The Theory of Schrödinger Operators: What's It All About?

by Barry Simon

CALTECH HAS BECOME a world center of the study of Schrödinger operators. What is the theory of Schrödinger operators anyway? In short, it is the rigorous mathematical study of the Hamiltonian operators of nonrelativistic quantum mechanics.

Schrödinger operators are a part of mathematical physics, an area that suffers the usual fate of interdisciplinary areas: Too often the mathematicians think it's physics, and the physicists think it's mathematics. (Chemical physics has the same problem. Many years ago so did biochemistry, but in a sense *it succeeded in absorbing a large part of* biology.) Perhaps the reader will appreciate that Schrödinger operators have useful insights to offer to both mathematics and physics.

In quantum theory, a fundamental role is played by the energy written as a function of the momentum and position of the constituent particles, but with the twist characteristic of quantum mechanics (due to the Heisenberg relations $[p,x]=-i\hbar$ that the momentum, p_j, is replaced by the elementary differential operator $-i\hbar \; \partial/\partial x_j$ (where $\hbar$ is the rationalized Planck's constant, $h/2\pi$; having been careful to put $\hbar$ here, for typographic simplicity I henceforth will use units with $\hbar = 1$ and with electron mass equal to ½). Thus, for a classical system of N electrons with momenta p_j ($j=1,...,m=3N$ for the N 3-vector momenta) and classical energy

$$\sum_{j=1}^{m} p_j^2 + V(x),$$

the Hamiltonian operator becomes $H=-\Delta+V$ where the Laplacian Δ equals

$$\sum_{j=1}^{m} \partial^{2}/\partial x_j^2.$$

The Hamiltonian plays a crucial role in the dynamics of quantum systems since wave functions satisfy the partial differential equation $i\partial\psi/\partial t = H\psi$ called the time-dependent *Schrödinger equation*.

In physics textbooks, H is called the Hamiltonian since it is the quantum analog of the object introduced in classical mechanics by William Hamilton. The same name is also used for the quantum energy operator of relativistic quantum field theory. To emphasize that one is looking only at the nonrelativistic case, the Hamiltonians of nonrelativistic quantum mechanics are called Schrödinger operators. This honors Erwin Schrödinger, one of the founding fathers of quantum mechanics. Thus the theory of Schrödinger operators is simply the study of the differential operators $-\Delta+V$.

At first sight, one would not think that such an innocent-looking object could have a very interesting structure. Of course, one might conclude the same thing about the clas-

sical analog, $m\ddot{x} = -\nabla V$, which is rich enough to describe phenomena as varied as the dynamics of binary stars or of water in a waterfall. In the same way, Schrödinger operators describe the full richness of quantum dynamics. Although there are results involving general V (perhaps there has been too much work on the general and not enough on the specific classes), much of the recent thrust has focused on the Vs relevant for atomic and molecular physics, the description of perfect solids, and more exotic objects such as amorphous materials and quasicrystals. Here I will describe several results, all related to atomic physics, which I hope will give a flavor of the scope and significance of the field.

The first question that a mathematician might ask concerns the existence of solutions of the time-dependent Schrödinger equation $i\dot{\psi}_t = H_t$. For some reasonable class of Vs, including that describing N electrons moving in the Coulomb field of M protons (taken infinitely massive) in fixed positions $\vec{R}_1, \ldots, \vec{R}_M$ (equation 1)

$$V(\vec{x}_1, \ldots, \vec{x}_N) = -\sum_{\substack{1 \le i \le N \\ 1 \le j \le M}} \left|\vec{x}_i - \vec{R}_j\right|^{-1} + \sum_{1 \le i < j \le N} \left|\vec{x}_i - \vec{x}_j\right|^{-1} + \sum_{1 \le i < j \le M} \left|\vec{R}_i - \vec{R}_j\right|$$

(now the xs are three-vectors), one would like to know that the time-dependent Schrödinger equation has a unique solution, if not for all initial conditions, at least for a very large set of initial conditions.

Formally, the solution is given by

$$\psi_t = exp(-itH)\psi_0,$$

but what does $exp(-itH)$ mean? Expanding the exponential and considering terms like $\Delta^2 V\psi_0$, where the derivatives of V produce non-square integrable singularities, will show the problem with a naive approach. On a slightly more sophisticated level, students of quantum mechanics learn that they cannot study the quantum motion of a particle in a box without specifying boundary conditions. How does one know that boundary conditions are not needed at infinity or at the Coulomb singularities in solving the Schrödinger equation with potential (1)?

There is one attitude about these questions that needs to be addressed — that is, that you don't need existence proofs for a physical theory because nature is an existence proof. This idea of nature as a grand analog computer misses the whole point of existence proofs: We are not testing nature but rather our theories, which could, after all, be wrong or incomplete. A paradigm of this phenomenon concerns the changing attitude about quantum electrodynamics (QED). After the formulations of the theory by Feynman and Schwinger in the 1940s, a common attitude was that the impressive agreement with experiment proved that the theory had to be consistent. The current opinion among elementary particle theorists is quite different: As an abelian gauge theory, QED is most likely mathematically inconsistent. Rather, high energy physicists hope that the non-abelian gauge theory associated to the unified treatment of electromagnetic and weak interactions is consistent. The agreement with experiment is because the formal perturbation expansion of the putatively inconsistent theory is quite close to that of the putatively consistent theory.

In response to the development of the new quantum mechanics (1925-28), John von Neumann developed a theory of unbounded operators in Hilbert space precisely to deal

with foundational questions in quantum mechanics. What von Neumann realized is that the key to solving the time-dependent Schrödinger equation was the proof that H has a abstruse mathematical property called essential selfadjointness.

But here progress on this particular problem stalled for 25 years, in part because von Neumann was convinced that the problem of proving essential selfadjointness of $-\Delta + V$ for V of the form (1) was an impossibly hard problem. And he was not shy about informing others of his opinion. Indeed, I am told that as late as 1950, he insisted that the problem was not trivial even for the hydrogen atom. (Of course, physicists do know how to write down all the eigenfunctions, including continuum eigenfunctions, in this case; in fact, by using the rotational symmetry of the hydrogen atom, one can reduce this case to some ordinary differential equations where selfadjointness was completely analyzed by Hermann Weyl in 1912. Using his methods it is not hard to prove essential selfadjointness for the hydrogen atom Hamiltonian.) This is ironic, because the two main inputs that turn out to be essential for the proof of this fundamental property are an abstract result in operator perturbation theory proven by Franz Rellich in Germany in the mid-1930s and certain inequalities now known as Sobolev inequalities, developed in the 1930s by Sergei Sobolev in the Soviet Union and Salomon Bochner in the United States, among others. These tools were available by 1936 or so. But nobody put them together, perhaps because of von Neumann's well-known attitude that the problem was too hard to work on.

The essential selfadjointness of atomic Hamiltonians was established by Tosio Kato (since the late 1950s professor of mathematics at UC Berkeley). Here is his description of the history of his work, taken from his acceptance speech for the Wiener Prize (which appeared in the November 1980 issue of the *Notices of the American Mathematical Society*):

"During World War II, I was working, in the countryside of Japan, on the spectral theory of Schrödinger operators and perturbation theory. As a physics student I had been led to study these problems on my own, since no one seemed to pay attention to them in spite of the existence of the general principle given by von Neumann. My first efforts were directed toward establishing the essential selfadjointness of Schrödinger operators and laying a mathematical foundation for perturbation theory. (At that time I did not know of Rellich's work.)

"These works were more or less completed by the end of the war, but I was not very lucky with their publication. A couple of years later I submitted two papers on the subject to *Physical Review*. They were soon forwarded to the *Transactions of the American Mathematical Society*, where the manuscripts were passed from one referee to another without success, eventually to be lost. I had to resubmit new manuscripts. After three years from initial submission, the papers were finally saved by the last referee."

Interestingly, the analog of Kato's theorem for classical mechanics is open. That is, one can ask about global (in time) solutions of Newton's equations for point masses interacting gravitationally (or electrostatically). Even in the two-body case, global solutions will not exist for *all* initial conditions, since collisions are possible in finite time. But in the two-body case, only initial conditions with zero angular momentum can have collisions, so for almost all initial conditions, the classical two-body problem has global solutions. The analog of this was proven for the three-body problem by George Birkhoff (using results of Painleve and Sundman) in 1927, and for the four-body problem by Donald Saari (Northwestern) in 1977. It is open for $N \geqslant 5$ where there are indications of a new phenomenon: initial conditions where particles travel infinitely far in finite time (the infinitely large velocities are possible because of pairs spiraling into each other). It is not clear whether these only occur for a set of initial conditions of measure zero. The quantum result can be viewed as an indication that for general N, the classical problem does have solutions for most initial conditions. The reason that quantum mechanics is nicer than classical mechanics is the uncertainty principle, in this case expressed through the Sobolev inequalities.

Kato's work can be viewed as the birth of the modern theory of Schrödinger operators. Once we know this fundamental result, we can begin to ask many detailed questions. Some of the most subtle involve an area I call "quantum potential theory," which is the rigorous study of exact Coulomb binding energies in quantum mechanics.

A basic result in quantum potential theory is "the stability of matter." This involves a problem first raised by Lars Onsager (known

for his work on the Ising model, and on non-equilibrium thermodynamics, for which he received the Nobel Prize in chemistry) in the 1930s. It is a basic fact of astrophysics that bulk matter in the absence of nuclear effects undergoes gravitational collapse. Onsager asked how we know that bulk matter doesn't undergo "electrostatic collapse." We know that quantum mechanics yields stability of a system of one electron and one proton. In classical mechanics the electron would fall into the proton, but in quantum theory this doesn't happen. This implies that a system of 10^{26} protons and 10^{26} electrons won't collapse to zero size, but it certainly isn't obvious *a priori* that such an array won't collapse to a very small size indeed. Of course, since individual electrostatic forces are much stronger than gravitational, if there were electrostatic collapse, it would require much less matter than gravitational collapse, so we would observe it. Since we don't, it must not occur, but this doesn't explain *why* it doesn't occur and whether the fact that it doesn't occur is just due to quantum mechanics and electrostatics.

The realization came quite early that lack of collapse is implied by the binding energy of a large system of particles being an extensive quantity, that is, that a system with potential (1) has a total energy bounded from below by $-c(N+M)$ for some constant, c. Freeman Dyson (Institute for Advanced Study) and Andrew Lenard (Indiana University) first proved that this is the case in 1967. There is one especially striking aspect of their work: It is critical for their proof that electrons are fermions, that is, that they obey the Pauli exclusion principle. And we know now that this is essential; if both electrons and protons were bosons, electrostatic collapse would take place. While the precise rate of collapse is not known, it is likely that in the neutral *(N=M)* case, volume shrinks as the inverse fifth root of N. A system of 10^{26} "Bose" electrons and protons would live in a volume a very small fraction of a single hydrogen atom.

The relevance of the Pauli principle, even for qualitative features, is ubiquitous in quantum potential theory. For example, last year Elliott Lieb (Princeton), Israel M. Sigal (UC Irvine), Walter Thirring (University of Vienna), and I showed that the number of electrons, $N(Z)$, that one can bind to a charge Z nucleus grows as Z in the sense that $N(Z)/Z$ goes to 1. But if electrons were bosons, Lieb and Raphael Benguria (University of Chile) showed that the analog $N_b(Z)$ grows at least as fast as $1.21Z$. The observed fact of nature that there seem to be no negative ions with net charge larger than 1 is critically dependent on the Pauli principle.

As realized by Lieb and Joel Lebowitz (Rutgers), an important consequence of the Dyson-Lenard theorem combined with a study of shielding is the existence of thermodynamics for bulk matter, that is, of the extensive nature of basic quantities such as pressure in the quantum statistical mechanics of Coulomb systems.

The Dyson-Lenard result is not the end of the story, because the constant c in their bound on the energy is roughly 10^{14} Rydbergs (a Rydberg is the binding energy of hydrogen). Thus, while matter cannot shrink indefinitely, it could shrink so that interparticle distances were roughly 10^{-14} Bohr radii and still not violate the theorem of Dyson-Lenard. The large number 10^{14} in their proof occurs in part because of complexity. Sacrificing some truth to humor, one could say that their proof has 14 steps, each of which introduces a factor of 10 error. Dramatic progress was made by Lieb and Thirring in 1975, who (counting some later refinements of their ideas) obtain a constant of roughly 20 Rydbergs.

The physics behind the Lieb-Thirring proof is quite illuminating; it is perhaps worth describing a part of it. There is an old, quasiclassical approximation to quantum mechanics called the Thomas-Fermi (TF) approximation. In 1973 Lieb and I rigorously proved that this approximation is exact in the large Z limit in the sense that it properly describes the total binding energy and the electron density of the core. It does not correctly describe electrons near the outer shell, nor the ionization energies relevant to chemistry. Quantum chemists in the 1950s tried to compute numerically molecular binding energies using the TF approximation and were unable to find any binding. In response to this, Edward Teller proved in 1960 that molecules never bind in TF theory. (Lieb and I later supplied some points of mathematical rigor, especially the existence of solutions of the TF equation, but Teller's argument is quite close to the rigorous one.) Thus, stability of matter is easy in TF theory: By Teller's result, the energy of an array of protons and electrons is always bounded from below by the total binding energy of well-isolated protons and an appropriate number

of electrons. Since it can be shown that in TF theory a proton binds only one negative charge, in TF theory the binding energy of N protons and M electrons is bounded by cN, where c is the binding energy of hydrogen in the TF approximation.

Lieb and Thirring prove this by using their extension of the Sobolev inequalities (the Pauli principle enters in this step) and an additional trick that the total quantum energy of a system of N protons and M electrons is bounded below by $-dM+TF'$, where d is a constant related to c and TF' is the TF energy, but in a theory with the wrong value of h. Thus the bound

$$TF' > -c'N$$

implies stability of matter.

In a real sense, their proof has the right physics behind it. Stability is a statement of lack of collapse. This collapse is prevented by the interaction of "atomic cores," and Teller's theorem is precisely an assertion that cores (which, in the large Z limit, are described by TF theory) repel. Of course, $Z = 1$ is not large Z, so that it isn't clear that TF theory will apply here; Lieb and Thirring's discovery is that, if one is willing to make a small sacrifice in constants, it does apply.

The last example I will discuss is slightly more technical in detail. Among other things, it illustrates that in mining for pyrite, you can occasionally discover gold. Simple-minded, two-body scattering theory breaks down precisely at Coulomb potentials. There is a logarithmic infinity that must be handled. From a time-dependent point of view, this was done in 1964 by John Dollard (University of Texas). But, typically, mathematicians are not satisfied with treating only the physically relevant case, but want to understand where the modified scattering theory breaks down, so literature has developed on scattering for potentials decaying more slowly than Coulomb. I must confess that, while I recognize the validity of this area, I have not found it especially attractive or interesting. When I was starting out in the early 1970s, two other bright, young mathematical physicists, Rick Lavine (University of Rochester) and Jean Michel Combes (University of Toulon), separately proposed studying long-range scattering using a mathematical discipline called C*-algebras. Not only was the problem of limited interest, but I was convinced that this was the wrong approach to the problem. So, being a brash young man at the time, I

didn't hesitate to tell both Lavine and Combes that they were wasting their time. In a sense my opinion was correct: The C*-algebra approach to the problem has not gotten very far, and there are now much better ways of analyzing the problem. But fortunately Lavine and Combes didn't listen to me, because each of them ran into technical problems that forced them to develop striking new methods. Lavine's ideas were a major element in a key breakthrough by Eric Mourre (CNRS Marseilles) in 1979. Interestingly enough, while the methods of Lavine and Combes seem to be unrelated, recent work by Peter Perry, the Bantrell Research Fellow in Mathematical Physics here at Caltech (with Arne Jensen at the University of Kentucky and Mourre) has shown an intimate relation between Mourre's descendent of Lavine's ideas and the Combes idea.

Combes's ideas originally appeared as an appendix to a paper on the C*-approach to long-range scattering. After some reflection, Combes decided to perform an appendectomy and throw the patient away — his paper never appeared, but he pursued the appendix. In 1971 Combes published two papers developing this approach: one with Jean Aguilar (CNRS Marseilles) on the two-body case, and one with Eric Balslev (University of Arhus) on the more subtle multiple-particle case. The main result of this analysis was a proof of the non-occurrence of a mathematical pathology called singular continuous spectrum in atomic and some other multiparticle quantum systems. Shortly thereafter, in several papers, I exploited the method to study resonances and, in particular, the mechanism whereby an embedded bound state turns into a resonance. Then quantum chemists and calculational atomic physicists, led by John Nutall (University of Western Ontario), got interested in the method as a practical way of computing resonances from first principles. Some of the significant work on molecular resonance was done by Bill McCurdy, now at Ohio State, and Tom Resignio, now at Lawrence Berkeley Laboratory, who learned the method while working (as graduate student and postdoc respectively) in Professor of Theoretical Chemistry Vince McKoy's group here at Caltech.

As we shall see, the group of dilations continued analytically plays a critical role in the Aguilar-Balslev-Combes theory. Combes, being French, dubbed the method "dilatation

analyticity," later shortened to "dilation analyticity." The atomic physicists and quantum chemists, not liking fancypants language, called the method "complex scaling," the most common name now.

We begin by describing the idea for the case of the hydrogen Hamiltonian,

$$H = -\Delta - |r|^{-1},$$

whose spectrum has a continuous part

$$[0,\infty),$$

and a discrete part, the eigenvalues at

$$-\frac{1}{4}n^{-2} \ (n=1,2...).$$

Under the scale transformation

$$r \rightarrow e^\theta r,$$

H goes into

$$H(\theta) = -e^{-2\theta}\Delta - e^{-\theta}|r|^{-1}.$$

This operator has a natural analytic continuation to complex

$$\theta = \varphi + i\eta \ (\varphi,\eta \ real).$$

What is the spectrum of $H(\theta)$? The continuous spectrum should come from states near infinity where $|r|^{-1}$ doesn't count; that is, the continuous spectrum should be that of

$$-e^{-2(\varphi+i\eta)}\Delta$$

(this can be made precise by a theorem of Weyl). Since this operator is just a multiple of $-\Delta$, its spectrum is just
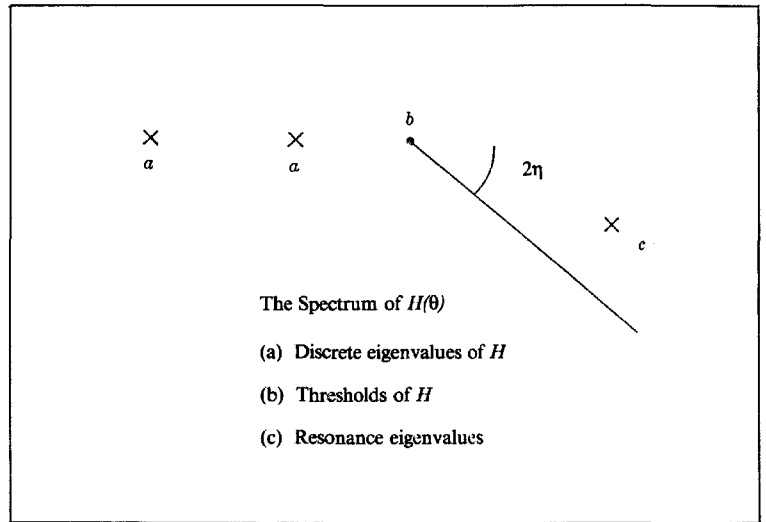
$$\{e^{-2(\theta+i\eta)}\alpha | \alpha \ in \ [0,\infty)\} = \{e^{-2i\eta}\alpha | \alpha \ in \ [0,\infty)\}.$$

Thus, as η varies, the continuous spectrum moves away from the real axis. However, a separate argument shows that the discrete eigenvalues

$$-\frac{1}{4}n^{-2}$$

don't move. As the continuous spectrum swings down, new eigenvalues can appear, but only out of the continuous spectrum. Put differently, if $\eta > 0$ is decreased to 0, any eigenvalue of $H(\theta)$ persists, except that if the continuous spectrum hits them, they can be gobbled up and disappear (the Pac Man theory of resonances??). Clearly, one should associate these new eigenvalues with resonances. Such eigenvalues do not occur for the hydrogen Hamiltonian, but for multiparticle Coulomb Hamiltonians there is a similar theory: $H = T + V$, the kinetic and Coulomb energy

$$H(\theta) = e^{2\theta}T + e^{-\theta}V.$$

The Spectrum of $H(\theta)$

(a) Discrete eigenvalues of H

(b) Thresholds of H

(c) Resonance eigenvalues

Now the continuous spectrum is in various rays,

$$\{t + e^{-2i\eta}\alpha | \alpha \ in \ [0,\infty)\}$$

where t is a possible scattering threshold of the system. Eigenvalues of $H(\theta)$ do occur and can be computed accurately by variational methods. Moreover, this precise mathematical definition of resonance allows the rigorous study of a wide variety of objects. For example, I was able to prove the convergence of time-dependent perturbation for Auger states, and Evans Harrell (Georgia Tech) and I, using the extension of the theory to Stark Hamiltonians (due to Ira Herbst of the University of Virginia), were able to make the Oppenheimer-Lanczos formula for the leading asymptotics of Stark resonances in hydrogen a mathematical theorem.

What then does the theory of Schrödinger operators accomplish? There is an occasional contribution to other parts of physics, especially in areas such as the theory of resonances or the theory of random impurities, where there are mathematical subtleties with real physical significance. Also, there are important spinoffs to mathematics. To name two areas, operator theory, especially spectral analysis, and the theory of path integrals have been illuminated in significant ways during the recent past by the theory of Schrödinger operators. But in the end, there is the internal dynamic connected with intellectual honesty — precisely what can and what cannot be explained rigorously from first principles, starting with the basic formalism of nonrelativistic quantum mechanics? There results a theory of great beauty with intimate connection to a wide variety of aspects of both mathematics and theoretical physics. □