (24)$$

and we can use the first form of the lemma to construct a function g(θ) from which

$$u(P) = g(\theta) + \int_A v ds/B \qquad (25)$$

where the integral is extended from the point $P$ along $A$ backward to the cut $C_1$ (Figure 2). The solution $u$ is single-valued and unique within an added constant (and the integral can be extended back to intersect $C_1$ any number of times).

All the above is predicated on a "good" value of $\alpha$ for which $g(\theta)$ can be obtained as in Eq. (17) from $f(\theta)$. Taking a family of surfaces $\psi$ = const on which $\alpha'(\psi) \neq 0$, we can solve for $u(P,\alpha)$ for a set of positive measure of $\alpha$ ($\alpha$ is used to parametrize $\psi$).

If the restrictions Eq. (23) and Eq. (24) are not made, define

$$q(\psi) = \int_S d\theta ds/B = \int_S dS/|\nabla\psi| \qquad (26)$$

integrating over the entire surface $S$ of the torus, and form the mean value

$$\langle v \rangle = (1/q) \int_S v d\theta \, ds/B. \qquad (27)$$

The solution $u$ will now not be single valued, but we can require that $\nabla u$ (restricted to the toroidal surface) be single valued so that the periods

$$\int_{C_i} \nabla u \cdot dx = [u]_i \qquad (28)$$

are path independent ($C_2$ is a cut the long way round the torus). It is easy to see that for "good" $\alpha$, a multivalued solution for $u$ exists which is unique provided that one of the periods $[u]_i$ is fixed (Grad 1967b, see Appendix). Alternatively, if $u_0$ is any special solution, the general solution is

$$u(P) = u_0(P) + a\theta(P) + b. \qquad (29)$$

The solution to the general inhomogeneous equation Eq. (21) can be expressed in terms of the solution of the special equation

$$\mathbf{B} \cdot \nabla\zeta = 1 \qquad (30)$$

by noting that $u$ (as in Eq. (21)) can be written

$$u = \hat{u} + v_0\zeta \qquad (31)$$

where $v_0 = \langle v \rangle$ and $\hat{u}$ is the single-valued solution to

$$\mathbf{B} \cdot \nabla \hat{u} = v - v_0. \qquad (32)$$

The periods of $u$ are not arbitrary (only one can be specified) and they are related by

$$\alpha[\zeta]_1 + [\zeta]_2 = 0, \text{ or}$$
$$\alpha[u]_1 + [u]_2 = v_0 q = \int v \, dS/|\nabla\psi|.$$

The mathematical distinction between rationals and irrationals or the separation out of "good" irrationals (excluding the rational covering) in the above theory or in the more sophisticated KAM theory requires interpretation before it can be considered to be relevant to any physical problem. Mathematical terms are used that demand arbitrarily precise distinctions that can have no physical meaning, per se. The mathematical pathology arises from the compression into a single expression, as in the Moser twist mapping theorem, of what are, in principle, an *infinite number of physically useful results*. It is conventional in mathematics (pure or applied) to simplify a problem in which a large parameter occurs by allowing the parameter to approach infinity. Unfortunately, this process is not always a simplification. To be specific, consider the confinement of a single particle in an accelerator or storage ring. The practical question is whether the particle is confined within a certain error for a specific number, $n$, of circuits. The mathematical theory says there is one answer if $n = 10^4$ (say, three important resonances), another answer if $n = 10^6$, still another if $n = 10^8$, and so on. There is no simplification, only complication, in the physical and mathematical problems as $n$ becomes larger. The mathematically pathological theorem results from the possibility of combining all these results in a single mathematical statement applying to $n = \infty$.

## Explicit Examples of Nonexistence

### *(A) The Vacuum Field Limit*

Consider a toroidal domain with a smooth, perfectly conducting boundary. We *postulate* the existence of a smooth family of equilibria with nested $p$-surfaces, $\mathbf{J} \times \mathbf{B} = \nabla p$, where

$$p = \varepsilon\, \hat{p}(\psi), \quad I = \varepsilon \hat{I}(\psi). \qquad (33)$$

The two profiles are $p$ and $I$ (toroidal current) given as functions of the toroidal flux $\psi$ and the parameter $\varepsilon$. It is easily shown that curl $\mathbf{B} \to 0$ uniformly in $\varepsilon$; further, the unique limiting vacuum field curl $\mathbf{B} = 0$ has nested flux surfaces $\psi$ = const (we postulate whatever smoothness this requires of the family of solutions). This gives a contradiction for many (most) domains, since the vacuum field properties are known and usually do not have nested surfaces.

This counterexample is extremely important, since *most equilibrium expansions in the literature take* $\beta = p/\mathbf{B}^2$ as an expansion parameter and expand smoothly in the neighborhood of a vacuum.

### *(B) The Field-Excluded Domain with Free Boundary*[4]

Consider a toroidal domain with a perfectly conducting boundary enclosing a toroidal plasma with $p$ = const, $\mathbf{B} = 0$; the region intervening between the plasma boundary and the fixed walls is a vacuum with $p = 0$, curl $\mathbf{B} = 0$. The interface supports the boundary condition $\mathbf{B}_n = 0$ and the classical free boundary condition $|\mathbf{B}|$ = const. In two dimensions, this problem is classical and admits many methods of solution, including conformal mapping (Berkowitz et al. 1958; Kadish and Stevens 1974; and Gilbarg 1960). One method is that of the inverse problem, to start with a *given*, analytic free boundary, leaving the external fixed boundary to be determined. The method is to use the Cauchy-Kovalevsky theorem to expand the analytic solution to the harmonic problem in a neighborhood, taking $\psi = 0$, $\partial\psi/\partial n = 1$ as Cauchy initial data. To implement this method in three dimensions requires use of the potential, $\mathbf{B} = \nabla\phi$, $\Delta\phi = 0$, since there is no flux function. The boundary (Cauchy initial) conditions are:

$$\text{i) } |\nabla\phi| = 1, \quad \text{ii) } \partial\phi/\partial n = 0. \qquad (34)$$

If such analytic $\phi$ can be found on the initial manifold, then we are ready to use Cauchy-Kovalevsky to solve for $\mathbf{B} = \nabla\phi$ in a neighborhood.

The structure of Eq. (34) on a toroidal surface is similar to a Hamilton-Jacobi system. Starting from a cross-cut, we follow wave fronts, $\phi$ = const and hope that they return to the original wave front after passing around the torus. Fortunately this can be recognized as a previously solved problem. On a free boundary, $|\mathbf{B}| = 1$, the magnetic lines are geodesics.[5] The problem of fields of geodesics smoothly covering a torus reduces to the Moser twist mapping theorem (KAM theory) (Arnold and Avez 1968). The annulus being mapped has as angle coordinate an angle around the cross-cut on the torus; the radial coordinate in the mapped annulus represents the initial angle that the geodesic makes with the cross-cut. The theorem states that if we start with a configuration known to have a complete set of geodesics (that is, a smooth direction field for each value $\alpha$ of the geodesic rotation number), then in a perturbed neighborhood of the original toroidal surface there will be geodesic fields for a set of $\alpha$ of positive measure, excluding neighborhoods of rational $\alpha$. In the positive (irrational) case, an admissible set of initial data Eq. (34) is supplied for use in Cauchy-Kovalevsky. A second use of the twist mapping theorem gives a set of positive measure, near the free boundary, of locations for the fixed boundary with $\mathbf{B}_n = 0$.

The result is more complete than promised (for the inverse problem). It gives positive results (existence) as well as negative (nonexistence). But no standard methods of perturbational or numerical analysis can make use of the pathological result.

The obvious conjecture for the direct problem would be to expect a pathological set of plasma free boundaries of positive measure. In three dimensions we cannot expect a simple continuum of free boundary solutions as in two dimensions.

### *(C) The Screw Pinch: An Incomplete Example*[6]

The two previous examples were somewhat abstract. The present example is strictly computational, Fourier series plus expansion in a parameter, and the results are essentially explicit. The interpretation of the results is not obvious. The linear expansion is inherently unreliable; it does not easily distinguish pathology (nonexistence) from the mere inadequacy of linear representation for an essential nonlinearity such as a helical island. The appearance of a singularity in a linear solution, in principle, casts doubt on the linearization only.

The screw pinch geometry is a straight circular cylinder with two ignorable coordinates and therefore allows explicit expansion of a linear perturbation in Fourier series. Let $(r,\theta,z)$ be cylindrical coordinates, with $\mathbf{B} = (B_r, B_\theta, B_z)$ dependent on $r$ only. From div $\mathbf{B} = 0$, $B_r = 0$. The pressure balance Eq. 3 takes the form

$$B_z B_z' + B_\theta B_\theta' + B_\theta^2/r + p' = 0 \qquad (35)$$

$(' = \partial/\partial r)$. Given two of the three profiles of $B_\theta, B_z, p$ the third is essentially determined.

In a linear perturbation,

$$\mathbf{B} = \mathbf{B}_0(r) + \varepsilon \mathbf{B}_1, \quad p = p_0(r) + \varepsilon p_1,$$
$$\hat{p} = p + \tfrac{1}{2}B^2 = \hat{p}_0(r) + \varepsilon \hat{p}_1. \qquad (36)$$

Here $\mathbf{B}_1$ (three components) and $p_1$ take the form

$$A_1 = A(r) \exp(im\theta + ikz). \qquad (37)$$

The Fourier modes are left uncoupled (depending, of course, on boundary conditions). After elimination (Appendix) a pair of coupled first-order ordinary differential equations in $r$ for $\hat{p}_1$ and $\mathbf{B}_1{}^r$ results, with each first derivative multiplied by the factor

$$\mathbf{k} \cdot \mathbf{B}^0 = \frac{m}{r} B_\theta{}^0 + k B_z{}^0. \qquad (38)$$

A singularity (*resonance*) occurs at a radius where $\mathbf{k} \cdot \mathbf{B}^0(r) = 0$. For the continuous spectrum to extend to the origin, making a linear perturbation singular, it is necessary and sufficient that $\mathbf{k} \cdot \mathbf{B} = 0$. Both $\hat{p}_1$ and $B^r{}_1$ behave at least as badly as $x^{1/2}$ where $x = r - r_0$ is the distance from the resonance and, for $B_1{}^\theta$ and $B_1{}^z$, at least as $x^{-1/2}$. The singularity can also be infinitely oscillatory, depending on the sign of $p'$.

A dense set of singularities can be generated in two ways: first, in the linear problem Eq. (36) by summing over Fourier components (as for a general perturbation of the boundary); second, if $\varepsilon$ is a finite parameter (no matter how small), and a formal power series in $\varepsilon$ is taken with two or more first order terms incommensurable in $(k,m)$.

One difficulty in interpretation arises from the double ignorable coordinate. Removing only one, for example, as in axial symmetry where only $\theta$ is ignorable, or in helical symmetry where $\theta - kz$ (for some fixed $k$) is ignorable, still leaves an ignorable coordinate, therefore a solvable equilibrium problem. The singularity exhibited by the linear perturbation has an "ordinary" interpretation, failure to approximate a nonlinear effect linearly. To be more precise, to every choice of ignorable helical coordinate, $\theta - kz$, there corresponds a unique flux function $\psi$. Because of the degeneracy of the cylinder, any of these serves as a flux function for the cylinder, but only one remains after the choice of a helical coordinate. The resonance condition $k \cdot \mathbf{B} = 0$ is equivalent to $\partial\psi/\partial r = 0$ where $\psi$ is the flux function that corresponds to $k/m$. Figure 3 ("tilted volcano") shows how a resonant perturbation gives a change of topology (island) which is seen

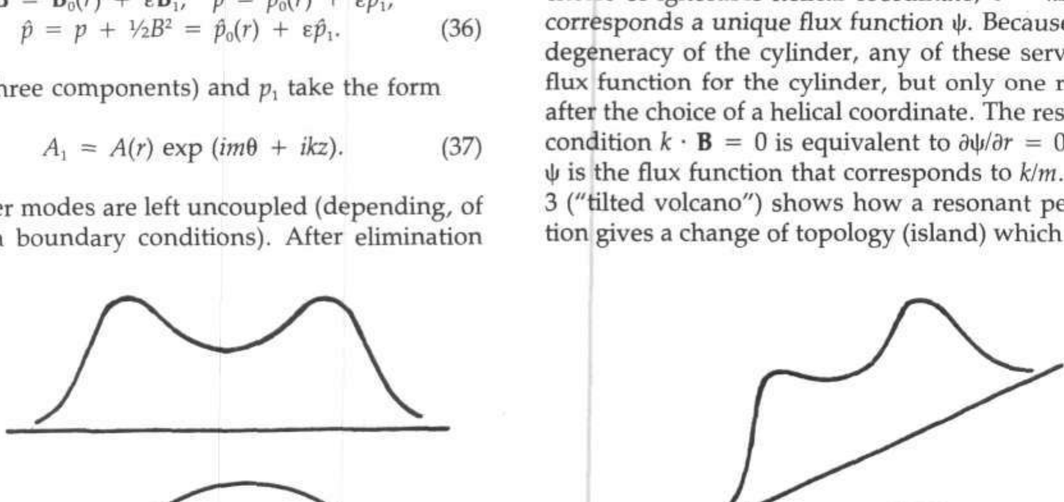


**Figure 3.** Tilted volcano

by the linear perturbation as a singularity. It is possible that under certain circumstances even a single helical resonance can destroy the existence of a solution to the particular perturbation.

A nonlinear helical (legitimate) solution with change in topology corresponds to a strict Fourier series, in multiples $n(m\theta + kz)$ of the fundamental $m\theta + kz$. If we insist on a solution with nested surfaces (that is, the topology of the original cylinder) and remove every Fourier coefficient with a singularity, or alter the profiles $\mathbf{B}^0(r)$, $p^0(r)$ to remove every singularity, then we end up with either a strictly helical perturbation without islands or $p \equiv 0$.

The question of weak solutions, with dense island pathology and worse, was raised in two previously cited works (Grad 1967b; 1970). The answer has not yet been given. It is a simple matter to make the Fourier coefficients approach zero fast enough so that the individual islands that arise from each *linear* term do not overlap and cover a set of small measure. The step from this result to a proof of existence with dense resonances for *finite* $\varepsilon$, *no matter how small*, is at least as great as from the proof of the lemma in the previous section to the proof of the Moser twist map theorem.

Island creation is only *one* consequence of loss of symmetry in an equilibrium, as emphasized in 1967b, but this failure has gradually dominated the scene at the expense of public awareness of the totality of pathological possibilities.

## Guiding Center Equilibrium

Guiding center theory derives many distinct mathematical models for the plasma from guiding center kinetic orbit theory. We mention only two: lowest order (macroscopic) guiding center fluid theory, and a version of drift kinetic theory. For our purposes, the essential similarity between guiding center fluid theory and MHD resides in the composite structure with magnetic lines as doubly counted real characteristics. (In addition, guiding center fluid theory contains more conventional cases of nonexistence which are less intrinsic and can be remedied by including more sophisticated physics in the model; see Grad 1961; 1967a).

For the drift theory we shall be concerned only with the equations of motion of the particle orbits (more precisely of the particle guiding centers). No complete equilibrium theory has been carried out, but any such formulation would certainly include this problem: given a field configuration, to assign a time-independent particle distribution (that is, density in phase space) on each magnetic line; in other words, to find a time-independent solution of Liouville's equation. To complete the theory, it would have to be made self-consistent so that the charges and currents induced by the particle flow generate the hypothesized fields. To demonstrate *nonexistence*, the second (self-consistent) half of this program is not necessary. The principle purpose of this demonstration is to show that *one should not expect to be able to find particle distributions that are time-independent*. It is a historical accident (though a very valuable one) that so many classical mathematical models of physical systems do allow static equilibria and steady flows.

The coordinates of a guiding center are the geometrical coordinates $(\alpha, \beta, s)$, where $s$ is arclength along a field line, and $(\alpha, \beta)$ are the field line coordinates

$$\mathbf{B} = \nabla\alpha \times \nabla\beta \tag{39}$$

also the velocity $V$, along a field line, and the magnetic moment, $\mu = \tfrac{1}{2}mV_\perp{}^2/B$ (an adiabatic invariant arising from the rapid circulatory motion around a field line). In equilibrium, $(s, V)$ is replaced by $\varepsilon$, the total particle energy

$$\varepsilon = \tfrac{1}{2}mV^2 + \mu B. \tag{40}$$

For present purposes, we postulate a time-independent particle distribution function in the form $f(\alpha,\beta,\varepsilon,\mu)$.

The "drift" velocity of a guiding center perpendicular to a magnetic line is given in Grad 1961:

$$\mathbf{V}_D = (1/eB^2)\,\mathbf{B} \times (mV^2\kappa + \mu\nabla B) \tag{41}$$

where $\kappa$ is the curvature of the field line, arising from $\mathbf{B} \cdot \nabla\mathbf{B}$. The motion across $\alpha$ and $\beta$ is given by

$$d\alpha/dt = \mathbf{V}_D\cdot\nabla\alpha, \quad d\beta/dt = \mathbf{V}_D\cdot\nabla\beta. \tag{42}$$

In a closed line (or open-ended) system the net drift across lines can be shown to take the form

$$\begin{aligned} \int \frac{d\alpha}{dt}\,dt &= \int \frac{d\alpha}{dt}\frac{ds}{V} = \frac{m}{e}\frac{\partial \hat{J}}{\partial\beta} \\ \int \frac{d\beta}{dt}\,dt &= \int \frac{d\beta}{dt}\frac{ds}{V} = -\frac{m}{e}\frac{\partial \hat{J}}{\partial\alpha} \end{aligned} \tag{43}$$

where

$$\hat{J}(\alpha,\beta) = \int V ds = \int \left[\frac{2}{m}(\epsilon - \mu\mathbf{B})\right]^{1/2} ds. \tag{44}$$

This determines *drift surfaces* $\hat{J}$ = const that contain particle drift orbits for circulating particles in closed line systems or for trapped particles in general. This *second adiabatic invariant* $\hat{J}$ cannot be used in the large for trapped particles except in the simplest systems. As a particle drifts, it can change its drift state (Figure 4), which changes the definition of $J$. If in the course of its motion it returns to its original drift state, the new value of $\hat{J}$ will, in general, have no relation to the original value. The result is an ergodic drift path in a

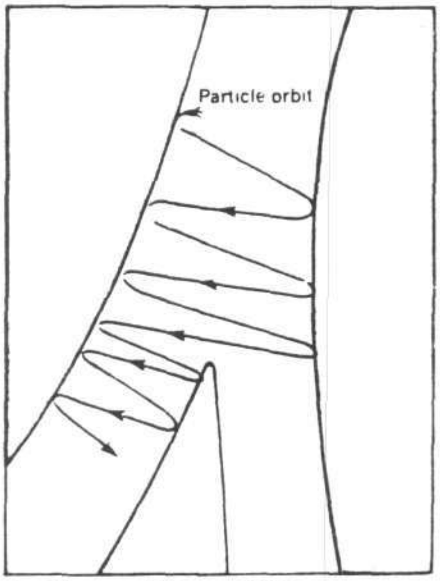


**Figure 4.** Change of Trapped State

certain volume of phase space, with $f$ = const on this region (Grad 1968). In principle this presents no insurmountable difficulties, but in practice the complications are very formidable and have not been dealt with. There are similarly no difficulties in principle with closed line systems, but very great difficulties in practice.

A difficulty, even in principle, occurs in a sheared system with dense sets of closed line surfaces. For a circulating particle on an irrational surface, the net drift velocity off the surface is zero (Grad 1961; 1967b). On a rational surface the net velocity on each closed line is

$$\int \frac{d\psi}{dt}\frac{ds}{V} = -\frac{m}{e}\frac{\partial \hat{J}}{\partial \theta}, \quad \hat{J} = \int \left[\frac{2}{m}(\epsilon - \mu \mathbf{B})\right]^{1/2} ds \qquad (45)$$

where $\theta$ is the flux coordinate conjugate to $\psi$. For a time independent equilibrium, $\hat{J}$ must have the same value on each closed line of the rational $\psi$-surface for every pair $\varepsilon$ and $\mu$. For example, for particles with $\varepsilon$ very large compared to $\mu B$, this implies that the geometrical length of all closed lines on a rational surface must be equal. For some other $\varepsilon$ and $\mu$, the relation will be different. For this reason, unless there is symmetry, it is too much to expect a particle distribution to be time independent for each $\varepsilon$ and $\mu$. There is no *a priori* reason to expect such a miracle. The best that can be hoped for is that by adjusting the electrostatic potential, the net charge (current) will not accumulate. That even this cannot be arranged is shown by the macroscopic analysis. Considering the wide variety of orbits, what the present microscopic analysis shows, *even without taking self-consistency into account,* is the impossibility of finding a particle distribution which will not fluctuate.

For specialists in plasma confinement we add the note that except in very special cases, the formulation of guiding center equilibrium in terms of $f(\varepsilon,\mu,\hat{J})$ or in terms of omnigenity is overly naive; each concept adds to the nonexistence difficulties.

## Anholonomic Linear Instability

The nonexistence of equilibria without very special symmetry raises an obvious question with regard to *stability:* What is the effect of nonexistence of equilibrium on the time evolution of small asymmetric perturbations in the neighborhood of a legitimate, symmetric, equilibrium? The answer is unambiguous: Any perturbation that is not accessible to the symmetric equilibrium by a smooth displacement (that is, which is *anholonomic*) is unstable. Furthermore, all such (anholonomic) perturbations are excluded by the standard ($\delta W$) treatment of MHD stability![7] The last statement is apparent, since the $\delta W$ construction is premised on the existence of a smooth Lagrangian displacement, which is generally believed to imply that there is no change in the field line or magnetic surface topology.[8] We see then that the conventional Lagrangian displacement (holonomic) stability theory is unduly restrictive with regard to the class of admitted perturbations, since it is easily verified that the magnetic topology can be altered by an arbitrarily small perturbation of $\mathbf{B}(x)$ (which does not arise from a displacement, $\xi$). The essential point is that all perturbations that destroy the topology are unstable.

Another peculiarity of the conventional variational, $\delta W$, theory, is that the natural norm, $\int \rho_0 \xi^2 < \infty$ with suitable $L_2$ restrictions on derivatives of $\xi$, does admit $\xi$, which can change the topology (for example, a cusped perturbation $\xi$). We shall see, however, that the particular unstable, anholonomic perturbations that we consider are not included by this completion of the space of allowable $\xi$ (and are as smooth as desired).

First, we establish the nonexponential instability of anholonomic perturbations of a given sheared toroidal equilibrium with nested $p$-surfaces. Then we show that this instability does not correspond to an isolated point eigenvalue.

A flux surface is defined as a global, single-valued solution of

$$\mathbf{B} \cdot \nabla\psi = 0. \qquad (46)$$

Given a field $\mathbf{B}_0$ and associated family of flux surfaces $\psi_0$ = const., $\mathbf{B}_0 \cdot \nabla\psi_0 = 0$, by perturbing, $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$, we get the linearly perturbed equation for $\psi_1$,

$$\mathbf{B}_0 \cdot \nabla\psi_1 + \mathbf{B}_1 \cdot \nabla\psi_0 = 0 \qquad (47)$$

$$\psi_1 = -\int (\mathbf{B}_1 \cdot \nabla\psi_0)\frac{dl_0}{B_0}$$

where the (indefinite) integral is taken along unperturbed field lines $\mathbf{B}_0$. The theory of this equation is discussed above in the section "An Ergodic Lemma." The necessary and sufficient condition that $\psi_1$ be single-valued is that

$$\oint \mathbf{B}_1 \cdot \nabla\psi_0 \frac{dl_0}{B_0} = 0 \tag{48}$$

for every closed line on a rational surface.

Next, consider the classical circulation

$$C = \oint \mathbf{u} \cdot d\mathbf{x} \tag{49}$$

on each closed line of a rational surface in a sheared equilibrium with $dp/d\psi \neq 0$. A linearized calculation gives (after some manipulation)

$$\begin{aligned} \frac{dC}{dt} &= \oint \left(\frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{B}_0\right) \frac{dl_0}{B_0} \\ &= -\frac{1}{\rho_0} p_0'(\psi_0) \oint \mathbf{B}_1 \cdot \nabla\psi_0 \frac{dl_0}{B_0} \end{aligned} \tag{50}$$

Thus $dC/dt = 0$ if and only if the field perturbation due to $\mathbf{B}_1$ allows flux surfaces to first order. Continuing,

$$\begin{aligned} \frac{d^2C}{dt^2} &= \frac{1}{\rho_0} \oint \frac{\partial \mathbf{B}_1}{\partial t} \cdot \nabla p_0 \frac{dl_0}{B_0} \\ &= \frac{1}{\rho_0} \oint \operatorname{curl} (\mathbf{u} \times \mathbf{B}_0) \cdot \nabla p_0 \frac{dl_0}{B_0} \\ &= \frac{1}{\rho_0} \oint \operatorname{div} [(\mathbf{u} \times \mathbf{B}_0) \times \nabla p_0] \frac{dl}{B_0} \\ &= \frac{1}{\rho_0} \oint d(\mathbf{u} \cdot \nabla p_0) \\ &= 0 \end{aligned} \tag{51}$$

In other words, $C = a + bt$ with $b \neq 0$ if the perturbation $\mathbf{B}_1$ changes the topology and $b = 0$ if it does not. It is easily verified that there is no double eigenvalue at the origin, $(\rho_1, \mathbf{u}, \mathbf{B}_1)$ = linear in $t$, except for certain very special cases. Therefore, the anholonomic perturbation is associated with a part of the essential spectrum that includes the origin, $\omega = 0$. The perturbation is unstable, and a certain property of the appropriate generalized eigenfunction (but not the eigenfunction itself) is linear in time. This part of the spectrum is not contained in the self-adjoint spectrum that belongs to $\delta W$.

Although discovered in 1971, there appears to be no further work concerning this anholonomic instability. The fact that the growth is linear, not exponential, has no great practical significance when one recalls that the noise level in most plasmas is on the order of 1 to 10 percent.

The first order linearized MHD equations take the form

$$\partial u/\partial t = L_1 v, \quad \partial v/\partial t = L_2 u \tag{52}$$

where $u$ is the velocity and $v = (p, \mathbf{B})$; the symmetric hyperbolicity implies that the differential parts of $L_1$ and $L_2$ are formally adjoint. The standard second order self-adjoint form (which then leads to $\delta W$) can be derived by differentiation,

$$\partial^2 u/\partial t^2 = L_1 L_2 u \tag{53}$$

where $L_1L_2$ is formally self-adjoint. This is *not* the conventional derivation. Rather, $\xi$ is introduced as

$$\partial \xi/\partial t = u. \tag{54}$$

We drop the integration constant in the second equation Eq. (52):

$$v = L_2 \xi \tag{55}$$

from which

$$\partial^2 \xi/\partial t^2 = L_1 L_2 \xi. \tag{56}$$

These elementary operations look trivial, but they are not. For example, if there is a solution $u \sim t^{-1/2}$ of Eq. (53) (stable), it corresponds to $\xi \sim t^{1/2}$ [unstable for Eq. (56)]. The relation of either $\xi$ or $u$ to the original (and correct) variables $(u,v)$ is not so trivial as mere differentiation. Comparing sets of initial values for Eqs. (52), (53), and (56),

$$\begin{aligned} &(\xi,\ \xi_t = u)\ (56) \\ &(v,\ u)\ (52) \\ &(u_t,\ u)\ (53) \end{aligned} \tag{57}$$

we see that for a given norm applied uniformly to all three equations, the spaces are related by

$$(\xi, \xi_t) \subset (u, v) \subset (u, u_t) \tag{58}$$

The two outer sets are formally self-adjoint; the original one $(u,\ v)$ appears not to be. The exponential spectra and eigenfunctions are identical, but the spectra associated with $\omega = 0$ are not. The initial values also differ; for example, $(u,\ v)$ includes anholonomic, unstable initial values that are not included in $(\xi,\ \xi_t)$ and are not unstable in $(u,\ u_t)$.

The (physically correct) first order system in $(u,\ v)$ can be related to the more restricted Lagrangian $\xi$ formulation by keeping the integration "constant" in Eq. (55):

$$v = L_2 \xi + f(x) \tag{59}$$

which implies

$$\frac{\partial^2\xi}{\partial t^2} - L_1L_2\xi = L_1f. \qquad (60)$$

The anholonomic terms are contained in the inhomogeneous term $L_1f$ which generalizes Eq. (56) (Lortz and Rebhan 1971). To obtain a solution growing in time as a result of the inclusion of $L_1f$ clearly depends on a nontrivial nullspace of $L_1L_2$.

The special nullspace which leads to anholonomic instability arises from the elementary (linearly) neutral perturbation of the homogeneous equation

$$\xi = \alpha(\psi)\mathbf{B} + \beta(\psi)\mathbf{J} \qquad (61)$$

where $\alpha$ and $\beta$ are arbitrary functions of $\psi$. This is a doubly infinite degeneracy at $\omega = 0$. This displacement, which is neutral for $\delta W$, becomes unstable for the original, first order, MHD formulation. The instability can be considered to be a resonance ($\omega = 0$) with the initial disturbance $L_1f$ [$f$ is expressed in terms of initial values of $v$ and $\xi$ in Eq. (59)].

## Discussion

It has already been mentioned that, from the first discovery of the impossibility of exactly balancing plasma by magnetic pressure (Grad 1967b), the emphasis was on what positive steps to take in order to salvage the large, but somewhat ambiguous literature on equilibrium and stability. One attitude (though unconstructive) is immediate: Since all the mathematical models are known to be only crude approximations to the real world, crude answers should suffice. To this, there are several immediate objections: (1) crude "answers" to what questions? (2) the physical consequences of lack of pressure balance are observable and frequently important; (3) the classical, almost universal technique (and the only one, if mathematics is to be used as a tool) is to replace the "real" imprecisely formulated problem by a precise model (or many precise models) from which precise conclusions are drawn and then interpreted in the real world; (4) the nonexistence results in equilibrium theory are mimicked by instabilities (anholonomic); (5) there should be structural stability, that is, invariance of the conclusions with respect to the choice of model. Nonexistence of solutions for the equilibrium equations, in the face of an urgent psychological need for them, has evoked strong aversion within the plasma physics community toward the conclusions presented here. Yet, most concepts spawned by nonexistence have gained wide acceptance under other names while the central result, that *solutions do not exist*, is sidestepped.

Consider the question of plasma stability. The question is almost impossible to formulate except in terms of the time-dependent behavior of solutions that are initially perturbed from an equilibrium or steady state.[9] This problem was reformulated by the author in 1970 (Grad 1970). For example, given a formal power series that is to some order a "solution" of the desired equations, truncate the series, assign a value to the expansion parameter, and take the resulting expression as an initial value for the equations of motion. The initial value problem (certainly in MHD and, in most cases, in guiding center fluid theory) is well posed. If this initial state does not change too much in an appropriate amount of time, then the "equilibrium" is stable. No such procedure has been followed, but it could be, in principle. This formulation is evidently fuzzy. The answer may very well depend on the number of terms in the series, the value of the expansion parameters, and the duration of observation. We are led to a formulation which bears a resemblance to the rather imprecise physical interpretation of the pathological KAM theory described in the section on an "Ergodic Lemma."

Another way of approaching the problem of nonexistence is familiar from the theory of asymptotic series: Don't ask for too much accuracy, and don't take too many terms. The situation is somewhat different, however, since there is no solution to which the series which is used is asymptotic.

The same advice would be relevant to a numerical calculation. Refining the mesh and otherwise "improving" the accuracy may be a mistake beyond some point. Fortunately, since all the relevant problems are fully three-dimensional, this concern is, at the present level of computational power, usually unnecessary.

There are a number of papers that bear on the question (in Horton et al. 1983). In one, a selection is made of a few "important" resonances; this selection is clearly a matter of judgment (but our attitude is that this is unavoidable). Another employs an ad hoc weak stochastic mechanism; this serves to mask most of the KAM pathology; of course, the pathology will reappear if the amount of stochasticity is made arbitrarily small—again unavoidably a matter of judgment. Still another paper eliminates resonances one by one; the claim is made (empirical) that the procedure is convergent. It almost certainly is not; nevertheless the procedure remains very useful, but only up to a point. Another common procedure which may be useful is to consider only two pathologies: islands and ergodic regions. This is an oversimplification, but the true picture may be too complex for practical application.

A psychological block to the easy acceptance of this theory is the very large analytic literature (for example, stellarators, perturbations of tokamaks, use of nonexistent Hamada coordinates, and so on) in which the formalism gives no hint of trouble. The reason is almost always that one of the small parameters is the rotation number (or rotation number per period). In

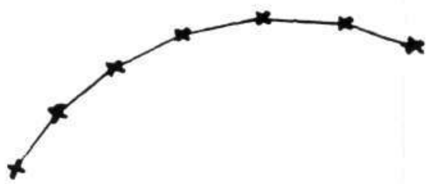


**Figure 5.** Poincaré map with small rotation number

a Poincaré map with small $\alpha$ (Figure 5), an obvious first aproximation is to replace the discrete difference equation by a differential equation. Almost any such approximation has the property that if it is continuous to first order, it is continuous to all orders. Also, as is easily shown, an incompressible flow in two dimensions with approximately closed streamlines has exactly closed lines. *Both the discreteness and the correct pathology of the map are invisible as a consequence of the mathematical technique.* On the other hand, once one is sensitized to the possibility, one can find intrusions of pathology in the analytical literature (for example, stellarator expansion near the magnetic axis), and in the numerical literature (evidence of islands, once the possibility is pointed out). Once the question is raised, it becomes clear that there is a choice of analytic techniques; some oblivious to all pathology, others that give "resonances" one at a time (with succeeding terms); and the most accurate (and most difficult) that give all the pathology at once. One can take advantage of this choice of the degree of blindness of analytic procedures. Similarly there are numerical techniques for either obscuring or emphasizing the pathology; the first to allow one to obtain an equilibrium, and the second to study its limitations (Bayliss et al. 1981).

The suggestion has been made that true mathematical solutions may exist in some weak sense (Grad 1967b). This cannot be ruled out, and a step toward formulating such a theory variationally has been made (Grad 1970). But it must be remarked that there is no evidence at all for such a conclusion. The classical theory of distributions is not available since it is strictly linear. Also, there would be very little practical consequence of such a theory (beyond the reduction of psychological tension) since the mathematical methods involved of necessity would be abstract. There is certainly no experimental evidence for quiescent confined plasma behavior, nor is it needed for successful operation of experiments (only for successful operation of theoretical tools).

A paradoxical question of interpretation arises between what we could term prototypes of mathematician ($M$) and physicist ($P$). A formal series, without error estimate, would be considered qualitative by $M$; a rigorous estimate, $x < a$, would be considered qualitative by $P$. The series would be considered quantitative by $P$ since two people would (generally) compute the same coefficients; the estimate would be considered qualitative by $P$ since a second person could conclude (also rigorously), $x < b$. Turn now to the rigorous KAM theory and the sometimes related "asymptotic" (but unproved) series. In practice, the rigorous (therefore quantitative) KAM theory is used to obtain a *qualitative* picture of what phenomena might arise; the qualitative (incorrect) series is the only practical way (other than numerical) to obtain actual numbers; it is therefore quantitative! A practical example is the use of the Birkhoff technique of successively removing resonances to obtain what is claimed (empirically) to be a convergent, pathology-free answer (Horton et al. 1983). As we have said, the procedure is not convergent but, up to a point, it is useful.

We close with a few suggestions (out of very many) for relatively elementary studies in this field. The first concerns the simplified force-free field, curl $\mathbf{B} = \lambda \mathbf{B}$, with constant $\lambda$. This has the enormous advantage of being linear and strictly elliptic. One can *superpose* two finite explicit solutions (for the same value of $\lambda$) and study (1) flux surfaces; and (2) $\int dl/B$ on rational surfaces. Explicit solutions are available with any of the usual symmetries.

The second problem is that of small nonlinear perturbations of the cylindrical screw pinch. The question here is: What do the singularities of the linear solutions (explicit from the Appendix) imply with regard to weakly nonlinear perturbations?

It has been conjectured that many non-MHD instabilities in which the growth rate vanishes with a parameter as one approaches ideal MHD (for example, resistive, drift) should leave a detectable residue of the spectrum near the origin in ideal MHD (Grad 1973). The study of anholonomic instabilities which arise from the vast assortment of ideal spectra which extend to the origin may provide a fertile ground to find such connections.

To sum up this summary: The most important point is to recognize that there are inherent limitations in obtaining accurate equilibria balancing plasma and field. Quantitative techniques for recognizing and handling these phenomena are beginning to be developed analytically and numerically. An occasional obfuscation arises from a sloppy solution that may be only careless, or may serendipitously be ingenious. The interpretation of techniques and solutions is paramount.

## Acknowledgment

The author wishes to express his sincere thanks to Albert A. Blank for many useful and stimulating conversations on the entire content of this paper and for his unstinting efforts in improving the manuscript.

## Appendix: Cylindrical Perturbation

Consider the original, unperturbed cylinder $0 < r < a$, $0 < z < 2\pi$ with given $B_\theta(r)$, $B_z(r)$, $p(r)$ satisfying

$$B_z B_z' + B_\theta B_\theta' + B_\theta/r + p' = 0. \qquad (62)$$

Perturb, making the following substitutions,

$$\mathbf{B} \rightarrow \mathbf{B}(r) + \epsilon\beta(r,\theta,z), \qquad (63)$$

$$\mathbf{p} \equiv p + \frac{1}{2} B^2 \rightarrow \mathbf{p}(r) + \epsilon\phi(r,\theta,z)$$

to obtain

$$\mathbf{B} \cdot \nabla\beta + \beta \cdot \nabla\mathbf{B} = \nabla\phi, \;\; \mathrm{div}\,\beta = 0 \qquad (64)$$

and set

$$\beta = \beta(r)\exp(im\theta + ikz), \qquad \phi = \phi(r)\exp(im\theta + ikz). \qquad (65)$$

Some manipulation (including algebraic elimination of $\beta_\theta$ and $\beta_z$, which are not differentiated) yields (for $k^2 + m^2 \neq 0$)

$$\delta\beta_\theta = \frac{m}{r}\phi + i(B_\theta' + B_\theta/r)\beta_r \qquad \delta\beta_z = k\,\phi + iB_z'\beta_r \qquad (66)$$

$$\delta\phi' = -(2mB_\theta/r^2)\phi + \frac{i}{r}[\delta^2 - \frac{2B_\theta}{r}(B_\theta' + B_\theta/r)](r\beta_r)$$

$$\delta(r\beta_r)' = -ir(k^2 + m^2/r^2)\phi + [kB_z' + (m/r)(B_\theta' + (B_\theta/r)](r\beta_r) \qquad (67)$$

where

$$\delta \equiv k\mathbf{B}_z + \frac{m}{r}\mathbf{B}_\theta = \mathbf{k} \cdot \mathbf{B}. \qquad (68)$$

Equations 66 and 67 for the cylindrical perturbation have a number of interesting properties (including the fact that the trace of the coefficient matrix on the right is $\delta'$ and the determinant has the sign of $p'$ (near $\delta = 0$). A constraint, such as adiabatic or $p(\psi)$ = given, and so on, will be observed only in the $k = m = 0$ mode for which the equations take a different form.

## References


Arnold, V.I. and A. Avez. 1968. *Ergodic Problems of Classical Mechanics*. New York: W.A. Benjamin.

Bayliss, A., H. Grad, and J. Hogan. 1981. "Resistive Diffusion in High-β Tokamak with Finite Ripple." *Bull. Am. Phys. Soc.* **26 (7)**: 1002.

Berkowitz, J., K.O. Friedrichs, H. Goertzel, H. Grad, J. Killeen, and E. Rubin. 1958. "Cusped Geometries." In *Second United Nations Conference on the Peaceful Uses of Atomic Energy*. Geneva. **31**: 171-176.

Bineau, M. 1972. "On the Existence of Force-free Magnetic Fields." *Comm. Pure Appl. Math.* **25**: 77-84.

Chandrasekhar, S. 1956. "On Force-free Magnetic Fields," *Proc. Nat. Acad. Sci.* **42** (1):1.

Gilbarg, D. 1960. "Jets and Cavities." In *Handbuch der Physik*, Vol. IX. S. Flügge, ed. (Berlin: Springer).

Grad, H. 1961. "Microscopic and Macroscopic Models in Plasma Physics." In *Proceedings of the Symposium on Electromagnetic and Fluid Dynamics of Gaseous Plasma*. Polytechnic Institute of Brooklyn. Pp. 37-64.

———. 1967 (a). "The Guiding Center Plasma." In *Symposia in Applied Mathematics*. Harold Grad, ed. **18**: 162-248. American Mathematical Society.

———. 1967 (b). "Toroidal Containment of a Plasma." *Phys. Fluids* **10**: 137-154.

———. 1968. "Anomalous Particle Losses in Toroidal and Complex Mirror Geometries." *Bull. Am. Phys. Soc.*, Ser. II. **13**: 298.

———. 1969. "Problems in Magnetostatic Equilibrium." Presented at International Colloquium on MHD, Lille, June, 1969; also AFOSR 70-1134 TR, MF-62, April 1970.

———. 1970. "Mathematical Problems Arising in Plasma Physics." *Actes, Congrès International Math.* Nice. **3**: 105-113.

———. 1973. "MHD Stability Theory." *Bull. Am. Phys. Soc.* Series II. **18 (10)**: 1320.

———. 1978. "Reconnection of Magnetic Lines in an Ideal Fluid." Courant Institute Report MF-92.

Grad, H. and H. Rubin. 1958. "Hydromagnetic Equilibria." In *Second United Nations Conference on the Peaceful Uses of Atomic Energy*, Geneva, **31**: 190-197.

Horton, C.W., Jr., L.E. Reichl and V.G. Szebehely (eds.). 1983. *Long-Time Prediction in Dynamics*. New York: John Wiley, pp. 45-70.

Kadish, A. and D.C. Stevens. 1974. "Equilibria and Adiabatic Compression of Free-boundary Belt Pinches." *Nucl. Fusion* **14**: 821-829.

Lortz, D. 1970. "Über die Existenz toroidaler magnetohydrostatischer Gleichgewichte ohne Rotationstransformation." *Z. Angew. Math. Phys.* **21**: 196.

Lortz, D. and E. Rebhan. 1971. "On the Sufficiency of the Energy Principle for Magnetohydrodynamic Stability." *Physics Letters A*. **35**: 236.

Moser, J. 1962. "On Invariant Curves of Area-preserving Mappings of an Annulus." Akad. Wiss., Göttingen, *Nachrichten*, Abt. II, Math. Phys. Kl. **1**: 1.

Schechter, Martin. 1959. "Some Axially Symmetric Problems in Magnetohydrodynamics." In *Proceedings of the Conference on Theoretical Aspects of Controlled Fusion Research*, Gatlinburg, TN, April 27-28, 1959, 176-186. ORNL 2805 (TID-7582).

Taylor, J.B. 1975. "Relaxation of Toroidal Discharges to Stable States and Generation of Reverse Magnetic Fields." International Atomic Energy Agency, Fifth Conference on Plasma Physics and Controlled Nuclear Fusion Research (Nov. 1974). Tokyo, Japan. **1**: 161.


## Notes


[1]This work was supported by U.S. Department of Energy Contract No. DE-AC02-76ER03077.

This manuscript is essentially the same as one presented at the Workshop on Mathematical Aspects of Fluid and Plasma Dynamics at the University of Trieste, Italy, June, 1984.

[2]H. Grad, "EBT Transport with Anisotropic Core," unpublished.

[3]For more detail, see Appendix in Grad 1967b.

[4]The "published" account (Grad 1969) shows this as a blank page (the submitted manuscript, it is claimed, was never received). The complete paper is available as the report (MF-62) cited in this reference.

[5]From

$$0 = \text{curl}\,\mathbf{B} \times \mathbf{B} = \mathbf{B}\cdot\nabla\mathbf{B} - \nabla \tfrac{1}{2}B^2,$$

we see that the principle normal to a field line is normal to the surface; hence the line is a geodesic.

[6]This example is unpublished in the equilibrium context; but it *is inherent in the stability literature.*

[7]The material in this section was derived by the author in 1971 and was, at that time, discussed with D. Lortz and E. Rebhan who had independently obtained and published related results (Lortz and Rebhan 1971). Their interpretation and emphasis are different; they treat the insufficiency of $\delta W$ for stability rather than the topological or equilibrium-related consequences. Moreover, the class of counter-examples is not the same.

[8]For a counter-argument, see the author's 1978 work.

[9]An exception (the result of another type of *nonexistence*) is described in Grad 1967a. Here nonexistence defines instability in a general, nonsteady flow, without perturbing from equilibrium.

# Reports

## On the Controversy over Whether Classical Systems Like Plasmas Can Behave Like Superconductors (Which Have Heretofore Been Supposed to Be Strictly Quantum-Mechanically Dominated)

**by Winston Bostick**

The Meissner-Ochsenfeld effect is one of the crucial distinctions between a perfect conductor (zero resistance) and a superconductor. In a perfect conductor the magnetic field is frozen in so that

$$\partial B/\partial t = 0$$

(except at the surface). In a Type I (soft) superconductor, if the applied magnetic field is below a certain critical value $B_C$, the magnetic field is expelled and $B = 0$ except at the surface. In a perfect conductor in which electrons are freely accelerated by the electric field $E$, the rate of change of current density $\mathbf{J}^s$ carried by those electrons is

$$\mathbf{E} = \partial/\partial t\ (\Lambda\ \mathbf{J}^s), \qquad (1)$$

where $\Lambda$ is a temperature-dependent constant of the material with dimensions of time and where $s$ denotes the superconducting part, and where

$$\Lambda = 4\pi\lambda_L{}^2/c^2 = m/h_s e^2$$

is a phenomenological parameter. This London equation plus the second London equation

$$\mathbf{h} = -c\nabla \times (\Lambda \mathbf{J}^s) \qquad (2)$$

put forth in 1935, describes the two foregoing basic electrodynamic properties that give superconductivity its unique interest. The value of magnetic flux density on a microscopic scale is denoted by **h**. The second London equation, when combined with the Maxwell equation,

$$\nabla \times \mathbf{h} = 4\pi J/c \qquad (3)$$

gives the equation

$$\nabla^2\mathbf{h} = \mathbf{h}/\lambda_L{}^2, \qquad (4)$$

where $m$ is the electron mass, $e$ its charge, $n_s$ the superconducting electron density per cubic centimeter and $\lambda_L$ is the penetration depth in centimeters; **h** (and **B**) thus decays exponentially with distance into the superconductor with a mean distance

$$\lambda_L = \sqrt{mc^2/4\pi ne^2}$$

and is expelled (except for the surface) in a type of "dc skin effect." In 1935 there was no explicit involvement of Planck's constant in Eq. (3) and no attempt was made to derive Eq. (3) from any fundamental theory of superconductivity: None existed at that time. It was always recognized, however, that the justification of Eq. (3) was imbedded in quantum mechanics which somehow imparted a rigidity to the overall wavefunction of the superconducting electrons. This recognition was justified with the Bardeen, Cooper, Schrieffer theory of superconductivity, which showed that for the Type I superconductors, when the free energy released in the formation of a superconductivity blanket of Cooper-paired electrons was greater than the energy of the expelled magnetic field, the Meissner-Ochsenfeld effect would occur.

### The Edwards Argument

There has been a series of papers during the last three years on the subject of whether these originally phenomenologically theorized but now quantum-mechanically canonized London equations can be given a purely classical derivation. W. Farrell Edwards (Edwards 1981), the first speaker for the affirmative in this recent debate, claims to show that the London equations do indeed have a classical derivation that applies to superconductors and to some collisionless plasmas as well (for example, filamentation of the current sheath of the plasma focus and the magnetic flux ropes in the Venerian ionosphere). Edwards leads off with transforming well-accepted, classical Lagrangian actions into Eulerian notation for collisionless fluids at absolute zero temperature and then applies the principle of least action to obtain the equations of motion. Energy is assumed to remain constant; as a result the system is assumed to be dissipationless, which means that collision times are long, compared with the persistence times of the phenomena under scrutiny. The appropriate-current-density variation of the Euler-Lagrange equations (4-vector notation, which is very complicated) yields the important equation

$$A_\sigma + (m/q)u_\sigma = 0 \qquad \text{(Edwards's 7)}$$

where $m$ is the electron mass, $q$ the charge, $u_\sigma$ the appropriate fluid velocity component, $A_\sigma$ the appropriate vector potential component. In using Eq. (Edwards's 7) to eliminate $A_\sigma$ from the electromagnetic tensor the three-dimensional, low-velocity London equations become

$$\mathbf{B}_z = -(m/q)\, \nabla \times \mathbf{u}$$

and

$$\mathbf{E} = (m/q)(\partial \mathbf{u}/\partial t + \tfrac{1}{2}\nabla u^2).$$

Edwards points out that free-energy considerations predict that plasmas

$$\beta \equiv nkT/(B^2/2\mu_0) > 1$$

(that is, high beta) should exhibit the Meissner effect, but low-beta plasmas should show the familiar frozen-in fields. Edwards's work has not gone unnoticed: his claim is so unusual that the sacred columns of *The Times* directed their attention in his direction (*The Times* 1982). And the following artillery-barrage response has been so intense as to blow almost anyone off the map.

### The Opposition

The first man up to bat for the opposition is Frank S. Henyey, who states that the result Edwards has derived is that for an ideal plasma where the flux *will* be frozen in (Henyey 1982). And if the initial flux is zero, the flux *will* be *frozen out*. Henyey asserts that Edwards treated the action as if it is a quantity with equality implying equivalence, whereas, in fact, the action is a function of the dynamical variables, and equivalence requires more than equality at the extrema. Henyey proceeds to carry out the variation of the action integrals the "proper way according to Courant and Hilbert" and comes up with the corrected equation which should replace Eq. (Edwards's 7):

$$A_\sigma + (m/q)u_\sigma + \lambda_\sigma = 0 \qquad \text{(Henyey's 5)}$$

which differs from Eq. (Edwards's 7) by the presence of $\lambda_\sigma$. The solution $\lambda_\sigma = 0$ gives zero flux, which would make Eq. (Henyey's 5) the same as Eq. (Edwards's 7). However, the general solution requires only the constancy of $\lambda_\sigma$ along particle trajectories, not the vanishing. The more general solution, Henyey declares, is the well-known freezing of magnetic field lines into the plasma.

The next players up to bat form a triumvirate: B. Segall, L.L. Foldy, and R.W. Brown (Segall, Foldy, and Brown 1982). They assert that if Edwards's approach were valid it should lead to correct results when it is applied to a system of *neutral* noninteracting particles. Thus with the charge $q = 0$ the variation with respect to the velocity $\mathbf{u}$ (since the correct density $\mathbf{j} = 0$) leads to $\mathbf{u} = 0$ instead of the correct Euler equation. They further state that Edwards's equations #7 and #10 can be shown to be not valid classically, by taking their curl and then the limit $q$ goes to zero. The result, independent of gauge, is that curl $\mathbf{u} = 0$. But there are many systems that violate this condition. They caution against attempting to obtain a proper action principle with $\mathbf{u}$ a variational coordinate simply by naively transforming a particle Lagrangian to a corresponding density. This caveat is suggested by the verifiable fact that $v_n$ cannot be used as a variational coordinate in the particle description. To obtain a valid Hamilton's principle for the continuum, one can introduce constraints such as the continuity condition with the use of Lagrangian multipliers, or appropriately modify the Lagrangian density. In addition, they state that any simple generalization of Edwards's procedure involving only the fluid velocity (and $\rho$) in the Lagrange density cannot account for the contribution to the stress tensor term on the right-hand side of Euler's equation, which appears even for a system of noninteracting particles. This procedure requires knowledge of the local deviation of particle velocities from their local mean value (the fluid velocity) and that is absent in Edwards's formulation.

### More Opposition

The next man up to bat (for the opposition), who apparently hits a homerun, is Paul G.N. deVegvar, who notes that Edwards's four-current vector (his Eq. #5) involving the worldline of the *i*-th particle is a highly singular object, vanishing off worldlines and being $\delta^3$-like on them. To get the usual Euler-Lagrange equations by setting $[\cdots] = 0$, all terms must exist and be continuous. This criterion fails when applied to Edwards's work. The current densities in the London equations are smoothed, and not the delta-function variety, which Edwards has constrained his solutions to be. Finally, deVegvar points out that Edwards's Eq. #7, even if justifiable, could not describe a superconductor, because it is a *local* relationship between currents and fields. As is well known, the current response of an electron gas to a field is nonlocal: $j_\beta$ at a point depends upon the spatial average of extent $\sim l$ and a time average of $\sim l/v_F$ ($l$ is the mean free path, $v_F$ the Fermi velocity). Thus even if one could use Eq. (Edwards's 7) at all in superconductors, it would be restricted to the case of static uniform fields, precluding its application to penetration phenomena.

After the spine-jolting shocks of this artillery barrage, Edwards bounces back (Edwards 1982) for his next inning, proving himself to be tough, resourceful, and not nearly so naive as some of his opposition might have assumed him to be:

Edwards recollects that treatments of collections of particles as continuous systems are common, and literature covering action-integral approaches using Eulerian variables, where variations are taken with respect to the velocity fields, is extensive. In fact, without suggesting that the derivation is classical or that there might be classical applications, Geurst (Geurst 1980) recently used an action principle very similar to Edwards's $I_2$ to derive the Ginzburg-Landau equations as well as the London equations. Edwards apologizes for not being aware of Geurst's paper when he wrote his earlier paper. Geurst's work apparently avoided the ferocious fire of the opposition, probably because he did not claim that his derivation was classical.

Edwards asserts that the key question is whether his Eulerian action $I_2$ properly describes certain classical systems, and he says that his adversaries who have attempted to negate this possibility have actually aided in the understanding of the physical systems to which his $I_2$ applies.

Considering deVegvar's comments about the singularities which arise, Edwards uses a proof of convergence based on the Schwartz distribution theory: he approximates the worldlines of each particle $i$ by a tube extending a small radius $\xi_i$ from the worldline and he approximates $j_\sigma$ by a smooth function $J_\sigma$, which is nonzero within the tube, falls to zero quickly and smoothly immediately outside the tube, and remains zero beyond the radius $\zeta_i$ where $\zeta_i > \xi_i$. Within the tube where $I_2$ is differentiable with respect to $J_\sigma$, the Euler-Lagrange equations are satisfied, and consequently

$$A_\sigma + (m/q)U_\sigma = 0$$

where

$$U_\sigma = J_\sigma/\sqrt{J_\alpha J^\alpha}.$$

As we shrink the tube by letting $\zeta_i$ become small, $J_\sigma$ approaches $j_\sigma$, and $U_\sigma$ approaches $u_\sigma$, the velocity of the particles.

Concerning the nonlocal nature of the response of an electron gas, Edwards points out that the London theory is a local description and is valid to first order whether one considers its derivation to be classical or quantum mechanical. The nonlocal theory of Pippard, later derived from the Bardeen, Cooper, Schrieffer theory, is a refinement yet to be investigated. So much for Mr. deVegvar's comments.

### Edwards's Reply

Now here is Edwards's reply concerning Segal, Foldy, and Brown's attempt to demonstrate that the results from $I_2$ are invalid unless one introduces constraints, because when $q = 0$, the variation in $I_2$ with respect to $u_\sigma$ results in $\mu_\sigma = 0$. Edwards argues that with $q = 0$, conservation of mass is lost and must be inserted as a constraint; consequently, no longer do we get $\mu_\sigma = 0$ but

$$\mu_\sigma + \partial\eta = 0$$

where $\eta$ is an undetermined multiplier.

However, as Segal, Foldy, and Brown point out, even that result is unacceptable for a neutral fluid because the curl of the velocity is zero. The apparent dilemma has been extensively discussed in the literature and has resulted in the introduction of the Lin constraint (Lin 1963), which removes the contradiction, and when applied to $I_2$ for a charged fluid leads directly to the Lorentz force equations without the London equation restrictions.

Then Edwards raises the unanswered question: "But does the neglect of the Lin constraint necessarily lead to absurdities and, if not, under what circumstances can it be neglected?" Thus, Edwards has really held his own against the adversary criticism of Messrs Segal, Foldy, and Brown.

Concerning Henyey's comments, Edwards gratefully points out that Henyey's derivation clarifies the differences between $I_2$ (unconstrained) and $I_1$. Using $I_2$ without the Lin constraint (as Edwards did) is the same as $\lambda_\sigma = 0$ in Henyey's equation of motion. The result is the London equations, the solutions of which are a subset of the solutions of the Lorentz force relation.

Edwards now attempts, and successfully this author believes, to advise when to neglect the Lin constraint: the Lin constraint refers to the ability to follow the path of a fluid element from one position to another. For most fluids this is possible; hence, the constraint applies. However, because of the Heisenberg uncertainty principle, one is unable to follow the trajectory of a quantum fluid as discussed in a recent paper by Putterman (Putterman 1982). Thus, at least for such fluids, the Lin constraint cannot be imposed. Putterman agrees that $I_2$ in Edwards's paper is classical and that the steps of the derivation from it, including the neglect of the Lin constraint, are permitted. However, Putterman maintains that neglecting the Lin constraint is a subtle but crucial quantum-mechanical assumption in an otherwise classical derivation.

Edwards proceeds to sketch the reasons why collisionless classical fluids require that the Lin constraint be neglected and consequently are also governed by the London equations: For a charged fluid whose particles are in collective motion but whose collision cross-sections are small, the momentum transferred by a photon to an element of volume would remove it from the collective system and hence its path could not be followed. The Lin constraint would not apply, but the London equations would. This argument can be extended to collisionless plasmas where the particle energy is high and the organizing magnetic fields are high.

Edwards calls upon us all to study the extensive work on collective motion by Pines and Bohm (1952) in order to develop quantitative criteria for systems to which the Lin criteria would not apply. Now, here is where Edwards lands a high-scoring swat!

> The possibility that there exist classical systems which require the London equations, as I have suggested, justifies considerable effort to experimentally establish or disconfirm the idea, because, if such systems do exist, e.g., thermonuclear plasmas, attempts to explain or control their behavior without using the London equations would be as contrived and ultimately unsuccessful as were early efforts to explain superconductivity without the London theory. In addition, the classical explanation of effects in superconductors and superfluids would contribute to our understanding of the foundations of quantum theory.
>
> Positive experimental evidence is accumulating including flux ropes in the Venus ionosphere, Saturn's ring structure [Edwards, to be published], and filamentation in the plasma focus [Vahala, to be published]. It is hoped that a direct laboratory test will soon be made.

There appears yet another criticism, by J.B. Taylor (Taylor 1982), who briefly reviews work of Edwards and is not satisfied with Edwards's handling of the Lin constraint. Of Edwards's reference to the existence of magnetic structures in laboratory-produced plasmas (as occur in Type II superconductors), Taylor states, "There are several other explanations for these structures." This reviewer will return to this point soon.

Now Prof. V. Nardi, a colleague of this reviewer, has called attention to the 1965 paper by H. Fröhlich (Fröhlich 1966). This paper is so short and sweet that it should be read in its entirety by all interested parties. Fröhlich shows that the hydrodynamic equations of compressible fluids together with the London equations lead to the macroscopic Ginzburg-Landau equation, and that in the presence of many fluxons, all relevant equations can be expressed with the aid of the velocity potential $\Phi$ and the macroscopic parameter $\mu$ (= electric charge density/mass density) without involving either phase or microscopic constants. Thus, while in the 1980s Edwards and his adversaries have been vigorously engaged in the battle to decide theoretically how many angels can dance on the head of a pin, Fröhlich had already settled the question theoretically in 1965. Nardi discusses this matter at some length (Nardi 1983).

Now, to return to the remarks of J.B. Taylor that "there are several other explanations for these structures" (that is, the paired flux tubes that are seen in plasmas carrying a high current): this reviewer cannot restrain himself from making the remark, "Oh, are there now?!" This author earnestly encourages Taylor to step forward with the details of such explanations, which will undoubtedly provide a particularly succulent appeal for some of the relativistic electron beam and plasma focus physicists who have never themselves observed vortex structures with their apparatus or found these structures in their computer simulations or faced them in their theoretical analyses. Prof. Taylor, who has already captured single-handedly the cheering sections of the spheromak and field-reverse-pinch community (apparently without giving reference to or being slowed down by the competition from the prior work of Beltrami, Chandrasekhar and Woljer, Nardi, Wells, and Norwood) could now considerably enlarge his clientele into these electron beam and plasma focus communities. We urge him not to pass up such a golden personal opportunity, while at the same time he could be enlightening us all about the proper explanations for the experimental effects discussed in the last chapter of *The Morphology and Cosmology of the Electron* (Bostick 1985a) and in the references contained in that chapter, and in the recent article "Comments on the Experimental Results of the Plasma Focus Group at Darmstadt, FRG" (Bostick 1985b). In particular we are interested in J.B. Taylor's explanation for the filaments observed in the current sheath of the plasma focus and the "vacuum" diode of a relativistic electron beam machine, and for the damage patterns produced on witness plates when a relativistic electron beam passes through gas at about 1 torr pressure.

### The Author's Hypothesis

The author and his colleague at Stevens, V. Nardi, believe that the paired filaments observed in the current sheaths of the plasma focus and "vacuum" diodes and the structures form by propagating relativistic electron beams through gas at $\sim 1$ torr are actually paired Beltrami-morphology vortex structures that are carrying a current far in excess of the Alfvén limit, because collectively they have created a magnetic structure where everywhere the electrons are flowing parallel to the local magnetic field. We claim that these vortices display the paired flux-tube morphology of Type II superconductors. We do not claim that the current in these vortices will flow *dissipationless* for $10^{10}$ years as they might in an ideal Kamerlingh Onnes experiment, but these vortices have *less* resistance than any other morphology and they spring into life in a picosecond time scale and live to do the job for many microseconds if necessary.

Indeed, when they all suddenly chop each other off, like a meat-cleaver action, at one value of $z$ (where $z$ is the coordinate parallel to their axes) we have the "plasma focus." It is equivalent to a Type II superconductor going suddenly to normal conduction: This action suddenly (on a picosecond time scale) vastly increases the resistance across the cloven gap left by the "meat cleaver" through the instrumentality of the magnetic insulation produced by the large displacement current that takes the place of the conduction current. The voltage across the gap can go up to 15 MeV and accelerate deuterons to 15 MeV through this "electromagnetic ram" action.

This text also suggests that the damage patterns produced by a relativistic electron beam traversing a gas at 1 torr or traveling next to a dielectric guide tube surface are the result of vortex structures (similar to Type II superconductor morphology) that can travel for several meters without completely losing their properties.

This author further hypothesizes that the diamagnetic vortices (whose axes line up *parallel* to the background magnetic field) are the macroscopic embodiment of a Type I superconductor: If there is rotational free energy available, the plasma will condense into rigid-body, minimum-free-energy, rotating, islandlike vortices that expel some of the magnetic field from their own volume into the vacuum region which has been created between the vortices by the condensation of the plasma into vortices. If the available change in free energy associated with the condensation of the plasma into rigid-body, minimum-free-energy vortices is great enough a plasma vortex can expel a large part (if not all) of the magnetic energy contained within its volume.

It is asserted that the plasma vortex structures are able to simulate the morphology of Type I and Type II superconductors because the "organized energy" of the ions and electrons in these structures far exceeds the "disorganized" or thermal energy, and that the transition from disorganized turbulence to organized vortex structures is a phase transition involving condensation without the rise of temperature.

With the foregoing interpretation of the experimental observations, this text claims that the experimental effects are a justification of the concept that superconductivity in macroscopic, classical physics can essentially exist even though quantum effects have not been invoked to canonize the process. This author welcomes the opportunity for dialogue on competitive interpretations of the experimental observations.

## References


Bostick, W.H. 1985a. *International Journal of Fusion Energy* **3**:9 (Jan. 1985).

———. 1985b. *International Journal of Fusion Energy* **3**: 68 (Jan. 1985).

deVegvar, P.G.N. 1982. *Phys. Rev. Letters* **49 (6)** (Aug. 9).

Edwards, W. Farrell. 1981. *Phys. Rev. Letters* **47 (26)**: 1863.

———. 1982. *Phys. Rev. Letters* **49 (6)** (Aug. 9).

Edwards, W. Farrell, B.F. Edwards, and R.C. Thompson. To be published.

Fröhlich, H. 1966. "Macroscopic Wave Functions in Superconductors." *Proc. Phys. Soc.* **87**: 330.

Geurst, J.A. 1980. *Physica (Utrecht) B* **101B**: 82.

Henyey, Frank S. 1982. *Phys. Rev. Letters* **49**: 416.

Lin, C.C. 1963. In *Liquid Helium*. Ed. G. Careri. New York: Academic Press.

Nardi. 1983.

Pines, D., and Bohm, D. 1952. *Phys. Rev.* **85**: 338.

Putterman, S. 1982. *Phys. Lett. A* 89A: 146.

Segall, B., L.L. Foldy, and R.W. Brown. 1982. *Phys. Rev. Letters* 49 (6): 417.

Taylor, J.B. 1982. *Nature* 299: 681 (21 Oct.)

*The Times*. 1982. "Classical Physics Makes a Comeback." London, 14 January.

Vahala, G., and Vahala, L. To be published.


# The Electromagnetic-Ram Action of the Plasma Focus As a Paradigm for the Production of Gigantic Galactic Jets and Cosmic Rays

**by Winston Bostick**

An article appears in the January 1985 issue of the *International Journal of Fusion Energy* titled "Recent Experimental Results of the Plasma-Focus Group at Darmstadt, West Germany: A Review and Critique," in which this author sets forth a rationale for explaining the electron beams and ion beams produced by the plasma focus in terms of a stuttering electromagnetic ram. The author further points out that the measured energy spectrum of the deuterons in the deuteron beam produced by the Stevens plasma focus can be approximately represented by $N(E) \sim E^{-2.7}$ between 1 MeV $< E <$ 13 MeV. This is of the same form as the measured energy spectrum of the primary of the cosmic rays between 1 GeV $< E < 10^6$ GeV. The measured energy spectrum of the electrons accelerated in Raudorf's "electronic ram" (a somewhat different kind of electromagnetic ram from that of the plasma focus) in the range 1 MeV $< E <$ 14 MeV is $N(E) \sim E^{-3}$ (Raudorf 1951, 1974).

The article in the IJFE explains the operation of the plasma-focus electromagnetic ram in terms of a sudden interruption of a column of current that was otherwise peacefully flowing in pairs of force-free, Beltrami-like, vortex filaments: The bursting of one filament leads to a rapid burst of all the filaments in a cincture around the column. The electrons in this cloven region where the filaments have been severed must now face the Alfvén limit of 17,000$\beta\gamma$ A: The gap becomes magnetically insulated by the displacement current $I^{displ}$ within the gap and can conduct current only near the geometrical axis where the magnetic field $B_\theta \rightarrow 0$. The gap becomes a "plasma capacitor" that is charged to ~15 MeV by $I^{displ}$ in a matter of a few picoseconds. At the peak of the voltage across this plasma capacitor, the displacement current reverses sign; the electron and ion beams are accelerated by the voltage across the capacitor and discharge the capacitor with a kind of $RC$ decay, where $R$ is rapidly decreasing as a result of the rapid enlargement of the beam diameters as the reverse displacement current increases in magnitude. The result is an energy spectrum of the form $N(E) \sim E^{-3}$. When the capacitor is discharged, an interruption of the current again occurs and the cycle repeats a number of times, with a spacing between pulses of about 25 psec and a duration of the individual electron beam pulses of about 4 psec.

The author, who has been known for many years through his attempts to explain astrophysical and geophysical phenomena on the basis of paradigms that he observes nature to reveal in laboratory-produced plasma structures, quite expectedly comes forward with the suggestion that the electromagnetic-ram action of the plasma focus is trying to tell us how cosmic rays acquire their energy. It will be only natural for those theoretical astrophysicists who are steeped in statistical mechanics and turbulent processes, and who are now having a love affair with the black hole, to scoff at such a suggestion. But this author, undaunted, plunges even further into this cosmical question: He has the audacity to suggest further that the gigantic galactic jets in the active galaxies such as are now being observed by the computer-synthesized data of the radio signals at a number of wavelengths with the Very Large Array radio telescope in New Mexico, from radio galaxies like Cygnus A and Centaurus A (NGC 5128) (Burns and Price 1983, Burns et al. 1984), are being produced by an electromagnetic-ram action similar to that of the plasma focus; and further, that this action is producing not only these spectacular jets, but also the acceleration of the cosmic ray onta at the same time in the same accelerating gap. (See Figure 1.)

This author, at the risk of affronting both fusion plasma theoreticians and astrophysical theoreticians, is suggesting that the concept of the "stuttering electromagnetic ram" packs a triple whammy!

(1) The stuttering electromagnetic ram accounts for the beam-producing features of the plasma focus and for a large fraction of the neutrons it generates with the D-D reaction.

(2) The stuttering electromagnetic ram generates the gigantic jets observed in galaxies like Cygnus A.

(3) The stuttering electromagnetic ram accelerates onta (mostly protons) to cosmic-ray-spectrum energies in the same gap where it is accelerating the gigantic jets. And, of course, the stuttering electromagnetic ram simultaneously accelerates electrons to high energies and these high-energy electrons produce cosmic gamma radiation.

To justify and quantify these suggestions it is necessary to understand something about the genesis and morphology of galaxies. To this end, the author refers the reader to an article titled "Possible Hydromagnetic Simulation of Cosmical Phenomenon in the Laboratory" (Bostick 1958). This paper describes, among other things, how "barred-spiral" plasma configurations can be produced in the laboratory and how such results can suggest a hypothesis for astronomical galaxy genesis and dynamo action. The paper also suggests that magnetic repulsion between galaxies can account for the "expanding universe," in this respect rendering the "big bang" doctrine unnecessary. Although the paper is ancient and should be brought up to date in a few details, in this author's opinion it is the only seminal and acceptable hypothesis for the genesis of barred-spiral galaxies that has yet been advanced. The author has listened to some attempts by astrophysical theoreticians to explain the genesis of barred-spiral galaxies and the genesis of the gigantic galactic jets. Most of these attempts would remind a witness of the efforts of a mariner in a boat without a rudder, without charts, and without a compass. The mariner does not know how to get where he wants to go. This author is making the point that the plasma physics experiments in the laboratory will frequently point the way toward the winning hypotheses that will save the "mariner" a great deal of time and effort (perhaps an infinite amount of time and hence *his very life*) in achieving his goals. The experience of the great astrophysicist Birkeland should be noted in this connection (Peratt 1985). The Birkeland current filaments are the same as the Beltrami-type of vortex filaments we have been studying in the plasma focus for the last 20 years.

This article in the *Review of Modern Physics* (Bostick 1958) should, however, be brought up to date in one important respect: We have now achieved magnetic fields in the laboratory of the order of a gigagauss and, therefore, we can more properly scale magnetic fields, as well as time and distance in going from the laboratory to the cosmos.

There are, in addition, two important features of the "galaxies" produced in the laboratory that are not mentioned in the above paper.

First, stereo pictures show that the plasma jets that are fired across a magnetic field through a conducting medium are helical in shape. As a result, the "barred-spiral" galactic configuration formed by the jets coming together, head-to-head, exhibits spiral arms whose tips do not lie in the galactic plane but are deflected out of the galactic plane. An examination of astronomical photos of galaxies shows this also to be the case as it certainly is for Centaurus A in Figure 1.

Second, the laboratory-produced barred-spiral galaxies have spiral arms that are forked at the tips. On the photos of astronomical barred-spiral galaxies a filamentary forking of each arm into two tips is frequently observed.

Anthony Peratt has, 25 years later, produced in the laboratory in collaboration with Oscar Buneman and with computer simulation, barred-spiral configura-

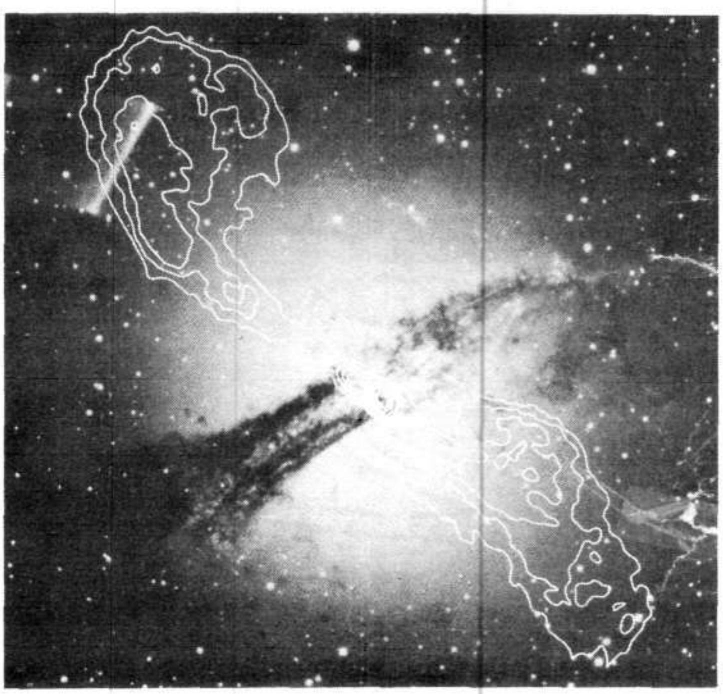


**Figure 1.** Map of radio emissions from the central region of Centaurus A is superimposed on a photo of the galaxy. A disk-shaped dust lane bisects the galaxy. The radio structure emerges from the center of the dust lane, more or less along its axis of rotation.
SOURCE: Jack O. Burns and R. Marcus Price, "Centaurus A: The Nearest Active Galaxy," *Scientific American*, Nov. 1983, p. 58.

tions similar to the author's (Peratt 1980, 1983, 1984, 1985). These three-dimensional computer simulations replicate the astronomical morphologies of barred-spiral galaxies with incredible fidelity, including the delicate features of helical spiral arms and forked tails. This author is deeply indebted to Peratt and Bunemann, whose high-fidelity simulations have resurrected, confirmed, and validated the 30-year-old galactic genesis hypothesis that appears in the 1958 paper.

As already stated, this hypothesis can explain, among other things, the expansion of the universe without the *ad hoc* invocation of the "big bang." Peratt et al. with their three-dimensional code also explain quantitatively the 3° K. background cosmic microwave radiation in terms of continuous synchrotron radiation by electrons. With these master strokes of their three-dimensional code, they have brought into question the Aristotelian logic of the "big bang" hypothesis. They are shaking the astrophysical scientific community with a shock wave as potent as the "big bang" itself. They can rightfully take their place among the foremost astrophysicists and cosmologists of all time!

With their three-dimensional code, Peratt et al. (1985) can explain the anomalously large red shift of quasars by the entraining of gas into the Birkeland current filaments: As the gas flows inward toward the center of the filament, it will always produce a red shift to the observer, if radiation from the far side of the filament is obscured by dust. This entrapment of gas into the vortex filaments (Birkeland-Beltrami filaments) of the plasma focus (the filaments act like vacuum cleaners) has been known for the past 20 years to occur in the plasma focus.

Figure 2 [Figure 8(d) of the author's 1958 paper] shows only the top hemisphere of the galaxy. The lower hemisphere would be a replication of the top hemisphere, and each hemisphere will have a column of current along the galactic axis near the plane of the galaxy. These currents will be in opposite directions in the two hemispheres, and will far exceed the 17,000$\beta\gamma$-A Alfvén limit. The current in these columns will be carried by large clusters of paired, force-free Beltrami-like vortex filaments (as in the plasma focus). The columns of current will be most highly concentrated near the galactic plane. It is there that each current column (one on each side of the plane) will undergo the action of the stuttering electromagnetic ram effect. Each one will project positive ions (mostly protons) in one direction and electrons in the other direction. Since the current columns are flowing in opposite directions, the beams of protons will be ejected in opposite directions along the galactic axis. The same is true of the electron beams. Both electron beams and proton beams are expected to be segmented by the cyclical, stuttering action of the ram, as in the plasma focus. The gigantic galactic jets are observed to have "beams" that are divided into segments, as does the plasma focus. The electron energy spectrum of the gigantic jets (as inferred from the

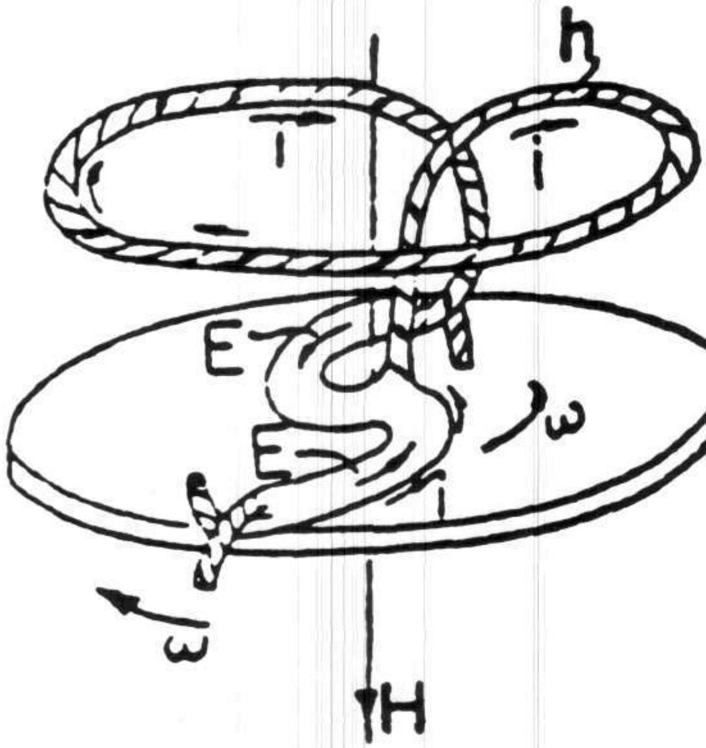



**Figure 2.** A suggested configuration for how regenerative dynamo action can occur in the galaxy to produce a galactic magnetic field that increases as time goes on. Electric current flow from galactic center to periphery via the galactic halo is indicated to show how the initial magnetic field $H$ can be reinforced.

synchrotron radiation) is of the form $N(E) \sim E^{-3}$, and this is approximately the same as that for the ions and electrons from ram effect in the plasma focus, and for the electrons in Raudorf's ram.

The galactic jets apparently extend to much higher energy levels for protons than the ram effect is producing in the plasma focus: In the electron beams from the plasma focus the electron energies are relativistic for electrons but the ion beam energies (<15 MeV) are not relativistic for the ions. In the galactic jets apparently some of the ions as well are relativistic ($\gamma \rightarrow 4$), but an average speed of the jets corresponds to a $\beta \cong 0.02$.

We would expect the same general type of organized force-free vortex structures to appear in the segments of the galactic jets as are observed in the electron beams from the plasma focus (Nardi 1980).

The energy for generating the galactic jets (if the ram effect is responsible) comes from the magnetic energy stored in the vast columns composed of paired, force-free, Beltrami-like vortex filaments. Most astrophysicists *a priori* prefer a more fashionable, exotic energy source, such as a stuttering, belching black hole. This author would prudently caution that no one has ever seen a belching black hole, or even a quiescent black hole, but we *have* observed the stuttering electromagnetic ram in the laboratory. Also, the Birkeland-Beltrami filaments are observed in the laboratory and appear in the three-dimensional simulations.

The author concedes that numbers must be applied to this ram hypothesis in the astrophysical context and he invites the collaboration of astrophysicists in this endeavor. It would be unfortunate if the astrophysical profession should choose to ignore the ram hypothesis, as apparently they have ignored the galactic genesis hypothesis in the 1958 Bostick paper. It would be dimaying, indeed, to this author to witness professional astrophysical mariners losing their very lives without achieving their goals because they would not pay heed to a hypothesis flashed to them by a foreign (but friendly), ancient mariner in an unregistered boat, but nevertheless a mariner who has seen the plasma physics laboratory paradigms of solar flares, galaxies, solar prominences, convection rolls of sunspot penumbras, striated tails of comets, density fluctuations in the ionosphere due to diamagnetic plasma vortices that cause the twinkling of radio stars, pairs of diamagnetic vortices produced by exploding barium canisters or a nuclear weapon at high altitudes in the Earth's magnetic fields, the paired flux tube structures of Type II superconductors, and the "CP" conservation of beta ray decay.

The stuttering electromagnetic ram model for the gigantic jets fits very well for the general geometric configurations observed. The crucial question is, where does the energy come from? Again, the plasma focus can be the paradigm for the answer to this question: The energy for accelerating the positive ions and electron beams in the plasma focus comes from the magnetic energy stored in the force-free current (~0.5 MA) column, in the electrodes and in the capacitors in which all the energy was originally stored. The kinetic energy of the electrons in the ionized current column is about 10 eV and the kinetic energy of the deuterons about 50 eV. When an electron or ion enters the accelerating gap of the ram it can, in a picosecond time frame, jump into an $N(E) \sim E^{-2.7}$ energy spectrum that extends to 15 MeV. For the ions and electrons that end up toward the top of this spectrum, there is a $\times 10^6$ factor in their energy increase! The energy for the many cycles of ion and electron acceleration at the gap is drawn from the "electrical power grid," the magnetic energy stored in the plasma current column, the electrodes, and the storage capacitors.

The energy for accelerating the protons and electrons in the galactic jets (in the stuttering electromagnetic ram model) comes from the "electric power grid" involving the enormous flywheel-kinetic-energy of the rotating barred spiral, which is a homopolar generator that is storing magnetic energy in the countless Beltrami-like, force-free, current-carrying filaments that compose the pattern of current circulation suggested in Figure 2. The original source of this electric-power-grid energy was accumulated over perhaps $10^{10}$ years while the self-gravitational process of attracting a dispersed mass of $M = 3\times10^{13}$ solar masses $= 6\times10^{46}$ g from a large volume to a smaller volume of radius $r = 20$ psec $= 20 \times 0.6 \times 10^{19}$ cm. The energy made available for storage in the electrical grid and flywheel by this gravitational transition is $E \cong GM^2/r = 2\times10^{66}$ ergs.

Now the electromagnetic energy radiated by an ordinary galaxy of $M = 3\times10^{13}$ solar masses is about $10^{41}$ ergs/sec. An active galaxy like Centaurus A with the large jets radiates $10^6$ times as much or $10^{47}$ ergs/sec. At this rate of radiation of energy (even with *no* further gross gravitational condensation), Centaurus A would be able to sustain this rate for $10^{19}$ sec $= 3\times10^{11}$ years, which is an order of magnitude greater than the age of the universe. Indeed, the percentage drain of this radiation on the resources of the galactic electrical grid (it should really be called a magnetic grid because homopolar generators are naturally low impedance devices) could be compared to the effect of one camel taking a drink on the flow of the Nile (that is, before the construction of the high Aswan dam).

The fact that the high-energy electrons—which are producing the synchrotron radiation that makes the gigantic galactic jets "visible" to the radio telescopes—are constantly being reenergized as they proceed outward with the segments of these jets should be no mystery to physicists who have worked with the plasma focus. The plasma focus is observed to manufacture plasmoids, which contain magnetic energy produced by circulating currents that are flowing in Beltrami-type, force-free filaments. These plasmoids are projected through space and connect this magnetic energy that can be released gradually or suddenly to the electrons and positive ions of which the plasmoid structures are composed (Nardi 1980). The same process is very likely occurring in the galactic jets.

The laboratory experiments with colliding plasma jets, discussed in the author's 1958 paper, also yield an explanation that accounts for the observed fact (both astrophysical and laboratory-produced) that the "bar" in the barred spiral has the morphology of a straight, taut cable: The two jets as they draw close to one another with their two induced electromagnetic frequencies start driving a current pattern that flows through self-organized Beltrami-type vortex filaments that have tensile strength. It is these tensile filaments under tension that are stretched into a straight line as they provide the centripetal force to hold the leading ends of the two spiral arms in orbit. This bar then becomes a major part of the armature of the homopolar generator, which supplies the electromagnetic frequency for the galactic current pattern. This current pattern requires a confluence of the current filaments at the center of the bar, where they turn 90°, some going north along the galactic axis and some going south. At this 90° turn, the filaments very likely are locked together in some kind of twist so that the entire diameter (that is, the entire length of the bar) has tensile strength. It is to be expected that there would be some clumping of the plasma and magnetic fields at the position of the 90° turn as well as the twisting, something like a knot in a rope. Such knots or nodules are frequently seen in the centers of the bars in astronomical galaxies and in Peratt's simulations.

## References


Bostick, W. 1958. "Possible Hydromagnetic Simulation of Cosmical Phenomena in the Laboratory." *Rev. Modern Phys.* **30**(3): 1090 (July).

———. 1985. "Recent Experimental Results of the Plasma-Focus Group at Darmstadt, West Germany: A Review and Critique." *Intl. J. Fusion Energy*. **3**(1): 68.

Burns, J.O., and R.M. Price. 1983. "Centaurus A: The Nearest Active Galaxy." *Sci. Amer.* **249**(5): 56 (Nov.).

Burns, J.O., et al. 1984. *Astron. J.* **283:** 515 (Aug. 15).

Peratt, A., et al. 1985. "Companion Galaxies Match Quasar Red Shifts: The Debate Goes on." Accepted for publication in *Physics Today*.

———. 1985. *Phys. Rev. Lett.* "Microwave Generation from Filamentation and Vortex Formation within Magnetically Confined Electron Beams." **54 (4):** 1167.

———. 1984. *Sky and Telescope*. **67:** 118.

———. 1983. *Astrophys. and Space Science*. **91:** 19.

———. 1983. *Sky and Telescope*. **66:** 19.

———. 1983. *J. Appl. Phys.* **54:** 11.

Nardi, V. 1980. *Phys. Rev. A.* **22:** 224 (Nov.)

Raudorf, W. 1974. "Acceleration of Electrons by an Electrodynamic Space Charge Effect." *Proc. of the International Conf. on Energy Storage, Compression, and Switching*, Asti Torino, Italy, Nov. (New York: Plenum), p. 381.

———. 1951. "An Electronic Ram." *Wireless Engineer* (July), p. 51.

## Nonlinear Pumping for X-ray Lasers: The Rhodes Experiments

by Charles B. Stevens

A totally new possibility involving nonlinear pumping processes for the X-ray laser is being explored by Dr. Charles Rhodes of the University of Illinois at Chicago to produce efficient X-ray lasers. Rhodes has attained inner shell excitations in atoms and molecules in a manner that seems to contradict all existing theoretical models of the atom and quantum physics. According to Dr. James Ionson of the Strategic Defense Initiative Office Division of Innovative Science and Technology, the Rhodes nonlinear pumping method could be demonstrated before the end of 1985.

Rhodes's experimental success opens up new vistas for X-ray lasers, but at the same time underscores the present dearth of knowledge on how these processes actually work. Normally, efficient pumping of the high energy level, inner atomic electrons needed for X-ray lasing is limited to energy inputs that are qualitatively comparable to the output. Utilizing longer wavelength inputs, such as radiation at wavelengths longer than that of X-rays, leads to large wastage. This is because the outer electron energy levels interact most strongly with the longer wavelengths and emit radiation in the longer wavelengths that escapes the lasing medium. That is, only a small portion, if any, of the input will end up pumping the required inner electron orbitals. One way around this is to ionize the atoms so that only the inner electrons remain. But this involves a lot of energy and makes X-ray lasing quite inefficient.

In a number of simple experiments, Rhodes showed that contrary to accepted theory, long-wavelength radiation could efficiently pump inner electron levels. Recent developments in producing high-power, extremely short pulses of extreme ultraviolet laser light (0.193-micron) provided the key technology for Rhodes's experiments. Once these experiments established the new phenomenon at this wavelength, it could then be shown that similar processes were occurring at longer, 1.06- and 0.53-micron wavelengths.

Rhodes irradiated various gases with 5-picosecond, 3-billion-watt pulses of 0.193-micron argon-fluoride laser light. The pulses were focused to intensities of 1,000 trillion to 100,000 trillion watts per square centimeter in the experimental volume of the gas. Stimulated emission from the irradiated gas at wavelengths shorter than 0.193 micron was measured. More than 1 percent of the input energy was measured in each of the stimulated emission wavelengths. This means that this pumping method could be *100 million times* more efficient than the Livermore collisional pumping method, where less than $10^{-11}$ of the input optical light ended up as X-ray laser output.

A wide range of elements were irradiated, demonstrating that the absorption-emission was highly dependent on the atomic shell structure.

### The Failure of Existing Theory

According to Rhodes, the data show that the atomic outer shells are absorbing many photons and conveying this input to inner shells in a manner apparently in contradiction to present theoretical models. Rhodes noted that:

> The data strongly indicate that an organized motion of an entire shell, or a major fraction thereof, is directly involved in the nonlinear coupling. With this picture, the outer atomic subshells are envisaged as being driven in *coherent* oscillation by the intense ultraviolet wave. An immediate consequence of this motion is an increase in multiphoton coupling resulting directly from the larger magnitude of the effective charge involved in the interaction. In this way, a multielectron atom undergoing a nonlinear interaction responds in a *fundamentally different fashion* from that of a single electron counterpart. The strong highly nonlinear coupling which develops between the radiation field and the atom can result in the transfer of energy by a direct *intra-atomic* process to inner-shell excitations. . . . Although all standard theoretical approaches fail to provide a description of the observed phenomena, a relatively simple model, valid at sufficiently high intensity, can be contemplated.

How is the phenomenon explained? First, the time scale of the interaction is such that interatomic processes like collisions must be excluded. Next, it appears that the interaction between the outer shell electrons and the ultraviolet laser light is like that found in plasmas, where the electrons are not bound to atomic shells but are free electrons, with the stipulation that the electrons are forced to follow a restricted path defined by the shell's orbit! The result, according to Rhodes's "simple model," is that the motions of the outer electrons produce giant current densities in the range of 100 to 1,000 trillion amps per square centimeter. These nonquantum outer-shell electron currents apparently produce the efficient excitation of the inner electron orbitals.

### Higher Levels of Interaction

Another possibility is that the Rhodes experiments are also revealing a new level of energy transfer on an atomic scale. Specifically, spectroscopic analysis of his findings shows that three excited levels of xenon gas coalesce to form another level that has a different spectroscopic line. This violates the Pauli exclusion principle, which holds that three electrons cannot have the same orbit. Apparently, however, three levels of xenon at outer orbits form a new orbit together at a higher energy level. With the increase in energy, this new, unstable orbit gets closer to the nucleus.

In particular, the high current, plasma-pinch-like atomic processes that Rhodes has derived from his experimental results may be taking place in all lasers, but the time scale for this process may have been too short to readily detect it.

One explanation suggested at a recent Fusion Energy Foundation seminar on astrophysics by Lyndon H. LaRouche, Jr., was that this higher level of electromagnetic interaction could be responsible for all lasing, and that the quantum processes, previously thought to be determinate, may be only acting as a medium to retard the evolution of this higher level of electrodynamic action.

Rhodes points out that when intensity levels of $10^{19}$-$10^{20}$ watts per square centimeter become available in the near future with the development of femtosecond rare gas halogen lasers, the irradiated electrons would act as completely free particles accelerated in the intense electric field of the incident laser beam. This could make possible the efficient pumping of a range of X-ray laser wavelength inner energy levels. Furthermore, Rhodes notes, this new process could be very important in providing a picture of how rotating neutron stars accelerate cosmic rays.

### References


T.S. Luk, et al. (undated). "Observation of Stimulated Emission in the 119 to 149 nm Range from HD Excited by Picosecond 193 nm Radiation." Submitted to the J. of Chem. Physics.

C.K. Rhodes. 1984. "Studies of Collision-Free Nonlinear Processes in the Ultraviolet Range." Invited Presentation for International Conference on Multiphoton Processes III, Crete, Greece. Sept. 5-12, 1984.

———. 1985. "Studies of Multiquantum Processes in Atoms." Invited Presentation for Seminar on Fundamentals of Laser Interaction, Bundessportheim, Obergurgl, Austria. Feb. 24-March 2, 1985. To be published.


## Major Beam Advances Announced at University of Rochester Conference

**by Charles B. Stevens**

The University of Rochester Laboratory for Laser Energetics (LLE) hosted a conference April 17-19, "Lasers and Particle Beams for Fusion and Strategic Defense," which drew 300 scientists and engineers. The conference was the occasion for the announcement of major advances across the full spectrum of directed energy technologies: Researchers from Los Alamos National Laboratory reported major successes with their White Horse neutral particle beam program, their Aurora krypton fluoride excimer laser, and their free electron laser facility. From Lawrence Livermore National Laboratory significant results on both the Experimental Test Accelerator and Advanced Test Accelerator electron beam accelerators were reviewed together with free electron laser experiments on the Experimental Test Reactor. From Sandia National Laboratories, major breakthroughs with light ion beam focusing were reported.

### Los Alamos KrF

Dr. Damon Giovanielli announced that the Los Alamos National Lab krypton fluoride Aurora laser had achieved a record output of 10,500 joules at a power level of 20-billion watts. This provides the key step toward demonstrating the technology for multimegajoule, 100-terawatt systems needed for inertial confinement fusion. Giovanielli outlined how the recent Los Alamos work showed that krypton fluoride could attain 10 percent efficiency levels. He also reviewed Los Alamos experiments with the Antares carbon-dioxide laser on beam-defense-related damage experiments and microwave generation. In addition, Dr. Robert Hunter of Wester Research Corporation reviewed developments in optical multiplexing that will be used to bring the krypton fluoride laser up to laser fusion-power levels.

### Sandia Light Ion Beam

Dr. J. Pace VanDevender, director of the Sandia Pulsed Power Sciences, and Professor Ravindra N. Sudan, director of the Cornell University Laboratory of Plasma Studies, detailed the experimental and theoretical status of the light ion beam program. Experiments on the Proto I and Particle Beam Fusion Accelerator or PBFA I indicate that the Sandia PBFA II will be able to develop 10,000 terawatts per sq cm of lithium ion pulses with several megajoules total energy when it is completed in early 1986. Actual fusion experiments are scheduled for 1988.

In May 1984, Proto I achieved 1.5 trillion watts per sq cm. In 1985, PBFA I delivered 8 trillion watts to a 4.0- to 4.5-mm diameter spot. Proto I has a 1.4 MV and .4 Mamp output, while PBFA I has 2 MV and 4 Mamp output. PBFA II will demonstrate voltage scaling when 30 MeV lithium ions are used. The recent success with focusing, which does not appear to have any theoretical explanation, strongly indicates that PBFA II will come at a power density 100 times that originally expected for the machine.

### A Unique Combination

An open house together with the presentation of poster papers was held at the laboratory just prior to the conference, providing a unique opportunity to obtain an overview of the work of this national user laser facility.

The University of Rochester Laboratory for Laser Energetics (LLE) was initiated by the combined efforts of the university, New York state, and private industry in the early 1970s. Its unique capabilities were recognized when the federal goverment designated LLE as a national laboratory of the Department of Energy's controlled fusion research effort and the only one open to nongovernmental users. In fact, LLE is the most powerful and versatile laser research facility readily available to scientists throughout the world.

The 24-beam Omega laser puts out more than 12 trillion watts of laser light. Recently it has been converted to the short wavelength, ultraviolet 0.351 micron—a laser wavelength that is called blue. The advanced laser and optical technology developed at the university has allowed the Omega to have among the highest repetition rates in the world for high-power lasers. The laser can be fired every 30 minutes. Other facilities, such as the 130-trillion watt, 10-beam Nova at Lawrence Livermore National Laboratory in California, generally are capable of only a couple of laser shots per day.

This makes the Omega very productive for research. And as the National Laser Users Facility, it has been made available to the entire scientific community, including researchers from academic and medical institutions, industry, and government laboratories. Over the past three years more than 21 user experimental programs have been carried out on the Omega.

Although it is primarily directed toward harnessing the unlimited energy potentials of laser fusion, the Rochester LLE has explored the full range of scientific and industrial applications of powerful lasers. Specifically, the lab's charter dedicates it "to investigate the interaction of intense radiation with matter."

### Current Work

The breadth of the LLE work can seen from these examples:

- Generation of bursts of radiation lasting less than a trillionth of a second (a picosecond) and ranging from the long wavelength microwave (inches) to the short wavelength X-ray (1 Angstrom ~ an atomic radius) region of the electromagnetic spectrum. When combined and synchronized with picosecond laser pulses, this provides a unique means for the examination and recording of fast, microscopic processes. This imaging diagnostic is utilized in observing solid-state physics processes—the science of computer microchips, imaging of biological specimens and metabolisms, and chemical reaction dynamics.
- Generation of picosecond electron beams and electric pulse detection, which provides a new way of imaging fast processes and resolving microscopic resonances in the structure of various materials.
- Development and application of ultra-high-speed (trillion times a second, trillion hertz) sampling systems.
- Creation of new types of optical materials and coatings for high-power, short-wavelength laser light and radiation.
- Short pulse, subnanosecond, high-resolution X-ray diffraction and crystallography.
- X-ray laser development and application.
- Demonstration of high-resolution X-ray lithography to print microchips.
- Development of high-repetition-rate, high-efficiency, high-power glass lasers.
- Applications of solid-state switching technology to picosecond, time-resolved spectroscopy for both biology and chemistry.

### Rochester Firsts

LLE has accomplished pioneering research in laser fusion itself:

- Early experiments demonstrating the presence of nonlinear "parametric" processes in the interaction of laser light and matter.
- First direct measurement of compressed fuel density in laser-driven targets.
- The first comprehensive measurements of harmonic and subharmonic emission from spherical laser fusion targets.
- Pioneering work on high-gain, low nonlinear index of refraction of phosphate glass for high-power lasers.
- Development of efficient nonlinear methods for upshifting laser light frequencies—making shorter wavelengths out of longer ones.

## The Biology of Light: New Insights into Life Processes

**by Wolfgang Lillge, M.D.**

Dr. Fritz A. Popp, a biological physicist from the University of Kaiserslautern, West Germany, published the book *Biology of Light, the Basis of Ultra-weak Cell Radiation*, in 1984 (Popp 1984). In it, he reviews specific experiments that he and other scientists around the world have conducted in the biology of light in the last 60 years.

Popp's view of biological processes differs significantly from that of most of his colleagues, who are mainly concerned with molecular phenomena; he stresses instead the importance of coherent electromagnetic radiation as a key regulating force inside and between the cells. His approach is interesting because it focuses on the developmental process of living matter and how this process changes, rather than narrowing the understanding of nature to simple molecular interactions, which at most roughly *describe* reality but cannot *explain* how the universe works.

Popp's work in this field lends new features to James Frazer's article, "New Frontiers in Biophysics" (IJFE, **3**: 63), if Popp's experimental results are put into the correct focus. Frazer demonstrated that the absorption and emission spectra of a variety of biological surfaces can be associated with longe-range coherence in numerous biological processes, especially membrane and DNA processes, and he proposed to study apparent complementarities (as in the DNA molecule) from the standpoint of logarithmic spiral work-function-type geometries.

In the effort to explain life on Earth as a process of achieving higher and higher states of order (negentropy) mediated by singularities at any branching point reached, simple mechanistic models cannot be relied on, as they are known from solid-body physics, quantum mechanics, or information theory. Thus, Popp's results—presented in this book and other earlier papers—are valuable exactly in so far as he avoids interpretation by means of that simplistic, misleading information theory of which I. Prigogine is a leading representative.

This article will review new insights into biological processes that represent highly structured, coherent biological work—the basis for life on our planet. Several other researchers have made valuable contributions to broadening knowledge about, as yet unexplainable phenomena in nature, including Prof. Philip Callahan of the Insect Attractants, Behavior, and Basic Biology Research Laboratory in Gainesville, Fla., and J.P. Biscar, Department of Physics, University of Wyoming. In their fields, these scientists have elaborated details of the broader reality that living processes are based on highly nonlinear resonance effects in narrow-band frequencies over a broad range of the electromagnetic spectrum.

### The Historical Perspective

One of the first experiments from which the hypotheses of specific light emission by plants resulted was conducted by the Russian biophysicist Alexander G. Gurvich in 1922. (Popp, p. 34). Gurvich observed a significant increase in cell divisions of an onion root when he brought a second onion root near to the first. He concluded that only ultraviolet emission could be the trigger for the increase, for this "mitogenic" effect could be stopped by ordinary ultraviolet-proof window glass, but not by quartz glass, which does not absorb in the ultraviolet.

Only after World War II were sufficiently sensitive photon counters developed for detecting the extremely weak emissions from animal and plant tissue. In laboratory work mainly plant seedlings such as wheat, beans, and corn were used. Apart from some Italian efforts, Russian researchers first recognized the significance of "ultraweak luminescence." In the 1960s and 1970s, Soviet science magazines regularly published reports of experiments with bioradiation. One of the experiments employed two cell cultures in separate quartz glass containers, one of which was either infected with a virus or poisoned with strong ultraviolet radiation or sublimate. In every case, the cells in the neighboring container showed the same symptoms of sickness soon afterward. When normal window glass was used, the transmission of the disease was not observed, which suggested the involvement of ultraviolet radiation in the process.

One Russian report on these experiments states (Popp, pp. 38-39): "The radiation of cells was measured with a photon multiplier. Normal living cells emit a steady flow of photons. This flow changes abruptly, when a virus enters the cell: explosion of radiation—silence—another explosion—slow decrease of emission in several waves, until the death of the cell . . . ."

Another field of intense Russian research was the effects of low intensity microwaves on biological samples (*Soviet Physics* 1974). Specific frequencies in the millimeter band were found to increase cell divisions in cell cultures by an order of magnitude, provided a specific threshold of intensity was reached. Characteristically, the low intensity radiation did not heat up the cells, and an increase in intensity beyond the threshold level did not result in a still higher yield. In West Germany, researchers associated with Prof. Genzel of the Max Planck Institute for Solid Body Physics confirmed these findings (Popp, p. 29). They observed drastic changes in the growth of yeast cultures irradiated with microwaves of specific frequencies. They concluded that nonthermal effects—biological resonances with unusually high resolution, that is, coherent electromagnetic couplings—were responsible for these effects.

These findings allowed a first, rough understanding of resonance effects in living matter because the microwaves absorbed by the cells induced molecules in some way to increase their specific internal vibration (resonance) and thus their ability to perform biological work.

### Popp's *Biology of Light*

In 1976, Popp and his coworkers began their own experiments in measuring the photon emissions of living plant cells, using chiefly cucumber seedlings (Popp, pp. 45-49). With an extremely sensible photomultiplier—which could detect a "fire-fly" at a distance of 7 miles—they essentially confirmed the findings of the Russian experiments. They poisoned the seedlings with heparin, for example, and found the same characteristic photon eruptions before the cell ceased to emit altogether. When, in the same experiment, the heparin antidote protamine was added, photon emission soon returned to its previous values.

Another experiment demonstrated the temperature dependence of the photon emission, showing that the photon intensity increased in the same way as other functions in the cell—membrane permeability, metabolism, and so on—with rising temperatures. Popp also irradiated seedlings with monochromatic light and found that the stored radiation was reemitted over a period of hours. He put his seedlings into a dark container, waited for two hours to allow any stored light photons to disappear, and then irradiated the sample for one hour with weak monochromatic light of varying frequencies (626, 695, 550 nm, and so on). The photomultiplier showed a decreasing curve whose half-life increased continuously over time. On average, photon emissions lasted for about two hours (Figure 1).

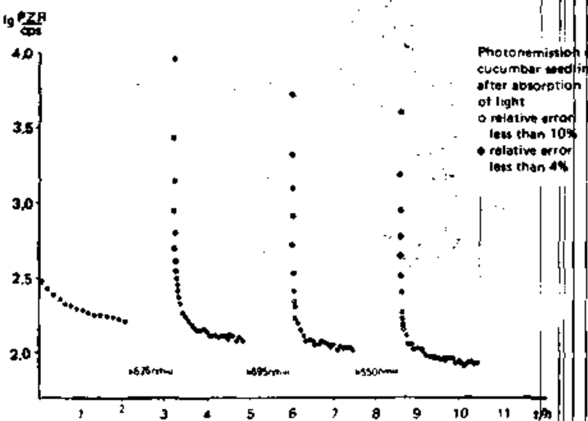



**Figure 1.** When cucumber seedlings, which are stored in a dark chamber, are then irradiated for two hours with weak monochromatic light (626 nm), the stored radiation decreases in a way that the half time becomes increasingly longer. This effect is also observed for different wavelengths (for example, 695 and 550 nm).

Popp compared the photon storage capability of the seedling to a Hohlraum resonator with efficiency (number of reflections) increased by a factor of 10 billion. Hohlraum resonators in microwave technology, according to Popp, may have half-lives of some milliseconds at most, which generally decrease exponentially. Assuming this kind of comparison from solid body physics really holds for living matter, this is an astounding phenomenon, which may be the basis for addressing several other questions.

With the same method of photon detection, Popp found a striking difference between normal and tumor tissue (Figure 2). After being irradiated with light of a specific wavelength, the tumor tissue not only exhibited a much faster decrease of photon emission than normal tissue over time, but also lost the ability of normal tissue to increase its rate of emission again after a second input of light of the same wavelength. Does tumor tissue then lack the ability to resonate coherent electromagnetic waves with the same efficiency as normal cells?

Frazer described a parallel phenomenon (IJFE, 3:63) when he reported that tumor tissue can be clearly differentiated from normal tissue, using Nuclear Magnetic Resonance technology or NMR. In NMR, strong

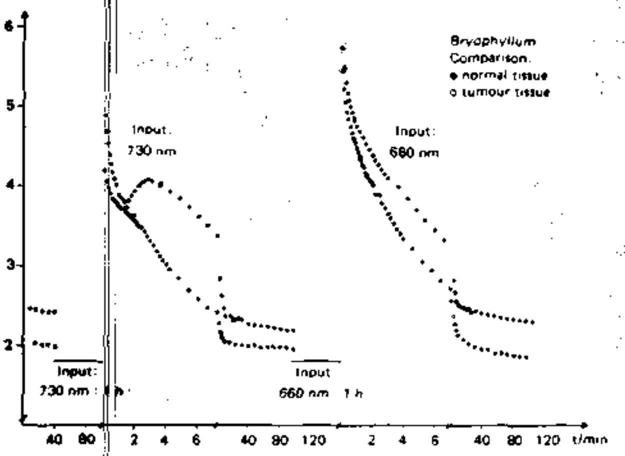



**Figure 2.** In contrast to what happens with dead matter, the intensity of emission after light excitation often shows a marked increase in biological objects. This seems to be also one of the most striking differences between normal and tumor cells of corresponding kinds. Tumor cells, besides showing a much faster decrease, are characterized by an almost total loss of resurgence of emission.

magnetic fields align the spins of hydrogen protons of water and other substances in the cells, and a characteristic oscillation of these protons is recorded when radio frequency beams are sent through the tissue. Generally speaking, the more regularly periodic the arrangement of protons, the more rapid the emission of radiation. Therefore, NMR measures something akin to long-range coherence of water and other substances present in living tissue. Frazer found that tumor tissue absorbs and emits at frequencies and with time distributions slightly different from normal tissue. He also seeks to use this same difference of absorption and emission for treating cancer by selectively generating hyperthermia in the tumor tissue.

Popp also studied the carcinogen 3,4-benzpyrene and noted that 1,2-benzpyrene, which differs from the former only through the arrangement of one benzene ring, shows no carcinogenic effect at all. According to "classical" theory, 3,4-benzpyrene damages DNA by way of a highly reactive intermediate product, an epoxide, which then somehow reacts with one of the bases in the DNA molecule and destroys the genetic code. It has never been demonstrated conclusively why the harmless 1,2-benzpyrene should not also function in this way. Popp and others reported another, much more fundamental difference between the two molecules: 3,4-benzpyrene exhibits a highly unusual energetic coincidence of its three lowest ultraviolet excitation states, while 1,2-benzpyrene does not. Other carcinogenic molecules are said to have similar properties.

3,4-BP may interfere with the ability of the cell to repair genetic damage. When cells are treated with weak ultraviolet radiation of a specific frequency, they themselves tend to repair DNA damage. If the cell itself has a similar kind of internal trigger for self-repair of DNA damage—perhaps much more efficient than crude external ultraviolet irradiation—it is possible that carcinogenic substances simply absorb the cells's internally produced ultraviolet light and block the self-repair of DNA damage.

Popp also studied the spectral pattern of ultraweak emissions from his cucumber seedlings over the visible light range, finding a nearly continuous band system with several species-specific maxima and minima (Figure 3). Even when the cells were manipulated with different substances, this characteristic band system remained, and just shifted up in intensity. This result, according to Popp, points to a large molecule (polymer) as the source of the photon emission that has strictly coupled spectral bands ("modes") and therefore has the same excitation energy and time of decay for all frequencies.

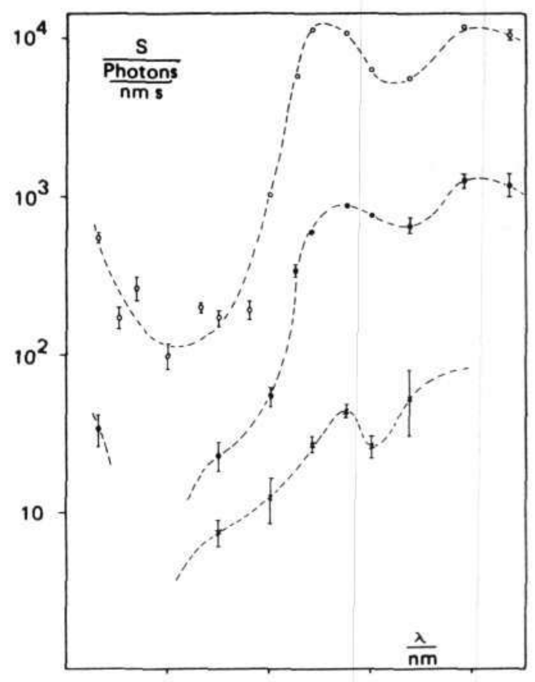



**Figure 3.** Ultra-weak cell emission shows an almost continous band system in the optical region of the spectrum, with several maxima and minima that seem to be species-specific. The lowest curve is untreated cucumber seedlings ("Chinese snakes"). It is notable that after application of cell poisons (middle curve Cialith, upper curve acetone) the intensity increases by up to the factor 1,000. However, the spectral distribution stays almost the same.

Popp and the theoretical physicist K.H. Li report that the curve of decrease of photon emission from tissue is hyberbolic rather than exponential, which can only be explained if oscillations can stabilize themselves on the basis of eigenfrequencies and store photons actively in a coherent way. Such a system, whose major component is suggested to be the DNA molecule and other macromolecules (proteins), is therefore able to perform work, using coherent electromagnetic waves (photons).

The ability to store photons and reemit them in a coherent way, proves to be one of the primary preconditions of life processes. The multitude of metabolic reactions requires a highly coherent flow of photons that provides the activation energy for these reactions at the exact location and point in time when they are needed. Such a biological photon field is superior to any thermal field in chemical reactions by the factor of $10^{40}$, according to Popp's calculations. These calculations are essentially based on the fact that in biological systems, the orbits of relevant molecules are always occupied with the same number of electrons each, regardless of their energy; that is, the occupation number is nearly independent of the wavelength. In this way, chemical reactions in the cell may be started and controlled on the basis of what is called the *least action principle*.

The Russian biologist Inyuschin even speaks of a *bioplasma* in the context of controlling photon fields (Popp, p. 62); in effect, this may be only a different way of describing the efficiency of the continuous transformation process taking place in living tissue. The loss of energy in biological matter is minimized to a point that, for example, light can be transported in plant tissue with an efficiency comparable to superconductivity in supercooled metallic fibers.

Another expression for such a work function is the observation made by various researchers that both chlorophyll and DNA molecules are capable of upshifting frequency by several orders of magnitude. A similar observation is reported by Popp: sometimes photons of shorter wavelength (that is, higher energy) were emitted by the cucumber seedlings than were used during the irradiation (Popp, p. 73).

### The Challenge of DNA

DNA, with its tremendous properties and geometries, is a natural object of interest in uncovering the actual processes of life. Not only has DNA the double-helical form—with sugars and phosphates constituting the two strands and the bases adenin, guanine, cytosin and thymin constituting the steps—but this threadlike molecule again is intertwined into other superstructures, super-helices, super-super-helices, and so forth, occupying a volume of only a billionth of a cubic centimeter in each human cell. If one were to unravel this structure, the DNA content of one cell would amount to approximately 2 meters in length, and the DNA of the human body would correspond to the diameter of our planetary system—about 10 billion kilometers.

The significance of the helical arrangement of DNA and its striking geometrical proportions in accordance with the golden mean, which give it the potential to mediate coherent electromagnetic action into biological work, has earlier been demonstrated ("Coherent Effects of DNA Explored," Ned Rosinsky, M.D., IJFE, **3**:1, p. 84). "Storing" photons and making them available for life processes in a coherent way is one aspect of the role of DNA in the organism.

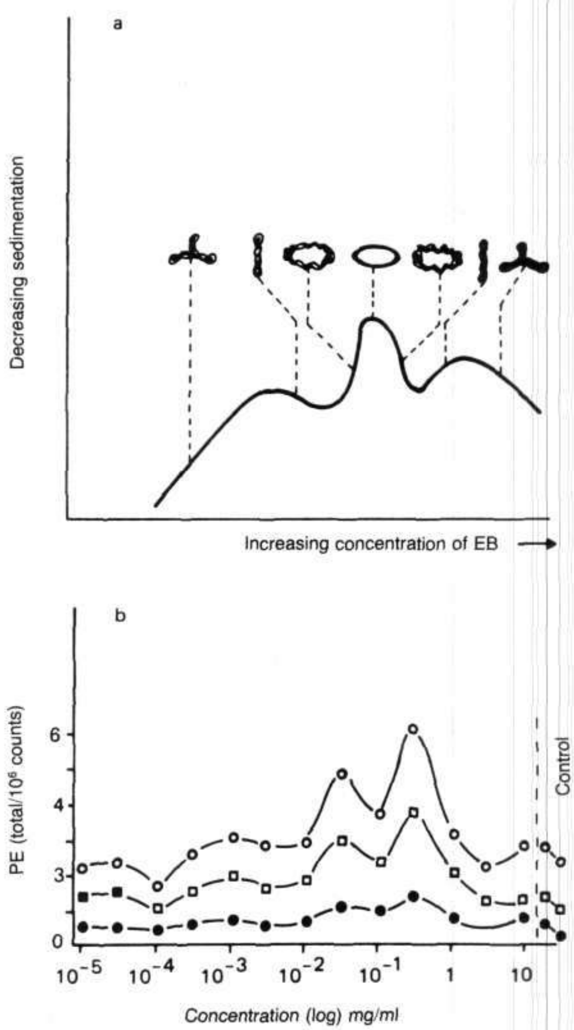



**Figure 4(a).** With growing concentration of ethidium bromide solutions more and more EB molecules are intercalated between the DNA base pairs. This intercalation leads to the unwinding of the helical DNA superstructures. Experimentally this can be established by DNA sedimentation. After the complete unwinding of DNA, a continuing intercalation does not lead to further dissolution of the DNA polymer, but to a renewed rewinding of the DNA helix structure; however the rewound structure has opposite spin.

**Figure 4(b).** When you observe ultra-weak cell emission after treatment with EB solutions, the total amount of emitted photons shows already after one hour (lower curve) indications of the same concentration dependency as in (a). This typical profile of (a) becomes even clearer after a longer time of registration (three and five hours), an indication that ultra weak cell emission depends on DNA conformation.

Popp and his collaborator M. Rattemeyer demonstrated this feature of DNA by means of a simple experiment: They correlated the unwinding and rewinding of DNA with a parallel increase and decrease of photon emission from DNA (Popp, pp. 86-89). Ethidium bromide, a red dye, is known to intercalate itself into the DNA molecule thereby unwinding its superstructure. This process depends on the concentration of the dye and does not lead to the total dissolution of DNA; rather, when a definite concentration is reached, DNA starts to rewind itself but with the opposite spin of its superspirals. In fact, the photon emission curve registered by Popp and Rattemeyer showed the same typical profile of that of the rate of DNA unwinding and rewinding effected by ethidium bromide (Figure 4).

The storing and emitting of light may have a definite physiochemical basis in the DNA molecule (Popp, pp. 91-93). It has been demonstrated that in DNA so-called exciplexes, or excited molecule complexes, are created when one of the molecules constituting the complex takes in a photon and stores it in a relatively long-living "metastable" state of excitation. This process was also described as an "excitation energy trap." As the first step, one of two neighboring molecules (one of the four bases in DNA) absorbs a photon and is transformed by that into an excited state. But rather than emitting the photon again, molecule $M_1$ reacts with neighboring molecule $M_2$ by emitting only a fraction of its excitation energy, forming a relatively stable complex. Only after a certain time period, the complex itself may decompose, emitting the rest of the energy it has stored.

This local exciplex formation is responsible not only for efficiently storing photons but also for stabilizing the DNA structure, since the attracting forces of the two excited molecules condense the bases of the whole macromolecule.

Not only DNA bases are capable of transforming themselves into exciplexes. The same is true also for RNA, proteins, chlorophyll, ATP, organic ring compounds and perhaps even oxygen and cell water. Given such a complex potential of coherent storage and emission of photons, it has been established by Li that transitions into exciplex states are synchronized and are thus capable of coherent photon emissions (Popp, p. 94). Popp and Li think the cell actually resembles a laser of very specific qualities, which continuously operates either just over, at, or below the so-called laser threshold.

Popp admits that there are strong counterarguments to this cell laser conception, yet in a simple experiment he demonstrates that biophotons at least have a high degree of coherence, the precondition for laser action. If sunlight hits a very small area on Earth, about $10^{-6}$ square centimeters—which roughly corresponds to the surface of a cell—this light becomes virtually coherent. To allow sunlight of the same intensity as that observed in photon emissions of biological samples to enter into the dark chamber of the photon multiplier, Popp cut a small slit in the chamber wall, not larger than the surface of a cell. The illumination of the interior space of the chamber thus amounts to illumination with laser light. Measurements showed that the biophotons from cucumber seedlings have at least the same—if not a greater coherence—than the relatively coherent incident light from the Sun (Popp, p. 141).

While the idea of a cell operating like a laser is fascinating, the complexity of organization on the microscopic scale of living processes may require a "technology" more sophisticated than a mere laser, even if one imagines it to work on the low-intensity level of biophotons.

A key aspect for the effectiveness of this highly coherent radiation, even if of very low intensity, is its spectral specificity, as Philip S. Callahan has demonstrated for the attraction of night-flying moths to a pheromone, candle light, or other substances that emit electromagnetic waves at a narrow-band frequency in the infra red. By some means, the organism must be able to perceive coherent electromagnetic waves and transform them into active work. Callahan found in insects highly specialized antenna sensillas, which function as receiving instruments for narrow-band infrared radiation emitted from scents (pheromones) of female insects.

Popp proposes a similar idea concerning the role of DNA. Assuming sunlight to be coherent when it hits an area of the size of a normal cell—based on the van Cittert/Zernike theorem—it is conceivable that the helical structure of DNA represents the unique geometry to act as a "receiver," or waveguide, for these coherent photons. Furthermore, such a "helix antenna" is able to self-adjust to specific frequencies and thereby might play a mediating role between coherent radiation and the nonlinear development of living matter.

Although Popp presents his conception of the role of DNA in a "quantum logic" form, he is on the right track when he looks for correlations between the evolution of matter and the role of electromagnetic energy. Two key parameters in the geometry of DNA conform to the geometry of power supply for life. First, the distance between two DNA base pairs in the helix structure is about 0.4 nm, which is also the "resolution power" of sunlight; that is, the smallest distance over which a photon can be released. Second, one full rotation of the DNA spiral has the distance of about 3.4 nm, which is an important measure if the spiral geometry is to work as an efficient antenna (Popp, p. 141).

These two parameters may be only suggestive. Yet the search for evidence of how life processes continuously generate new singularities will eventually lead to discoveries of more fundamental principles governing the evolution of life on Earth.

### References


Popp, F.A. 1984. *Biologie des Lichts, Grundlagen der ultraschwachen Zellstrahlung*. Berlin: Verlag Paul Parey.

*Soviet Physics* USPEKHI. 1974. **16:** 568, 571.


## The Gifts of Louis de Broglie to Science

**by Robert J. Moon**

*A review of* Quantum, Space and Time—The Quest Continues,[1] *Part I, 14 essays prepared in honor of de Broglie's 90th birthday anniversary (Aug. 15, 1982) by 18 well-known scientists.*

These studies and essays yield a wealth of insight, not only into the way scientists think, and much of the historical aspect of the development of scientific thought, but more important, into the conception of ideas from the spirit within a scientist. This always takes poetic form, with many facets that yield entrées into a more perfect description of God's creation. Indeed de Broglie described his discovery of wave mechanics in this way: "A great light suddenly appeared in my mind."

Ideas are buried within the individual's spirit and burst forth when the individual's freedom is not suppressed by worldly materialism and dogmatism. Ideas do not come from conscious mentation or reading, since ideas are part of the individual's spiritual makeup and must be searched for from within in order to be discovered. Ideas may flow contrary to the prevailing stream of human thought. The individual will most likely have to navigate upstream and avoid aimless drift, in order to find fertile soil in which to plant an idea for the benefit of mankind.

Such a navigator was de Broglie. Kind and gentle to all, but firm with his concepts, he "attempted to develop the most promising alternative to the orthodox version of quantum mechanics." He started with a model that involved a pilot wave or guiding wave vibrating within a particle, much like a radar on an airplane sees the entire topology ahead, and this in turn guides the plane by means of actions by the pilot. This pilot wave calls for a double solution to the equations of quantum mechanics.

De Broglie was pounced upon by members of the Fifth Solvay Physics Conference in 1927. The Congress did not like his concept of the pilot wave associated with a particle and the consequent double solution.

Wolfgang Pauli made important objections to de Broglie's concept and felt that it did not provide a consistent account of the many-body system or, in particular, a two-body scattering process. De Broglie felt that his idea had at least a germ of an answer. This was not appreciated by those present at the Solvay Conference, and de Broglie's friend Einstein did not speak up for the theory. These two rejections led to rejection by the Congress, which in turn caused de Broglie to close his books on this theory, giving up further work on it.

Einstein had in fact written to H.A. Lorentz on Dec. 16, 1924:

> A younger brother of de Broglie (the one we know) has undertaken a very interesting investigation (Paris Dissertation, 1924) to interpret Bohr-Sommerfeld quantum rules. I believe this is a first weak ray to illuminate this most serious of our physical riddles. I have also found something that speaks for his construction. (p. 41)

De Broglie learned of the letter only after Einstein's death in 1955.

In the introductory paper titled "Louis de Broglie—Physicist and Thinker," Jean-Pierre Vigier opens with a statement very characteristic of de Broglie, "Great physicists fight great battles." These essays, Vigier says, underline "his present position as forerunner, inspirer, and leader of a trend of research which is rooted in his dissent with the overwhelming majority of theoretical physicists—and his solidarity with Einstein in the famous Bohr-Einstein controversy." His scientific observations and interpretations opened new areas particularly on the "meaning and value of scientific knowledge itself."

There are four essential groups of problems with which these essays are concerned and in which de Broglie fought great battles.

(1) The first set is concerned with Heisenberg's dictum that microphenomena exist if and only if they are observable. De Broglie, on the contrary, held to his concept of the pilot wave, Ψ—a real microphenomenon wave that guided particles.

(2) The second set of problems has to do with Bohr's concept that quantum probabilities represent an ultimate limit to human knowledge. Contrary to this, de Broglie conceived of a random set of subquantal hidden variables in a real vacuum with which particles interact and exchange energy; that is, a vacuum alive with subquantal distributions of violent motions, so that particle energy changes when moving from one point to another, in accordance with the principle of least action. These new quantum forces reflect the "wholeness" of the surrounding universe. This concept is that of a new ether model. The vacuum state is the state of "empty space," vibrant with a covariant distribution of covariant spinning oscillators and with random jumps in the velocity of light. This ether is not the old ether-at-rest model, but is a "new description of nature's 'vacuum' that implies a Copernican revolution against the world vision of Newton and Laplace, since it organically combines causal motions with permanent randomness. It interprets quantum mechanics as a Markov process at the velocity of light," Vigier writes.

(3) The third set concerns "the physical origin of the laws of nature themselves." The Copenhagen School, according to Vigier, "regards Quantum Theory as a general form of knowledge that is final in its essence. If this is true, knowledge of nature will never change again but only eventually develop through the introduction of new elementary particles, new Lagrangians, new quantum numbers, and new forms of interaction."

De Broglie and Einstein's approach to theory is basically different, Vigier says. Reality is immense, and no description of the universe by means of a theory and experimental proof will ever be a total and final one. Rather, each new theory proved by experiment is just another thin layer of insight into the nature of the real world.

(4) The fourth set of problems deals with "the existence of causality in nature and covers the present controversy raised by the, now very probable, confirmation of the nonlocal character of quantum mechanical predictions, discovered by John Bell in the Einstein-Podolsky-Rosen type of experiment."

### Bohm Rediscovers the Pilot Wave

John S. Bell's contribution, "On the Impossible Pilot Wave," attempts to present the essential idea "so compactly, so lucidly, that even some of those who know they will dislike it may go on reading. . . ." Referring to the von Neumann impossibility proof, Bell "saw the impossible done" in David Bohm's papers (1952, 1952a) demonstrating "how parameters could indeed be introduced into nonrelativistic wave mechanics, with the help of which the indeterministic description could be transformed into a deterministic one." The pilot wave, ignored by Born and von Neumann, was not impossible. David Bohm had rediscovered the pilot wave!

Bell sets up a simple model of a system whose wave function is $\Psi(a, x, t)$ with one discrete argument, $a = 1, 2 \ldots N$, one continuous argument, $x$, of position, where $-\infty < x < +\infty$ as well as a continuous argument of time, $t$.

He then considers a particle with an "intrinsic spin" free to move in one dimension, and finds a solution of the Schrödinger equation that yields various wave packets Φ that "move apart from one another, and after a sufficiently long time, . . . overlap very little." This model is similar to that of a Stern-Gerlach experiment.

Then, by means of the ideas of de Broglie and Bohm, Bell adds to the wave function, $\Psi$, a particle position, $X(t)$. A particle always has a definite position, and the time evolution of the particle position after many repetitions of the experiment yields a probability distribution of $\rho(X(t),t)\ dX(t)$, which is the conventional quantum distribution for position. Thus the conventional predictions for the result of the Stern-Gerlach experiment obtain. The result is a position observation. Bell writes, "probability enters once only, in connection with initial conditions. . . . Thereafter the joint evolution of $\Psi$ and $X$ is perfectly deterministic." Thus in accordance with Bohr, the results are products of the complete experimental set-up, "system" plus experimental "apparatus" and are not to be regarded as "measurements" of preexisting properties of the "system" alone.

Bell concludes with these precepts so clearly emphasized in the de Broglie-Bohm picture:

(1) "Always test your reasoning against simple models."

(2) The only observations that must be considered in physics are position observations.

(3) In using the word "measurement" it is easy to expect that "'the results of measurement' should obey some simple logic in which the apparatus is not mentioned." "System and apparatus" are inseparable in probing the nature of God's creation. Bell favors banning the word "measurement" in favor of "experiment."

In order to best understand how an idea of de Broglie's had been shelved in 1927, forgotten, and then rediscovered by David Bohm in 1951, Bohm's own testimony of the sequence of events is most apropos. It is reproduced here in full, for it has many facets that should help any physicist to go forward in spite of the many vicissitudes that may intervene.

David Bohm is quoted (pp. 90-91) as follows:

> I wrote a book from Bohr's point of view, mainly in order to understand the quantum theory. But after I had written the book, I felt that I still didn't really understand the quantum theory, and so I began to look for new approaches. Meanwhile, I had sent copies of the book to Bohr, Pauli, Einstein, and other scientists. Bohr did not respond, but Pauli sent an enthusiastic reply, saying he liked the book very much. Einstein also got in touch with me, saying that though the book explained the quantum theory about as well as would ever be possible, he still was not convinced but wanted to discuss the subject with me.
>
> We had several discussions, the net result of which was that I was considerably strengthened in my feeling that there was something fundamental that was missing in quantum theory. This may perhaps have made me work with greater energy, but Y. Ne'eman's statement that I was "shaken" by my conversation with Einstein and "had not recovered to this day" is entirely false. In any case, what actually happened was that I soon came upon the trajectories-interpretation, and prepared a preprint, copies of which were sent to many physicists including de Broglie, Pauli, and Einstein. I learnt shortly thereafter from de Broglie that he had developed this idea much earlier and so, in later versions of the paper, I acknowledged this fact. Pauli was very negative in reply, saying also that de Broglie had developed the same model many years earlier, and that it had been shown by him to be wrong at the Solvay Congress.
>
> As a result of Pauli's letter, I developed a theory of the many-body problem answering his objections, which was incorporated in a second paper [(1952) *Phys. Rev.* **85:** 180]. I had several further discussions with Einstein, but he was not at all enthusiastic about the idea, probably mainly because of the feature of nonlocality of the quantum potential, which conflicted with his basic notion that connections had to be universally in the fundamental laws of physics.
>
> While I can understand Einstein's objections fully, I feel that it may have been a tactical error on his part to dismiss such ideas because they conflicted with his own notions as to the nature of reality. For though perhaps unsatisfactory in many respects, they made possible, as explained in the present paper [by Bohm and B.J. Hiley, pp. 77-92 of the work reviewed here; see below] certain important insights into the meaning of the quantum theory. I feel that a correct approach might have been to encourage such work as a purely provisional approach, but recognizing that it was not likely in itself to be a fundamental theory, without further radically new ideas. The result of not doing this sort of thing was that, for the most part, fundamental physics was reduced to its present state of relying almost exclusively on formulae and recipes constituting algorithms for the prediction of experimental results, with only the vaguest notions of what these algorithms might mean physically.

Bohm and B.J. Hiley ("The de Broglie Pilot Wave Theory and the Further Development of New Insights Arising Out of It") discuss de Broglie's approach in which he assumed a double-solution model to quantum mechanics. That is, (1) a real physical wave which satisfied Schrödinger's equation, (2) a particle following a well-defined trajectory, (3) the momentum, **p**, of this particle was related to the wave through the equation:

$$\mathbf{p} = \hbar \nabla \phi \tag{1}$$

where $\phi$ is the phase of the wave function. The particle is being guided by the wave ("pilot wave"). (4) Inside the particle there is a periodic process (a "clock") which, when at rest has a frequency $\omega_0 = m_0c^2/\hbar$, and the condition for the clock to stay in phase with the pilot wave was derived to be

$$\oint \mathbf{p}\cdot d\mathbf{x} = nh.$$

(5) The locking in phase, he suggested, is a nonlinear interaction, which is crucial in order to obey Schrödinger's equation, and this double solution described the guidance condition. De Broglie's model "provides at least a conceptual connection between quantum mechanics and Einstein's attempt at a unified field theory, in which the particle is also treated as a nonlinear singularity that merges with the background field."

Members of the Fifth Solvay Congress in 1927 objected to this idea, in particular Pauli, and not even Einstein spoke up for the theory. Twenty-five years after de Broglie cast the idea aside, David Bohm rediscovered the "double solution" with its pilot wave and showed it to be a consistent account of a one-body system. In a second paper he extended it to a many-body system in answer to Pauli's objection and this led to new insights as to the meaning of quantum mechanics. Bohm's exchange of ideas with de Broglie led the latter—then 60 years of age—to again take up his old ideas after 25 years, although his approach is not accepted by most physicists.

## The Trajectory Interpretation

Bohm and Hiley develop the trajectory interpretation for a many-body system as an extension of de Broglie's ideas. Their contribution here (pp. 80-87) is so significant that it merits a detailed account. They start with the $N$-body wave function as

$$\Psi(\mathbf{x}_1 \ldots \mathbf{x}_n) = R(\mathbf{x}_1 \ldots \mathbf{x}_n) \exp[iS(\mathbf{s}_1 \ldots \mathbf{s}_n)/\hbar]$$

and define the momentum of the $n$th particle (as did de Broglie) as:

$$\mathbf{p}_n = \nabla_n S \tag{2}$$

Equation (2) is substituted into the many-body Schödinger equation which yields the conservation equation in configuration space:

$$\partial P/\partial t + \sum_n \nabla_n \cdot (P\nabla_n S)/m = 0 \tag{3}$$

(where $P = \Psi^*\Psi$, the probability density in this space), and the modified Hamilton-Jacobi equation

$$\partial S/\partial t + \Sigma(\nabla_n S)^2/2m + V(\mathbf{x}_1 \ldots \mathbf{x}_n) + {}^{n}Q(\mathbf{x}_1 \ldots \mathbf{x}_N) = 0. \tag{4}$$

They conclude from this that "each particle will be acted on, not only by the classical potential, $V$, but also by the *additional quantum potential* $Q$" (emphasis added):

$$Q = (-\hbar^2/2m) \sum_n [(\nabla_n{}^2)R/R]. \tag{5}$$

This interpretation shows that new features of quantum mechanics arise basically from the quantum potential $Q$.

As an illustrative example they consider the case of a two-body system with a product wave function:

$$\Psi(\mathbf{x}_1,\mathbf{x}_2) = \phi_A(\mathbf{x}_1)\phi_B(\mathbf{x}_2) \tag{6}$$

where

$$\phi_A(\mathbf{x}) = R_A(\mathbf{x})e^{iS_A(\mathbf{x})/\hbar} \text{ and } \phi_B(\mathbf{x}) = R_B(\mathbf{x})e^{iS_B(\mathbf{x})/\hbar}$$

Thus:

$$Q = \frac{-\hbar^2}{2m}\frac{\nabla_1^2 R_A(\mathbf{x}_1)}{R_A(\mathbf{x}_1)} - \frac{\hbar^2}{2m}\frac{\nabla_2^2 R_B(\mathbf{x}_2)}{R_B(\mathbf{x}_2)} \tag{7}$$

The quantum potential, $Q$, is the sum of two independent functions. If the classical potential, $V$, is likewise a sum, $V_A(\mathbf{x}_1) + V_B(\mathbf{x}_2)$ then the Hamilton-Jacobi equation reduces to two separate parts:

$$\frac{\partial S_A}{\partial t} + \frac{(\nabla_1 S_A)^2}{2m} + V_A(\mathbf{x}_1) - \frac{\hbar^2}{2}\frac{\nabla_1^2 R_A(\mathbf{x}_1)}{R_A(\mathbf{x}_1)} = 0 \tag{8}$$

$$\frac{\partial S_B}{\partial t} + \frac{(\nabla_2 S_B)^2}{2m} + V_B(\mathbf{x}_2) - \frac{\hbar^2}{2m}\frac{\nabla_2^2 R_B(\mathbf{x}_1)}{R_B(\mathbf{x}_2)} = 0 \tag{9}$$

The conservation equation also apparently splits into two independent parts.

Bohm and Hiley note that "the one-body equation (as treated by de Broglie) arises as an abstraction and a simplification of that of the two-body system, and eventually of the $N$-body system. (It is clear moreover that ultimately these $N$-bodies must be extended to include the whole universe.)"

Note that quantum mechanics and classical mechanics are expressed in terms of the same language.

> [T]he quantum potential, $Q$, is not altered when the wave function is multiplied by a constant, so that it does not fall to zero at long distances, where the wave intensity becomes negligible. However, the classical notion of analyzability of a system into independent parts depends critically on the assumption that whenever the parts are sufficiently far removed from each other, they do not significantly interact. This means that the quantum theory implies a new kind of wholeness, in which the behavior of a particle may depend

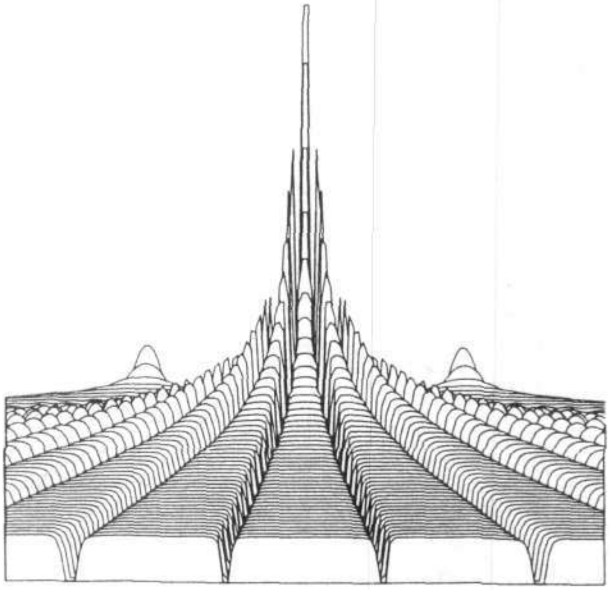


**Figure 1.** Quantum potential for a pair of Gaussian slits. The slits can be seen in the background. The fringes are formed in the foreground, the dark bands coinciding with the valleys of the quantum potential.

significantly on distant features of the over-all environment. This dependence produces consequences similar to those implied by Bohr's notion of unanalyzable wholeness, but different in that the universe can be understood as a unique and in principle well defined reality.

To illustrate in more detail what is meant here, . . . consider an interference experiment, in which a beam of electrons of definite momentum is sent through a two slit system. In Figure 1, we show the results of a computation of the quantum potential [C. Philippidis, C. Dewdney, and B.J. Hiley (1979) *Nuovo Cimento* **52B:** 15]; and in Figure 2, we show the trajectories resulting from the potential.

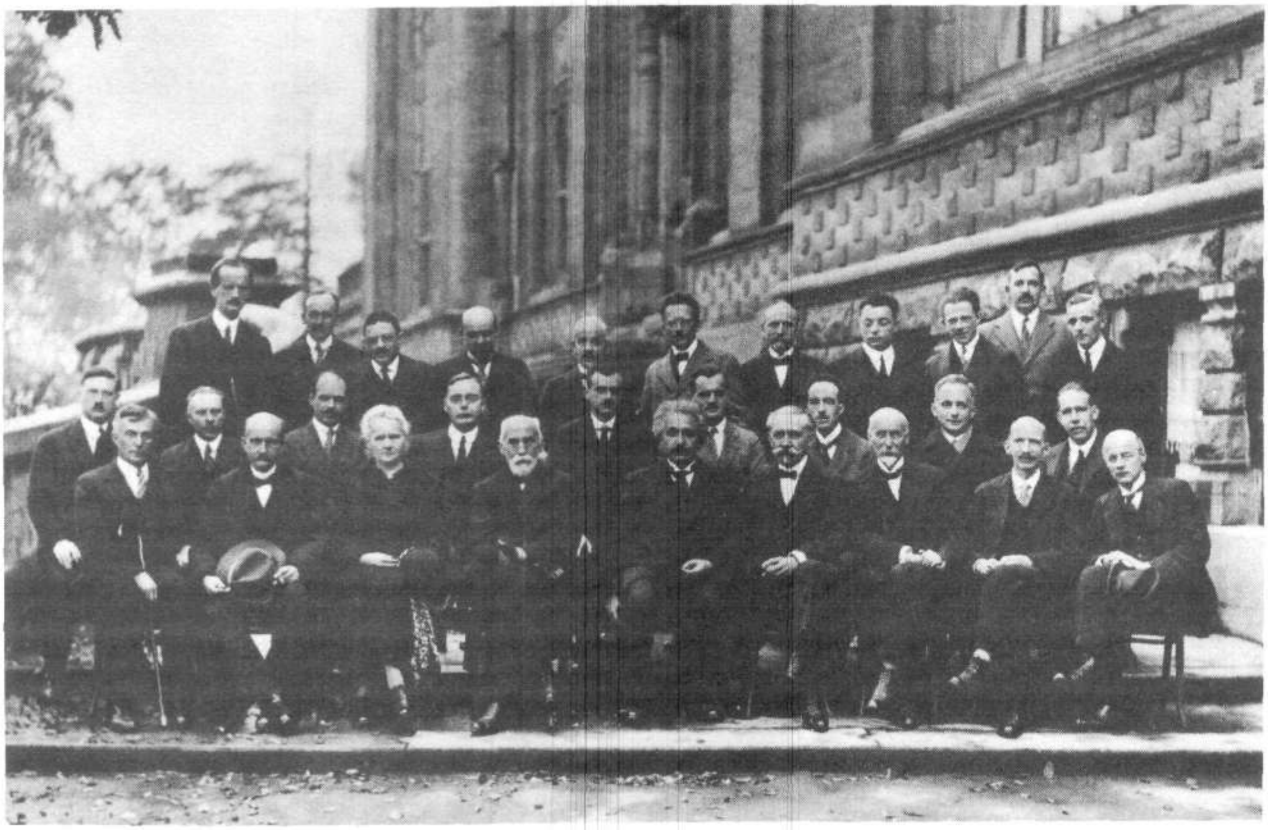
What is especially significant in Figure 1 is that the quantum potential remains large at long distances from the slits, taking the form of a set of valleys and high ridges, which latter gradually flatten out into broad plateaux. In Figure 2, one sees how the trajectories are ultimately bunched into these plateaux by the overall effect of the potential, and that this brings about the interference pattern. (So that, for example, if one of the slits had been closed, the quantum potential would have been a smooth parabolic function, which would produce no pattern of fringes). The fact that the quantum potential does not in general fall off with the distance is thus what explains interference and diffraction patterns, and this is clearly also what implies the kind of wholeness

**Figure 2.** The particle trajectories emanating from the Gaussian slits at the bottom of the figure. The fringes at the top result from the bunching of the trajectories.

Source: D.J. Bohm and B.J. Hiley, "The de Broglie Pilot Wave Theory and the Further Development of New Insights Arising Out of It" in A.O. Barut et al., eds., *Quantum, Space and Time—The Quest Continues* (Cambridge: Cambridge University Press, 1984), pp. 82-3.

of particle and environment to which we have referred above.

One may return here to the analogy of the airplane guided by radar waves. Evidently, it is not a case of mechanical pressure of these waves on the airplane, but rather, the information concerning the whole environment is enfolded by the waves, and carried into each region of space. The airplane thus responds actively to the *form* of the waves, and this *form* is not altered as the intensity falls off with distance. A similar response to the *form* of the quantum potential is seen to be characteristic of the behavior of the electron. This means that in the microworld the concept of active information is relevant (see Bohm



**The Fifth Solvay Physics Conference, held in Brussels, Oct. 23-29, 1927, sponsored by the Solvay International Institute of Physics. Among the 30 scientists who attended the conference were E. Schrödinger, W. Pauli, W. Heisenberg, W.L. Bragg, P.A.M. Dirac, A.H. Compton, L. de Broglie (middle row, third from right), M. Born, N. Bohr (middle row, far right), I. Langmuir, M. Planck, M. Curie, H.A. Lorentz, and A. Einstein.**

and Hiley [(1975) *Found. Phys.* **5:** 93] for more detail).

What has been said thus far about the new kind of wholeness implied by the quantum theory for the one-body system is further strengthened by a consideration of the many-body system. For here one finds that when the wave function is no longer separable as a product of functions of the coordinates of each particle, the quantum potential leads to a strong interaction between all particles of the system, that does not in general fall off to zero when the particles are distant from each other. This is evidently an extension of the dependence of the particle on its overall environment that characterizes the one-body system. But in addition, there is a yet more thoroughgoing breakdown of the possibility of analysis, because the force acting on each particle is no longer expressible as a predetermined function of the position of the other particles. Rather, the functional form of the force depends on the whole set of conditions in which the wave function is defined and determined (so that, for example, the form changes whenever this quantum state of the whole changes).

Let us take, as an example, the hypothetical experiment of Einstein, Podolsky, and Rosen [A. Einstein, B. Podolsky, and N. Rosen (1935) *Phys. Rev.* **47:** 777]. We consider there the original form of the experiment, in which we start with a quantum state of a two-particle system in which $(x_1 - x_2)$ and $(p_1 + p_2)$ are both determined. This is given by

$$\begin{aligned}\Psi(x_1x_2) &= f(x_1 - x_2 - a) \\ &= \sum_k C_k \exp[ik(x_1 - x_2 - a)] \qquad (10)\end{aligned}$$

where $f(x_1 - x_2 - a)$ is a packet function sharply peaked at $x_1 - x_2 = a$, while $C_k$ is its Fourier coefficient. Evidently, in this state $p_1 + p_2 = 0$ while $x_1 - x_2$ can be made as well defined as we please.

In this experiment, one can measure $x_1$ and immediately know that $x_2 = x_1 + a$ (to an arbitrarily high degree of accuracy). Alternatively, we can measure $p_1$ and immediately know that $p_2 = -p_1$. In both cases, the first particle is disturbed in the process of measurement and, of course, the disturbances can account for the Heisenberg uncertainty relations as applied to the particle

$\Delta p_1 \Delta x_1 \geqslant h$. But since the second particle is assumed not to interact with the first in any way at all, it follows that we are able to find its properties without its having undergone any disturbance whatsoever. Nevertheless, according to the quantum theory, the uncertainty principle, $\Delta p_2 \Delta x_2 \geqslant h$, must still apply. So Heisenberg's explanation of this uncertainty as due to a disturbance resulting from measurement can no longer be used. It was this which indeed led Einstein, Podolsky, and Rosen [1935] to argue that since both $x_2$ and $p_2$ were in principle measurable to arbitrary accuracy without a disturbance, they must have already existed independently in particle 2 as "elements" of reality with well-defined values before the measurement took place. And so, they concluded that quantum mechanics is an abstraction giving only an incomplete and fragmentary description of the underlying reality (as insurance statistics are abstractions that similarly yield an incomplete and fragmentary description of the people to whom they are applied).

As is well known, Bohr [N. Bohr (1935) *Phys. Rev.* **48:** 696] answered this argument by means of a further development of his notion that the measurement process is an unanalyzable whole, which led in this case to the conclusion that there is no meaning to the attempt to give a detailed description of how correlations of position and momentum are carried along by the movements of the parts of a many-body system. It is interesting, however, to go carefully into how the trajectory interpretation differs from that of Bohr, and yet comes to a similar notion of unanalyzable wholeness, though, of course, in another way. For this case, writing $f = Re^{iS/\hbar}$, we obtain for the quantum potential

$$Q = \frac{-\hbar^2}{2m}\left(\frac{\partial^2 R}{\partial x_1^2} + \frac{\partial^2 R}{\partial x_2^2}\right) \Big/ R$$

$$= \frac{-\hbar^2}{m}\,\frac{\partial^2 R}{\partial \Delta x^2}(\Delta x - a) \Big/ R(\Delta x - a) \qquad (11)$$

with $\Delta x = x_1 - x_2$. This function evidently remains large, even when the distance, $a$, separating the particles is not small. Therefore, when the properties of the first particle are measured, the quantum potential brings about a corresponding disturbance of the second particle. And from this, it can be shown [D. Bohm (1952) *Phys. Rev.* **85:** 180] that in a statistical ensemble of similar measurements, Heisenberg's uncertainty solutions, $\Delta p_2 \Delta x_2 \geqslant h$ will still be obtained.

## Karl Popper on Bohr and de Broglie

"The new gospel of irrationality," Karl Popper writes, "was first publicly preached by Bohr in Como at the International Congress of Physics 1927; and a few weeks later, in Brussels, at the [Fifth] Solvay Congress." Popper's contribution is "A Critical Note on the Greatest Days of Quantum Theory." He reports young physicists thinking Einstein had become prematurely old at the age of 48! Bohr became the favorite of the young brilliant physicists led by Heisenberg, Pauli, and Max Born into what the young considered a greater revolution than Relativity. Some thought Einstein an antediluvian. Popper thinks "the real break was . . . between a radical and dogmatic empiricism . . . and a critical realism." This empiricism was hidden under the "general usage of the almost incredible term 'observable'. . . . *There are, in fact, no observables in atomic physics.*" There are only indirect observations, that is, traces of the effects of particles on the environment through which the particles pass.

The de Broglie waves made Bohr's atom understandable. The advent of recording Geiger counters and photographic Wilson cloud chambers began the death of the "observer."

A new term, "hidden variable," arose to offset "observable," Popper writes. "In fact . . . *all* physical 'variables' are hidden." Hidden variables are a consequence of Heisenberg's interpretation of his indeterminacy formulae.

The Copenhagen school interprets Heisenberg's indeterminacy principle as *excluding:*

(a) all measurements which would be better than the product of the change of momentum with the change of position, $\Delta p_x \Delta x \geqslant h$;

(b) as well as all subjective knowledge better than this; and

(c) the existence of all particles that possess position and momentum to a greater precision than (a).

On the other hand, "a realist interpretation of quantum mechanics would interpret" (a) above "neither speaking about measurements nor about our knowledge," but rather "as speaking about the preparation of particles, and their position and momenta," independent of whether they are being observed or measured, though the realists recognize that the particles of course will respond to fluctuation in the environment mostly in a partially unpredictable fashion.

Einstein, Podolsky, and Rosen published their famous paper, "Can Quantum Mechanical Description of Physical Reality Be Considered Complete?" in 1935 "to show that a particle possesses both a precise position and a precise momentum." Popper considers the argument valid.

## De Broglie on the Poetry of Creativity

Georges Lochak of the Fondation Louis de Broglie ("The Evolution of the Ideas of Louis de Broglie on the Interpretation of Wave Mechanics") writes that de Broglie "always experiences creation as a dazzling poetic vision, and he cannot help feeling sad when he sees it weaken and fade as it is translated by himself or by others into a necessarily mathematical language."

O. Costa de Beauregard ("Reminiscences on My Early Association with Louis de Broglie") tells this related story of de Broglie's appreciation of Paul Valéry: "One spring afternoon, in those days bygone, I went to his [de Broglie's] home in the Paris suburb of Neuilly, with a work on physics I wanted to discuss with him. The weather was beautiful, and the chestnut trees in blossom. I don't remember how it happened that Louis de Broglie came to ask me which was, in my opinion, France's greatest poet. Somewhat hesitatingly, I answered that this was a question of personal taste, and that he might not agree with my choice of Paul Valéry. Well, this choice was also his. His following question was, among Valéry's masterpieces, which one would I select? Again with hesitation, I said that my selection was not the (very rightly) celebrated *Cimetière Marin* but rather the long, superb, philosophical poem with the understating title *Ebauche d'un Serpent* (Sketch on the Theme of a Snake). It is a sparkling theological address of Lucifer to God, starring the Garden, the Snake, Eve, the Tree—and what followed therefrom. Well, again de Broglie agreed. And we spent the rest of the evening reading and commenting on the wonderful poem, which finally has to do with the irresistible growth of knowledge from roots in the darkness beneath, to leaves in the brilliance above. . . . So it seems to me that there is some Leibnizian preharmony between Valéry and scientists."

As John Bell proclaims, "Long may Louis de Broglie continue to inspire those who suspect that what is proved by impossibility proofs is lack of imagination."

### Notes

1. *Quantum, Space and Time—The Quest Continues*, Asim O. Barut et al., eds. (Cambridge: Cambridge University Press, 1984, 680 pp., $49.50, paperbound). The 14 essays in Part I, covering 245 pages, are by the following authors: Jean-Pierre Vigier, Georges Lochak, Alwyn van der Merwe, O. Costa de Beauregard, Karl Popper, J. Andrade e Silva, J.S. Bell, D.J. Bohm and B.J. Hiley, L. de la Peña and A.M. Cetto, Stanley P. Gudder, Ph. Guéret and J.-P. Vigier, Mioara Mugur-Schächter, F. Selleri, and H.-H. v. Borzeszkowski and H.-J. Treder.

Part II is a collection of essays dedicated to Eugene Paul Wigner on the occasion of his 80th birthday, Nov. 17, 1982. Part III, in like manner, is dedicated to Paul Adrien Maurice Dirac on the occasion of his 80th birthday, Aug. 8, 1982.

### References

Bohm, D. (1952, 1952a) *Phys. Rev.* **85:** 166 and 180.

# Book Review

**Reversals of the Earth's Magnetic Field**
**By J.A. Jacobs**
**Bristol, UK: Adam Hilger Ltd.**
**1984**
**$35.00 US**

The author reviews the current state of reversal theory, that is, the theory of how the magnetic field of the Earth reverses and on what basis we know that the field does in fact reverse.

The basic problem the author faces, and to which he readily admits in the introduction, is that there exists no satisfactory theory of the Earth's magnetic field—a theory that explains not only how the field reverses but how the field came into being and continues to exist. Such a theory is, in fact, one of the greatest outstanding problems of the work of Riemann and Gauss.

The difficulty involved is mostly that nothing is known about the composition of the Earth beyond a depth of 20 km or so. We "guess" that it must be composed of some material with ferric properties. We also know that as we go deeper into the Earth, it grows warmer and, as is easily calculated, the pressure gets greater. This indicates that the material in the core of the Earth is hot and fluid. This idea agrees well with the material seen coming out of volcanos. Thus it seems that the inner parts of the Earth are made of fluid stuff with magnetic properties.

Fluid stuff with magnetic properties is also known as plasma, and what author J.A. Jacobs does not seem to understand is that he is dealing with a plasma under immense pressure. Therefore, one must turn away from mechanical models of the Earth's field and look at plasmas, the best sort of plasmas being those that are self-confining or confined under a high external field. The plasmas that fill this bill are z-pinches and Mather-type focuses. In the case of z-pinch, the best example is the Los Alamos ZT-40, which has intense field reversals at a semiregular rate. With the laboratory plasma focus, Winston Bostick et al. have demonstrated plasma conditions of the sort that occur on the astrophysical scale. The question as to how the field reversals occur should become a matter of what is the proper plasma model.

If the generator of the field is taken to be a plasma, one would expect reversal of the magnetic fields, shifts in structure, and even very short-term changes in strength and direction. It seems clear, although this is not the perspective of the author, that the core of the Earth is a very complex magnetohydrodynamic system.

—*Robert B. McLaughlin, Jr.*

# Information for Contributors

Manuscripts should be sent to David Cherry at the Fusion Energy Foundation, P.O. Box 17149, Washington, D.C. 20041-0149.

Manuscripts should be submitted in triplicate (with three sets of illustrations, of which one is an original.) They should be typewritten on one side of letter (quarto) paper and double spaced with at least 25mm (1 inch) margins. All pages must be numbered in sequence beginning with the title page.

TITLE PAGE of the manuscript should contain the complete article title; names and affiliations of all authors; name, address, telephone number, and cable or Telex number for all correspondence.

ABSTRACT of no more than 200 words should summarize the work and major conclusions.

TEXT should define all abbreviations at first mention. American measuring units should be accompanied by metric translation. In general, the meter/kilogram/second/ampere system of units should be used. Letters of permission should be submitted with any material that has been previously published.

FOOTNOTES should appear at the bottom of the respective page and must be numbered consecutively.

REFERENCES should be referred to in the text by the author's name and the date of the work. The references should also be listed alphabetically on a separate page. The references should include for articles: author, year of publication, title, periodical name, volume no., and pages; for books: author, year of publication, title, place of publication, and publisher's name.

TABLES, FIGURES, and ILLUSTRATIONS must be prepared each on a separate page, numbered in order of appearance, and have a title and descriptive legend. Photographs should be glossy black and white prints.

ACKNOWLEDGMENTS. Illustrations from other publications must be acknowledged. Include the following when applicable: author(s), title of journal or book, publisher and place of publication (if book), volume number, page(s), month and year of publication. The author is responsible for obtaining the publisher's permission to reprint.

REPRINT order forms will be sent to the author with proofs of the article.