Many-Body Stability Implies a Bound on the Fine-Structure Constant

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, P.O. Box 708, Princeton, New Jersey 08544

and

Horng-Tzer Yau [(a)]

School of Mathematics, The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 11 August 1988)

The Dirac equation for hydrogenic atoms has a well known instability when $z\alpha > 1$. A similar instability occurs for the "relativistic Schrödinger equation" with $p^2/2m$ replaced by $(p^2c^2+m^2c^4)^{1/2}-mc^2$ at $z\alpha=2/\pi$. These instabilities concern only the product $z\alpha$, but when the many-electron–many-nucleus problem is examined (in the relativistic Schrödinger theory) we find that a bound on α alone (independent of z) is then required for stability. If $\alpha < 1/94$ we find that stability occurs all the way up to the critical value $z\alpha=2/\pi$, whereas if $\alpha > 128/15\pi$ then the system is unstable for all values of z. Some implications of these findings are also discussed.

PACS numbers: 03.65.−w, 11.10.−z, 12.20.Ds, 31.10.+z

With the discovery of the relativistic Dirac equation for the hydrogen atom came the realization that quantum mechanics requires a bound on the fine-structure constant $\alpha=e^2/\hbar c$. More precisely the product $z\alpha$ (where z is the nuclear charge) cannot be greater than one for, as many textbooks tell us, the Dirac operator is ill defined (i.e, it has no self-adjoint extensions) when $z\alpha > 1$. There are, of course, corrections to this simple Dirac picture which might prevent collapse when $z\alpha > 1$; these include the nuclear motion (which smears out the Coulomb singularity) and numerous quantum electrodynamic (QED) effects. The point, however, is not that the collapse can be ameliorated but rather that the world of ordinary matter as we know it is radically different when $z\alpha > 1$. Instead of the Bohr radius being the relevant length scale, the radius of a nucleus or the Compton wavelength of the electron or some other small length scale becomes relevant. It is far from clear whether or not there is a true "phase transition" associated with some critical value of $z\alpha \simeq 1$, but in any case it is likely that there is a dramatic change in the physics of matter. Current experiments with the collision of heavy ions are now being undertaken to elucidate this realm of $z\alpha > 1$.

Now let us pose a question that should have been asked at the beginning, but to our knowledge was not —possibly because its precise formulation and its answer were difficult. Suppose that there are $K > 1$ nuclei of charge z and located at distinct points $R_1,\ldots,R_K \in \mathbf{R}^3$. Suppose each nucleus is subcritical, e.g., $z=1$, $\alpha=\frac{1}{137}$, but $Kz\alpha > 1$. We assume the customary approximation that the nuclear masses are infinite and we suppose there is only one electron. For each choice of distinct, $R_1,\ldots,R_K$ there is no problem, but there definitely will be a problem if we permit the nuclei to come together at a common point R, since the electron will then feel an attraction $Kz\alpha/r$ with $Kz\alpha > 1$. By length scaling one sees that, from the point of view of the electron's energy, it is favorable for the nuclei to do this.

The question is then, *within the framework of ordinary quantum mechanics (without invoking finite nuclear mass and QED effects) what prevents the nuclei from coming together and forming some kind of "sink"?*

The answer, presumably, is the nuclear-nuclear Coulomb repulsion which can be written (with $\hbar=c=1$, which we assume henceforth) as

$$V_{\rm rep}=\frac{1}{\alpha}(z\alpha)^2\sum_{1\le i<j\le k}|R_i-R_j|^{-1}. \qquad (1)$$

From (1) we expect that when $z\alpha$ is fixed it will be necessary to require α to be sufficiently small if $V_{\rm rep}$ is to prevent collapse. Thus, stability of ordinary quantum mechanics (with relativistic kinetic energy and without nuclear motion) somehow sets an *a priori* upper bound on the fine-structure constant. We emphasize that this bound is intrinsically a many-body effect. Suppose one allows z to be very small (nonintegral). Then one-body stability will only set a bound on αz and α can be very big provided z is correspondingly small. The above discussion suggests that α must be kept small independent of the value of z; in short, *many-body stability sets a bound on α alone which is independent of z.*

If α is larger than a certain critical value (which we estimate below) then electrons will bind nuclei together. The nuclei will be prevented from coming to a single point because of nuclear forces, but the implication of this binding for nuclear physics would be the following. For small α, nuclei can have only a finite size because of Coulomb repulsion, i.e., infinitely extended nuclear matter does not exist. With α sufficiently large, however, the presence of electrons would stabilize nuclear matter of arbitrarily large size.

It is the purpose of this Letter to announce several results of ours which make the above considerations more precise. The details appear elsewhere.[1]

We are not aware of the existence of any truly relativistic formulation of many-body quantum mechanics and among the many possible caricatures of such a theory we have adopted the simplest possible one which displays the essential feature of relativistic kinematics, namely the replacement of $p^2/2m$ by $(p^2+m^2)^{1/2}-m$. We do not try to deal with the Dirac operator $\not{p}$, not because it is too complicated but because the Dirac operator plus potential is not bounded below. We are unable to deal with the question of filling the "negative energy sea" —and it is not even clear how to do so precisely in the many-body context. We want to deal with a well defined Hamiltonian and be able to use the variational principle for the ground-state energy in order to decide unambiguously whether this ground-state energy is finite or whether it is $-\infty$.

Our choice for the kinetic-energy operator for N electrons is

$$\tilde{T}=\sum_{i=1}^{N}(p^2+m^2)^{1/2}-m\,, \tag{2}$$

with $p^2=-\nabla^2$ acting on ordinary scalar valued functions which satisfy the Pauli principle. This $\tilde{T}$ is well defined as a multiplication operator in Fourier space. It is then well defined in position space although it is highly nonlocal.

Our Hamiltonian is then

$$\tilde{H}=\tilde{T}+\alpha V\,, \tag{3}$$

where V is the Coulomb potential

$$V=\sum_{1\le i<j\le N}|x_i-x_j|^{-1}-\sum_{i=1}^{N}\sum_{j=1}^{K}|x_i-R_j|^{-1}+\sum_{1\le i<j\le K}z_iz_j|R_i-R_j|^{-1}. \tag{4}$$

Here, the x_i's are the electron coordinates and the z_i's are the nuclear charges. Neutrality is *not* assumed.

Our $\tilde{H}$ has many obvious shortcomings, but we remind the reader that a similar choice was made by Chandrasekhar[2] to explain the collapse of white dwarf stars. In that case one replaces α by $G(M/z)^2$ (with M = nuclear mass and z = nuclear charge) and V by the attractive gravitational energy $-\sum_{1\le i<j\le K}|x_i-x_j|^{-1}$. The resulting theory is at least qualitatively correct and it has been shown rigorously,[3,4] that the conventional mean-field analysis of the gravitational $\tilde{H}$ is indeed correct in the physical limit $G(M/z)^2\to 0$ and $N^{2/3}G(M/z)^2$ fixed.

Clearly, a more realistic model for white dwarfs is the Hamiltonian $\tilde{H}$ in (3) plus the gravitational energy. The critical mass in this model is thought to be[5] only slightly different from the Chandrasekhar value by a term of order α. Some rigorous bounds for this critical mass are given in Ref. 6. Our study of the Hamiltonian $\tilde{H}$ could also be viewed as a first step toward a rigorous understanding of this model.

Returning to (3), we define the ground-state energy for fixed R_i's by

$$\tilde{E}_{N,K}(R_1,\ldots,R_K)=\inf\langle\psi|\tilde{H}|\psi\rangle/\langle\psi|\psi\rangle\,, \tag{5}$$

where the infimum is over all ψ's satisfying the Pauli principle with q spin states per electron. Of course, $q=2$ in nature, but it is academically interesting to study the dependence of the critical α on q. In particular, $q=N$ is the case of "bosonic electrons." Next, we define $\tilde{E}_{N,K}$ to be the minimum of $\tilde{E}_{N,K}(R_1,\ldots,R_K)$ over all choices of the R_i's.

There are two simple remarks about $\tilde{E}_{N,K}$. (i) Since $|p|>(p^2+m^2)^{1/2}-m>|p|-m$ we see that replacing $\tilde{T}$ by

$$T=\sum_{j=1}^{N}|p_i|\,, \tag{6}$$

and $\tilde{H}$ by $H=T+\alpha V$ and $\tilde{E}_{N,K}$ by $E_{N,K}$ (analogously defined) does not change the stability problem. There is the bound $E_{N,K}>\tilde{E}_{N,K}>E_{N,K}-mN$. It is convenient to study $E_{N,K}$ instead of $\tilde{E}_{N,K}$ because, by simple length scaling, there are only two cases:

$$E_{N,K}=0 \text{ or } -\infty\,. \tag{7}$$

We say that *the system is stable if* $E_{N,K}=0$. (ii) It has been shown[7] that if there is some number z such that $z_i\le z$ for all $i=1,\ldots,K$ and if stability holds when all the nuclear charges are set equal to z, then stability holds for the original choice $z_1,\ldots,z_K$. Therefore, for simplicity, we can consider the case that all z_i have a common value z.

We say that *the Hamiltonian H (or $\tilde{H}$) is globally stable for a given z and α if $E_{N,K}=0$ for all choices of N and K. Otherwise H is unstable.* Our goal is to delineate regions in z,α space of stability and instability.

Suppose there is one nucleus and one electron (the hydrogenic problem). In this case it has been shown[8,9] that stability occurs for $E_{1,1}$ if and only if $z\alpha\le 2/\pi$. This is just like the situation in Dirac theory (except that 1 is replaced by $2/\pi$) and the underlying reason for instability is the same in both cases; the role of spinors is not central. In quantum mechanics $|p|$ scales like length^{-1} and so does the Coulomb potential. If $z\alpha\langle\psi|\,|x|^{-1}|\psi\rangle>\langle\psi|\,|p|\,|\psi\rangle$ for some ψ, then by length contraction we can drive $E_{N,K}$ to $-\infty$. From this we learn that in the many-body case we at least require $\beta\equiv(\pi/2)z\alpha$ to be at most 1 in order to have stability. The point that was far from clear is whether *global* stability can occur for $\beta=1$ and some $\alpha>0$, or whether global stability requires a further limit on β (possibly zero).

The first result in this direction was by Daubechies and Lieb[7] who showed that $E_{1,K}$ is stable for all K if $\beta\le 1$ and $\alpha\le 1/3\pi$. The first true many-body result was due to Conlon[10] who showed global stability if $z=1$, $q=1$ (q enters now), and $\alpha < 10^{-200}$. Fefferman and de la Llave[11] improved this to $z=1, q=1$, and $\alpha \leq 1/2.06\pi$; unfortunately, this does not cover the critical case $\beta=1$ and, more importantly, it does not generalize easily to arbitrary q.

We have succeeded in proving stability up to $\beta=1$ and all q by reducing the problem to a tractable one-body problem. Our main result is as follows.

Theorem 1.—If $z\alpha \leq 2/\pi$ and if $q\alpha \leq \frac{1}{47}$ then H and $\tilde{H}$ are globally stable.

What about instability? Is it really true that large α will cause collapse? We have proved two results in this direction which, when taken together, give a fairly complete picture of the collapse (except for numerical constants).

Theorem 2.—If $\alpha > 128/15\pi$ then for every q and every $z>0$, however small, H and $\tilde{H}$ are globally unstable. More precisely, $E_{1,K} = -\infty$ for some sufficiently large K (i.e., one electron can make a "bomb").

Theorem 3.—For arbitrary q and $z>0$, H and $\tilde{H}$ are globally unstable when $\alpha > 36 q^{-1/3} z^{-2/3}$.

A corollary of Theorem 3 is that H is always globally unstable for "bosonic electrons" for all α and $z>0$. This follows from the remark that $q=N$ for bosons.

Some slightly more refined estimates are given in Ref. 1, but the main point is that the Coulomb repulsion can stabilize the many-body system if and only if α is small enough. When $q=2$ and $z\alpha$ is at its maximum value $2/\pi$, we have proved that the critical α is between 1/94 and $128/15\pi$. The exact value is unknown but we would guess from the kinds of estimates we have used that the true value is about 1 in the model given by Eq. (3).

This work was partially supported by U.S. National Science Foundation Grants No. PHY-85-15288-A02 (E.L.) and No. DMS-8601978 (H.-T.Y.).

[a]Permanent address: Courant Institute of Mathematical Sciences, New York University, New York, NY 10012.

[1]E. H. Lieb and H.-T. Yau, Commun. Math. Phys. **118**, 177 (1988).

[2]S. Chandrasekhar, Philos. Mag. **11**, 592 (1931), and Rev. Mod. Phys. **56**, 137 (1984).

[3]E. H. Lieb and H. T. Yau, Commun. Math. Phys. **112**, 147 (1987).

[4]E. H. Lieb and H. T. Yau, Astrophys. J. **323**, 140 (1987).

[5]S. Shapiro and S. Teukolsky, *Black Holes, White Dwarfs and Neutron Stars* (Wiley, New York, 1983).

[6]E. H. Lieb and W. E. Thirring, Ann. Phys. (N.Y.) **155**, 494 (1984).

[7]I. Daubechies and E. H. Lieb, Commun. Math. Phys. **90**, 497 (1983).

[8]I. Herbst, Commun. Math. Phys. **53**, 285 (1977), and **55**, 316(E) (1977).

[9]T. Kato, *Perturbation Theory for Linear Operator* (Springer-Verlag, New York, 1966). See remark 5.12, p. 307.

[10]J. G. Conlon, Commun. Math. Phys. **94**, 439 (1984).

[11]C. Fefferman and R. de la Llave, Rev. Math. Iberoamericana **2**, 119 (1986). See also C. Fefferman, Commun. Pure Appl. Math. Suppl. **39**, S67 (1986).