A Solution to the Problem of Variational Collapse for the One-Particle Dirac Equation

Robert Nyden Hill and Carl Krauthauser
Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716
(Received 14 May 1993)

The problem of variational collapse for one-particle Dirac Hamiltonians H_{Dirac} is solved by applying the Rayleigh-Ritz variational principle to the operator $1/H_{\text{Dirac}}$ instead of to the Dirac Hamiltonian itself. The variational trial functions $|\tilde{\phi}\rangle$ are taken to have the form $|\tilde{\phi}\rangle = H_{\text{Dirac}}|\tilde{\psi}\rangle$, where $|\tilde{\psi}\rangle$ is a linear combination of basis functions. Only the matrix elements of H_{Dirac} and of H_{Dirac}^2 are then needed.

PACS numbers: 31.30.Jv, 03.65.Ge, 03.65.Pm, 31.15.+q

The conventional Rayleigh-Ritz variational principle as applied to the Dirac equation is merely a stationary principle. It is not a minimum principle because the Dirac Hamiltonian is not bounded below; it is not a maximum principle because the Dirac Hamiltonian is not bounded above. This is the well known problem of variational collapse, which causes difficulty when basis set methods are used to construct approximations to the bound states of a Dirac equation. In a number of cases of interest, this difficulty can be circumvented by applying the variational principle to an appropriately chosen function of the Dirac Hamiltonian H_{Dirac} instead of to the Dirac Hamiltonian itself. The function used maps the spectrum of the Dirac Hamiltonian in one to one fashion onto the spectrum of a bounded self-adjoint operator, for which the variational principle is both a maximum principle and a minimum principle.

A number of proposals, none of which we find completely satisfactory, have been made for avoiding the variational collapse problem. A review of these was given by Kutzelnigg [1] in 1984. The problem has been discussed more recently by Stanton and Havriliak [2], Baylis and Peel [3], Goldman [4], Talman [5], Rosenberg and Spruch [6], Grant and Quiney [7], Ishikawa and Sekino [8], Dyall and Fægri [9], and LaJohn and Talman [10].

We will solve the variational collapse problem for those Dirac Hamiltonians whose spectrum has a gap between the positive and negative energy states by applying the variational principle to $G(H_{\text{Dirac}}) = 1/H_{\text{Dirac}}$ instead of to the Dirac Hamiltonian itself. We will also consider the more general choice $G(W; H_{\text{Dirac}}) = 1/(H_{\text{Dirac}} - W)$, where W is a real number which is not in the spectrum of H_{Dirac}; by appropriate choices of W the energy eigenvalues can be bracketed by upper and lower bounds. Bounds on the errors in the wave functions in the L^2 norm can also be obtained.

We begin with the fact that the Dirac Hamiltonian is self-adjoint [11]. This allows us to define functions of the Dirac Hamiltonian via the spectral theorem for self-adjoint operators [12]. Specifically, for a function $F(x)$ of a real variable x which is not too wild [12] we define the corresponding function $F(H_{\text{Dirac}})$ of H_{Dirac} by

$$F(H_{\text{Dirac}}) = \sum_j F(E_j)\,|e_j\rangle\langle e_j|. \tag{1}$$

The numbers E_j and vectors $|e_j\rangle$ are the eigenvalues and eigenvectors of H_{Dirac}; the sum over j in (1) runs over *all* of the spectrum of H_{Dirac}, including the continuum. We have written our definition (1) in the notation used by most quantum mechanics texts for the benefit of the reader who is uncomfortable with the projection valued measures used by Reed and Simon [12]. If $F(x)$ is real for all x in the spectrum of H_{Dirac}, $F(H_{\text{Dirac}})$ is self-adjoint. If $F(x)$ is bounded for all x in the spectrum of H_{Dirac}, $F(H_{\text{Dirac}})$ is bounded. We will consider the specific choices $F(x) = G(x) = 1/x$ and $F(x) = G(W;x) = 1/(x - W)$, both of which give rise to self-adjoint operators. $G(H_{\text{Dirac}}) = 1/H_{\text{Dirac}}$ is bounded (by $1/\epsilon$) if there is a real $\epsilon > 0$ such that the region $|x| \le \epsilon$ contains no points of the spectrum of H_{Dirac}. Similarly, $G(W; H_{\text{Dirac}}) = 1/(H_{\text{Dirac}} - W)$ is bounded if $|x - W| \le \epsilon$ is not in the spectrum of H_{Dirac}.

The basic idea can be understood by looking at the picture of the spectrum of H_{Dirac} shown in Fig. 1 and the picture of the spectrum of $1/H_{\text{Dirac}}$ shown in Fig. 2. The pictures have been drawn for a general H_{Dirac} which has both positive and negative energy bound states; in particular cases either or both sets of bound states may be absent. The points in the spectrum of $1/H_{\text{Dirac}}$, which will be denoted by the Greek letter λ, are just the reciprocals of the points in the spectrum of H_{Dirac}, which will be denoted by E. In particular, the gap between the positive and negative energy states in the spectrum of the Dirac Hamiltonian H_{Dirac} is mapped to the exterior of the spectrum of $1/H_{\text{Dirac}}$, the lowest positive energy state is mapped to the top of the spectrum of $1/H_{\text{Dirac}}$, and the highest negative energy state is mapped to the

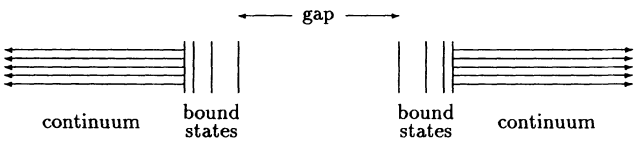

FIG. 1. The spectrum of the Dirac Hamiltonian H_{Dirac}.

0031-9007/94/72(14)/2151(4)$06.00

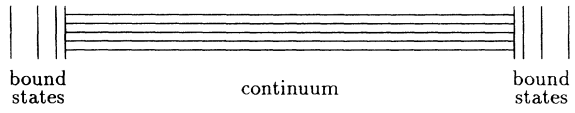

FIG. 2. The spectrum of $G(H_{\rm Dirac}) = 1/H_{\rm Dirac}$.

bottom of the spectrum of $1/H_{\rm Dirac}$. The Rayleigh quotient for an operator is an average over the spectrum of the operator and therefore cannot lie in the exterior of that spectrum. Thus the Rayleigh-Ritz variational principle applied to $1/H_{\rm Dirac}$ can never yield a value of λ for which the corresponding energy $E = 1/\lambda$ lies in the gap between the positive and negative energy states. If a positive energy bound state exists, maximizing the Rayleigh quotient for $1/H_{\rm Dirac}$ with respect to a variational trial function yields approximations $\tilde{\lambda}_+$ to the largest eigenvalue of $1/H_{\rm Dirac}$ which converge monotonically from below. The corresponding approximations $\tilde{E}_+$ to the lowest positive energy bound state of $H_{\rm Dirac}$, which are calculated via $\tilde{E}_+ = 1/\tilde{\lambda}_+$, converge monotonically from above, just as for the ordinary Rayleigh-Ritz principle applied to a nonrelativistic Hamiltonian. Similarly, if a negative energy bound state exists, minimizing the Rayleigh quotient for $1/H_{\rm Dirac}$ yields approximations $\tilde{\lambda}_-$ to the smallest eigenvalue of $1/H_{\rm Dirac}$ which converge monotonically from above. The corresponding approximations $\tilde{E}_-$ to the highest negative energy bound state of $H_{\rm Dirac}$, calculated via $\tilde{E}_- = 1/\tilde{\lambda}_-$, converge monotonically from below.

The difficult problem of calculating matrix elements of the inverse of an operator is avoided by taking the variational trial functions $|\tilde{\phi}\rangle$ to have the form $|\tilde{\phi}\rangle = H_{\rm Dirac}|\tilde{\psi}\rangle$ where $|\tilde{\psi}\rangle$ is a linear combination of basis functions. The Rayleigh quotient R_G for $G(H_{\rm Dirac}) = 1/H_{\rm Dirac}$ then takes the form

$$R_G = \frac{\langle\tilde{\phi}|G|\tilde{\phi}\rangle}{\langle\tilde{\phi}|\tilde{\phi}\rangle} = \frac{\langle\tilde{\psi}|H_{\rm Dirac}|\tilde{\psi}\rangle}{\langle\tilde{\psi}|H^2_{\rm Dirac}|\tilde{\psi}\rangle}, \tag{2}$$

and only matrix elements of $H_{\rm Dirac}$ and $H^2_{\rm Dirac}$ are needed. The choice of this form for the trial functions $|\tilde{\phi}\rangle$ does not impose any restriction on $|\tilde{\phi}\rangle$ beyond the obvious requirement that $|\tilde{\phi}\rangle$ be square integrable, because there exists a $|\tilde{\psi}\rangle$ such that $|\tilde{\phi}\rangle = H_{\rm Dirac}|\tilde{\psi}\rangle$ for every $|\tilde{\phi}\rangle$ in the space L^2 of square integrable functions if $|x| \leq \epsilon$ is not in the spectrum of $H_{\rm Dirac}$. It has been suggested that the reciprocal of the usual Rayleigh quotient R_H, which is

$$\frac{1}{R_H} = \frac{\langle\tilde{\psi}|\tilde{\psi}\rangle}{\langle\tilde{\psi}|H_{\rm Dirac}|\tilde{\psi}\rangle}, \tag{3}$$

could be recovered from the Rayleigh quotient R_G by using $|\tilde{\phi}\rangle = H^{1/2}_{\rm Dirac}|\tilde{\psi}\rangle$. This proposal fails because $H^{1/2}_{\rm Dirac}$ as defined by (1) is not self-adjoint (note that $x^{1/2}$ is imaginary for x negative). This lack of self-adjointness implies that the product of the ket vector $H^{1/2}_{\rm Dirac}|\tilde{\psi}\rangle$ with the corresponding bra vector is not $\langle\tilde{\psi}|H_{\rm Dirac}|\tilde{\psi}\rangle$, and that the expectation value of $1/H_{\rm Dirac}$ with respect to $H^{1/2}_{\rm Dirac}|\tilde{\psi}\rangle$ is not $\langle\tilde{\psi}|\tilde{\psi}\rangle$. Thus there is no simple relationship between the variational principle for R_G and the usual variational principle for R_H.

We will demonstrate the variational collapse, and illustrate our method of avoiding it, by considering

$$H_{\rm Dirac} = mc^2\left[Z_\alpha \boldsymbol{\alpha}\cdot\left(-i\frac{\partial}{\partial \mathbf{r}}\right) + \beta - \frac{Z_\alpha^2}{r}\right], \tag{4}$$

which is the Dirac Hamiltonian for a hydrogen atom with a point nucleus in dimensionless form. $\boldsymbol{\alpha}$ and β are the usual Dirac matrices. Z_α is the dimensionless number which is the product of the nuclear charge Z (in units of the electronic charge) with the fine structure constant $e^2/\hbar c$. The Bohr radius has been used as the unit of length in obtaining the form (4); this choice of length scale makes nonrelativistic intuition easily available. The solution to $H_{\rm Dirac}|\psi\rangle = E|\psi\rangle$ has the form

$$\psi_\pm(\mathbf{r}) = \begin{pmatrix} A_\pm(r)\,\mathcal{Y}^{j,m}_{j\mp\frac{1}{2}} \\ -ia_\pm(r)\,\mathcal{Y}^{j,m}_{j\pm\frac{1}{2}} \end{pmatrix}, \tag{5}$$

where, for positive energy states, $A_\pm(r)$ is the radial part of the large component and $-ia_\pm(r)$ is the radial part of the small component. For negative energy states, $A_\pm(r)$ is the radial part of the small component and $-ia_\pm(r)$ is the radial part of the large component. $\mathcal{Y}$ is the two component spinor

$$\mathcal{Y}^{\ell\pm\frac{1}{2},m}_\ell = \frac{1}{\sqrt{2\ell+1}}\begin{pmatrix} \pm\sqrt{\ell\pm m+\frac{1}{2}}\,Y_{\ell,m-\frac{1}{2}} \\ \sqrt{\ell\mp m+\frac{1}{2}}\,Y_{\ell,m+\frac{1}{2}} \end{pmatrix}, \tag{6}$$

where the $Y_{\ell,m\pm\frac{1}{2}}$ are spherical harmonics. We will use variational trial functions of the form (5) in which $A_\pm(r)$ and $a_\pm(r)$ are replaced by finite length basis set expansions.

We consider first a simple trial function of the form (5) with

$$A_\pm(r) = C\exp(-\mu r), \quad \text{and} \quad a_\pm(r) = 0. \tag{7}$$

Here μ is a (nonlinear) variational parameter and C is a normalizing constant. It is straightforward to show that, with this trial function, the Rayleigh quotient R_H for the usual variational principle is

$$R_H = \frac{\langle\tilde{\psi}|H_{\rm Dirac}|\tilde{\psi}\rangle}{\langle\tilde{\psi}|\tilde{\psi}\rangle} = mc^2\left(1-\mu Z_\alpha^2\right). \tag{8}$$

Obviously R_H has neither a maximum nor a minimum as a function of μ, and can be made to take on any value below mc^2 by an appropriate choice of μ (note that μ must be positive to preserve the square integrability of the trial function). This is the phenomenon of variational collapse. On the other hand, the Rayleigh quotient R_G for $G(H_{\rm Dirac}) = 1/H_{\rm Dirac}$, computed with the trial function $|\tilde{\phi}\rangle = H_{\rm Dirac}|\psi\rangle$, is

$$R_G = \frac{\langle\tilde{\phi}|G|\tilde{\phi}\rangle}{\langle\tilde{\phi}|\tilde{\phi}\rangle} = \frac{\langle\tilde{\psi}|H_{\text{Dirac}}|\tilde{\psi}\rangle}{\langle\tilde{\psi}|H_{\text{Dirac}}^2|\tilde{\psi}\rangle} = \frac{1-\mu Z_\alpha^2}{mc^2\left[Z_\alpha^2\left(1+2Z_\alpha^2\right)\mu^2 - 2Z_\alpha^2\mu + 1\right]}. \tag{9}$$

The Rayleigh quotient (9) is easily shown to have one maximum, at $\mu = \mu_+$, and one minimum, at $\mu = \mu_-$, where

$$\mu_\pm = \frac{1}{Z_\alpha^2}\left(1 \mp \sqrt{\frac{1+Z_\alpha^2}{1+2Z_\alpha^2}}\right). \tag{10}$$

The maximum μ_+ yields an upper bound $\tilde{E}_+$ to the lowest positive energy state which is

$$\tilde{E}_+ = -\frac{2mc^2}{Z_\alpha^2}\left(1+Z_\alpha^2\right)\left(1-\sqrt{\frac{1+2Z_\alpha^2}{1+Z_\alpha^2}}\right) = mc^2\left[1 - \tfrac{1}{4}Z_\alpha^2 + O\left(Z_\alpha^4\right)\right]. \tag{11}$$

The minimum μ_- yields a lower bound $\tilde{E}_-$ to the highest negative energy state which is

$$\tilde{E}_- = -\frac{2mc^2}{Z_\alpha^2}\left(1+Z_\alpha^2\right)\left(1+\sqrt{\frac{1+2Z_\alpha^2}{1+Z_\alpha^2}}\right) = mc^2\left[-\frac{4}{Z_\alpha^2} - 3 + O\left(Z_\alpha^2\right)\right]. \tag{12}$$

The upper bound $\tilde{E}_+$ gets the leading term mc^2 of the expansion about the nonrelativistic limit right. The next term differs from the correct nonrelativistic ground state energy $-\frac{1}{2}mc^2Z_\alpha^2$ by a factor of 2, but is obviously an upper bound as it must be. The lower bound $\tilde{E}_-$ is very bad, as is to be expected, since we have set the component $a_\pm(r)$, which is the large component for negative energy states, to zero in our variational trial function (7). We consider next a trial function of the form (5) in which $A_\pm(r)$ and $a_\pm(r)$ are given by basis set expansions of the form

$$A_\pm(r) = (\alpha r)^{-\gamma}\exp\left(-\tfrac{1}{2}\alpha^2r^2\right)\left[\sum_{m=0}^{M} a_m L_m^{(-\frac{1}{2}-\gamma)}\left(\alpha^2r^2\right) + \sum_{j=1}^{J} b_j\,(\alpha r)^{2j-1}\right], \tag{13}$$

$$a_\pm(r) = (\alpha r)^{-\gamma}\exp\left(-\tfrac{1}{2}\alpha^2r^2\right)\left[\sum_{n=0}^{N} c_n L_n^{(-\frac{1}{2}-\gamma)}\left(\alpha^2r^2\right) + \sum_{k=1}^{K} d_k\,(\alpha r)^{2k-1}\right]. \tag{14}$$

The trial functions (13) and (14) are bad enough to show what is important for rapid convergence and what is not. The expansions (13) and (14) need not be of the same length; M and J in (13) need not equal N and K in (14). Neither small r boundary conditions nor relations between large and small components need be imposed; the only requirement on the trial function $|\tilde{\psi}\rangle$ is that $\langle\tilde{\psi}|H_{\text{Dirac}}^2|\tilde{\psi}\rangle$ be finite. The $L_m^{(-\frac{1}{2}-\gamma)}$ and $L_n^{(-\frac{1}{2}-\gamma)}$ are Laguerre polynomials in the notation which is standard in the mathematics literature [13]. α and γ are nonlinear variational parameters. Table I illustrates the convergence to the exact hydrogenic ground state energy, which is 0.999 973 373 984 656 675 074 281 052 966..., when $\alpha = 1/\sqrt{2}$, $N = M/2$, and $K = J$. The value of γ has been chosen to get the leading term of the small r behavior right; nonzero values of $K = J$ get additional terms of the small r behavior right. The value 1/137.036 04 has been used for Z_α. Examination of the table shows that the variational approximation converges monotonically to the exact energy from above. A rate-of-convergence theory for the variational method advocated here can be developed which is similar to the theory given by Hill [14] for the nonrelativistic Schrödinger equation; a detailed discussion of this theory, which is too long to be presented here, will be published elsewhere.

The eigenvalues of the Dirac Hamiltonian can be bracketed with upper and lower bounds by applying the Rayleigh-Ritz variational method to the more general choice $G(W; H_{\text{Dirac}}) = 1/(H_{\text{Dirac}} - W)$. In this case the trial functions $|\tilde{\phi}\rangle$ are chosen to have the form $|\tilde{\phi}\rangle = (H_{\text{Dirac}} - W)|\tilde{\psi}\rangle$, where $|\tilde{\psi}\rangle$ is a linear combina-

TABLE I. Errors when $\alpha = 2^{-1/2}$, $N = M/2$, and $K = J$.

M	$K=J=0$	$K=J=1$	$K=J=2$	$K=J=3$
2	0.655×10^{-5}	0.871×10^{-6}	0.263×10^{-6}	0.802×10^{-7}
4	0.256×10^{-5}	0.130×10^{-6}	0.804×10^{-7}	0.212×10^{-7}
6	0.164×10^{-5}	0.598×10^{-7}	0.198×10^{-7}	0.112×10^{-7}
8	0.118×10^{-5}	0.161×10^{-7}	0.810×10^{-8}	0.422×10^{-8}
10	0.927×10^{-6}	0.879×10^{-8}	0.257×10^{-8}	0.185×10^{-8}
20	0.412×10^{-6}	0.558×10^{-9}	0.381×10^{-10}	0.302×10^{-10}
40	0.171×10^{-6}	0.531×10^{-10}	0.735×10^{-13}	0.458×10^{-13}
60	0.995×10^{-7}	0.135×10^{-10}	0.255×10^{-14}	0.251×10^{-15}
80	0.671×10^{-7}	0.506×10^{-11}	0.467×10^{-15}	0.290×10^{-17}
100	0.492×10^{-7}	0.236×10^{-11}	0.136×10^{-15}	0.647×10^{-19}

tion of basis functions. If W is chosen to lie in the gap between any two bound states, lower bounds to the states below W, and upper bounds to the states above W, are obtained (this is essentially the Lehmann-Maehley [15] approach to the Temple-Kato [16] lower bound). Errors in the wave functions in the L^2 norm can be bounded with the aid of the Eckart [17] bound to the overlap. Only the Gram (overlap) matrix and the matrix elements of H_{Dirac} and H^2_{Dirac} are needed for the computation of these bounds. The method outlined here can be applied whenever the spectrum of the operator whose eigenvalues are required contains gaps; the condition that W not be in the spectrum of the operator can then be satisfied by placing W in one of the gaps. For example, an attractive Coulomb potential which is so strong that the lowest bound state lies below zero but above $-mc^2$ can be handled by using $G(W; H_{\text{Dirac}}) = 1/(H_{\text{Dirac}} - W)$ with W chosen to lie in the gap between the lowest bound state and $-mc^2$.

Because matrix eigenvectors cannot be accurately determined when the corresponding eigenvalues are close together [18], eigenvector calculations should be performed by using $1/(H_{\text{Dirac}} - W)$, with W chosen close enough to the energy eigenvalue of interest so that the corresponding eigenvalue λ of $1/(H_{\text{Dirac}} - W)$ is well separated from all other eigenvalues. A "relativistic Sturmian" basis set which gives a discrete approximation to the Dirac spectrum suitable for performing the sums over intermediate states which arise in such problems as two-photon decay and the calculation of polarizabilities [19] can be obtained by taking *all* of the eigenvectors of the matrix eigenvalue problem which is obtained from the variational principle for the eigenvalues. However, the presence of the numerical difficulty noted above implies that calculations with such relativistic Sturmians should be carried out via analytically equivalent procedures which avoid the explicit matrix diagonalization. These issues will be discussed in more detail elsewhere.

The method outlined here is an adaptation of the Lehmann-Maehley approach to the Temple-Kato lower bound, which was explained to the first author (R.N.H.) by David W. Fox a number of years ago. The idea of applying this approach to the Dirac equation grew out of a conversation with S. P. Goldman, who suggested the trial function (7). The work has been supported by NSF Grant No. PHY91-06797.

[1] W. Kutzelnigg, Int. J. Quantum Chem. **25**, 107 (1984).
[2] R. E. Stanton and S. Havriliak, J. Chem. Phys. **81**, 1910 (1984).
[3] W. E. Baylis and S. J. Peel, Phys. Rev. A **28**, 2552 (1984).
[4] S. P. Goldman, Phys. Rev. A **31**, 3541 (1985).
[5] J. D. Talman, Phys. Rev. Lett. **57**, 1091 (1986).
[6] Leonard Rosenberg and Larry Spruch, Phys. Rev. A **34**, 1720 (1986). A generalization to the two-electron problem has appeared recently in Leonard Rosenberg, Phys. Rev. A **47**, 1771 (1993).
[7] I. P. Grant and H. M. Quiney, Adv. At. Mol. Phys. **23**, 37 (1987). See also H. M. Quiney, I. P. Grant, and S. Wilson, J. Phys. B **23**, L271–L278 (1990).
[8] Y. Ishikawa and H. Sekino, Chem. Phys. Lett. **165**, 243 (1990).
[9] K. G. Dyall and K. Fægri, Jr., Chem. Phys. Lett. **174**, 25 (1990).
[10] L. LaJohn and J. D. Talman, Chem. Phys. Lett. **189**, 383 (1992).
[11] Bernd Thaller, *The Dirac Equation* (Springer, New York, 1992), pp. 112–114. The history of the problem of self-adjointness of Dirac operators is discussed on pp. 305–306.
[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis* (Academic Press, New York, 1972), Chaps. VII and VIII. See Theorem VIII.6 on pp. 263–264.
[13] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, pp. 188–192; W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), 3rd ed., pp. 240–249.
[14] R. N. Hill, J. Chem. Phys. **83**, 1173 (1985).
[15] N. J. Lehman, Z. Angew. Math. Mech. **29**, 342 (1949); **30**, 1 (1950); H. J. Maehly, Helv. Phys. Acta **25**, 547 (1952). A nice review can be found in Sec. 3 of N. W. Bazley and D. W. Fox, J. Res. Nat. Bur. Standards **68B**, 173–183 (1964).
[16] G. Temple, Proc. R. Soc. London A **119**, 276 (1928); T. Kato, J. Phys. Soc. Jpn. **4**, 334 (1949).
[17] C. Eckart, Phys. Rev. **36**, 878 (1930). An extensive discussion of this and related bounds can be found in F. Weinhold, J. Math. Phys. **11**, 2127 (1970).
[18] Beresford N. Parlett, *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980), pp. 14–15.
[19] G. W. F. Drake and S. P. Goldman, Phys. Rev. A **23**, 2093 (1981); S. P. Goldman and G. W. F. Drake, Phys. Rev. A **25**, 2877 (1982); G. W. F. Drake and S. P. Goldman, Adv. At. Mol. Phys. **25**, 393 (1988).