# Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters

M. P. Nightingale
*University of Rhode Island*, nightingale@uri.edu

Vilen Melik-Alaverdian
*University of Rhode Island*

# Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters

Terms of Use
All rights reserved under copyright.

# Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters

M. P. Nightingale and Vilen Melik-Alaverdian

*Department of Physics, University of Rhode Island, Kingston, Rhode Island 02881*
(Received 26 October 2000; published 9 July 2001)

A quantum Monte Carlo method is introduced to optimize excited-state trial wave functions. The method is applied in a correlation function Monte Carlo calculation to compute ground- and excited-state energies of bosonic van der Waals clusters of up to seven particles. The calculations are performed using trial wave functions with general three-body correlations.

Weakly bound clusters display strong anharmonicity, and this makes solving the Schrödinger equation for such systems a formidable computational challenge. The discrete variable representation (DVR) method [1] was applied with success to compute the energies of vibrational states for systems with up to six degrees of freedom. Its computational complexity scales exponentially with dimension, a problem that Monte Carlo methods can avoid. Indeed, correlation function Monte Carlo [2,3] and the projector operator imaginary-time spectral evolution (POITSE) [4] methods are applicable to higher-dimensional systems, although in practice they are restricted to a smaller number of excited states.

The accuracy of Monte Carlo projection methods can be improved dramatically by employing approximate eigenfunctions. In fact, without good initial approximations, one rarely obtains results of sufficient accuracy. In this Letter, we discuss a systematic and efficient method to construct approximate eigenfunctions by optimization of many-parameter trial functions. We then use these functions in a correlation function Monte Carlo calculation. We expect that POITSE calculations can also be improved by the same means. A variant of the method described here was applied previously to study critical dynamics of lattice systems [5,6].

We compute energy levels of bosonic van der Waals clusters of atoms of mass $\mu$, interacting via a Lennard-Jones potential with core radius $\sigma$ and well depth $\epsilon$. In dimensionless form, the pair potential is $r^{-12} - 2r^{-6}$ and the Hamiltonian is $\mathcal{H} = P^2/2m + V$; $P^2/2m$ and $V$ are the total kinetic and potential energies. The only parameter is the inverse dimensionless mass $m^{-1} = \hbar^2/\mu\sigma^2\epsilon$, which is proportional to the square of the de Boer parameter [7].

We use the position representation, and denote by $R$ the Cartesian coordinates of a cluster of $N_c$ atoms. For the parameter optimization we generate a sample of configurations $R_\sigma$, $\sigma = 1, \ldots, \Sigma$, sampled with relative probabilities $\psi_g(R_\sigma)^2$; the choice of the guiding function $\psi_g^2$ will be discussed later. The sample typically consists of several thousand configurations, which are kept fixed during the optimization.

The trial functions are linear combinations of about one hundred *elementary basis functions*, each of which depends on nonlinear parameters. Correspondingly, we have a linear optimization nested in a nonlinear one. The result is a set of functions serving as *basis functions* in a correlation function Monte Carlo calculation. These basis functions are constructed one at a time, from the ground state up, as follows.

Suppose we fix at initial values the nonlinear parameters of the elementary basis functions denoted by $\beta_i$, $i = 1, \ldots, n$. Ideally, these functions span an $n$-dimensional invariant subspace of the Hamiltonian $\mathcal{H}$. Then there exists an $n \times n$ matrix $\mathcal{E}$ so that

$$\mathcal{H}\beta_i(R_\sigma) = \sum_j \beta_j(R_\sigma)\mathcal{E}_{ji}. \tag{1}$$

In that case, for $k = 1, \ldots, n$,

$$\tilde{\psi}^{(k)}(R) = \sum_i \beta_i(R)d_i^{(k)} \tag{2}$$

is an eigenvector of $\mathcal{H}$ with an eigenvalue $\tilde{E}_k$ equal to the exact energy $E_k$, if $d^{(k)}$ is a right eigenvector of $\mathcal{E}$ with eigenvalue $\tilde{E}_k$. We rewrite Eq. (1) in matrix form:

$$B' = B\mathcal{E}, \tag{3}$$

where $B_{\sigma i} = \hat{\beta}_i(R_\sigma)$ and $B'_{\sigma i} = \hat{\beta}'_i(R_\sigma)$, with $\hat{\beta}_i = \beta_i/\psi_g$ and $\hat{\beta}'_i = \mathcal{H}\beta_i/\psi_g$.

In practice, the subspace spanned by the basis functions is not invariant, so that, for given matrices $B$ and $B'$, Eq. (3) is an overdetermined set of equations for the unknown matrix $\mathcal{E}$. If one multiplies through, from the left by $B^T$, the transpose of $B$, one obtains by inversion

$$\mathcal{E} = (B^T B)^{-1}(B^T B') \equiv N^{-1}H. \tag{4}$$

As readily verified, this is the least-squares solution of Eq. (3); note that the rows of the matrices $B$ and $B'$ are weighted by the guiding function so that the elements of the matrices $N$ and $H$ approach the standard quantum mechanical overlap integrals and matrix elements in the limit of an infinite Monte Carlo sample. Equations (2) and (4) are usually derived from stationarity (in the linear

parameters) of the average energy. If the latter is estimated by a finite-sample average, requiring stationarity of this estimate yields Eq. (4) with $H$ replaced by its *symmetrized* analog. Since the exact quantum mechanical expression is indeed symmetrical, one might be inclined to use the symmetrized $H$. However, only the nonsymmetric expression $B^T B'$ in Eq. (4) satisfies the zero-variance principle of yielding exact results independent of the configuration sample if the basis functions span an invariant subspace of the Hamiltonian. As in the ideal case, Eq. (2) determines the linearly optimized trial functions, but now one has $E_k \lesssim \tilde{E}_k$, an inequality [8] which for a finite Monte Carlo sample may be violated because of statistical noise.

The solution for $\mathcal{E}$ as written in Eq. (4) is numerically unstable since the matrix $N$ is ill conditioned because of the near-linear dependence of the $\beta_k$. The solution to this problem [9,10] is to use a singular value decomposition to obtain a numerically regularized inverse $B^{-1}$ [11]. In terms of the latter, one finds, from Eq. (3),

$$\mathcal{E} = B^{-1}B'. \qquad (5)$$

With the linear variational parameters optimized for fixed nonlinear parameters in the elementary basis functions, we optimize—following Umrigar *et al.* [12]—the nonlinear parameters by minimization of

$$\chi^2 = \frac{\sum_\sigma [\hat{\psi}^{(k)\prime}(R_\sigma) - \tilde{E}_k \hat{\psi}^{(k)}(R_\sigma)]^2}{\sum_\sigma \hat{\psi}^{(k)}(R_\sigma)^2}, \qquad (6)$$

the variance of the local energy of the wave function given in Eq. (2); again, the guiding function is incorporated via $\hat{\psi}^{(k)} = \tilde{\psi}^{(k)}/\psi_g$ and $\hat{\psi}^{(k)\prime} = \mathcal{H}\tilde{\psi}^{(k)}/\psi_g$.

We now address the choice of the guiding function $\psi_g$. To obtain acceptable statistical errors, the sample has to have sufficient overlap with the desired excited states. In our case, this can be accomplished [2] with $\psi_g^p = \tilde{\psi}^{(1)}$ with a power $p$ in the range $2 \lesssim p \lesssim 3$, while the ground-state wave function $\tilde{\psi}^{(1)}$ is obtained after a few initial iterations.

The elementary basis functions [13,14] are the final ingredient of the computation. They are defined as functions of all interatomic distances $r_{\sigma\tau}$ and scaled variables $\hat{r}_{\sigma\tau} = f(r_{\sigma\tau})$. Here, $f$ maps the interatomic distances monotonically onto the interval $(-1, 1)$ such that most of the variation occurs where the wave function differs most from zero; the explicit form of $f$ is not important for the current discussion.

The elementary basis functions used for energy level $k$ have nonlinear variational parameters $a_j^{(k)}$, and are of the form

$$\beta_i(R) = s_i(R)\exp\left[ \sum_j a_j^{(k)} s_j(R) - \sum_{\sigma<\tau}\left(\kappa_k r_{\sigma\tau} + \frac{\sqrt{m}}{5r_{\sigma\tau}^5}\right)\right]. \qquad (7)$$

$\sigma, \tau, \upsilon = 1, \ldots, N_c$ are atom indices. The polynomial $s_i$ is characterized by three non-negative integral powers $n_{il}$:

$$s_i(R) = \sum_{\sigma<\tau<\upsilon} \prod_{l=1}^{3} (\hat{r}_{\sigma\tau}^l + \hat{r}_{\tau\upsilon}^l + \hat{r}_{\upsilon\sigma}^l)^{n_{il}}. \qquad (8)$$

The prefactor polynomial $s_i$ has bosonic symmetry, and contains general three-body correlations, since all polynomials symmetric in $x$, $y$, and $z$ can be written as polynomials in the three invariants $I_l = x^l + y^l + z^l$, with $l = 1, 2, 3$ and vice versa [13]. The number of elementary basis functions is limited by a bound on the total degree $\sum_l \ln_{il}$; the polynomials $s_j$ in the exponent are of the same form as those in the prefactor, and their number in Eq. (7) is similarly limited.

The constant $\kappa_k$ is determined self-consistently so that the wave function has the correct exponential decay in the limit that a single atom goes off to infinity. Assuming—as is plausible for the small clusters studied here—that the energy of a cluster is roughly proportional to the number of atom pairs [15], we find

$$\kappa_k = \frac{2}{N_c - 1}\sqrt{\frac{-m\tilde{E}_k}{N_c}}. \qquad (9)$$

The $r_{\sigma\tau}^{-5}$ term in Eq. (7) ensures that $\mathcal{H}\beta_i/\beta_i$ has a weaker divergence than with $r_{\sigma\tau}^{-12}$ in the limit $r_{\sigma\tau} \to 0$ [13].

States of higher energy are found with the same optimization scheme by using the appropriate eigenvector $d^{(k)}$ of the matrix $\mathcal{E}$ in Eq. (2). We use the same scaling function $f$ for all states, but different nonlinear parameters $a_j^{(k)}$ and $\kappa_k$. This scheme works as long as the trial functions possess the variational freedom accurately to represent the eigenstates. Otherwise, for the Monte Carlo samples of the size we are using, states may be skipped or spuriously introduced. We found it useful to check the consistency of eigensystems obtained with this basis set and one that includes the variational wave functions of previously determined, lower-lying states.

Instead of the hybrid method discussed in this Letter, which treats linear and nonlinear variational parameters differently, we originally attempted to use minimization of the variance of the energy for both types of parameters simultaneously. Although that works for statistical mechanical applications [6], it fails here, except for the lowest two or three energy levels. We could only make this work (and even then only for small de Boer parameters) for higher levels by first constructing approximate wave functions using a conventional approximation, and then fitting these functions by the basis functions used in this Letter. Finally, the parameter values obtained in these fits served as starting values for further optimization [16].

We used the optimized trial wave functions as basis functions in a correlation function Monte Carlo calculation [2,3]. Formally, this means that the $n$ elementary basis functions $\beta_i$ are replaced by a small number of functions $\exp(-\frac{1}{2}t\mathcal{H})\tilde{\psi}^{(k)}$. For this part of the computation, the analog of Eq. (4) is used to compute eigenvalues, rather than Eq. (5). The singular value decomposition (SVD)

cannot be used for the correlation function Monte Carlo because there are too many configurations $R_\alpha$ to store the required matrices $B$ and $B'$. However, since the optimized basis functions are few in number and roughly orthonormal—at least for small projection times $t$—the SVD is not essential in this case.

Before we present estimates of the excited state energies, we discuss the sources of error of this method [2]. In addition to the statistical errors, there are two systematic errors. For any finite projection time $t$ and in the limit of vanishing statistical errors, the energies computed by this method are upper bounds to the exact energies [8]. In practice, since the statistical errors increase with projection time, one should choose the smallest projection time such that the projection and statistical errors are of the same order of magnitude. To pinpoint that time, one has to distinguish real trends from false trends due to correlated noise. This is always tricky, but a troublesome detail is that at that point the results tend to have a non-Gaussian distribution [17], which makes it difficult to produce error bars with a well-defined statistical meaning. In addition, there is the time-step error, which arises because the imaginary-time evolution operator $\exp(-t\mathcal{H})$ has to be evaluated as the limit $\tau \to 0$ of $[\exp(-\tau\mathcal{H}) + O(\tau^2)]^{t/\tau}$, but this error is much easier to control.

Next we present results for excited-state energies for clusters with up to seven atoms. We computed energies for trimers of Ne, Ar, Kr, and Xe ($m^{-1} = 7.092 \times 10^{-3}$, $6.9635 \times 10^{-4}$, $1.9128 \times 10^{-4}$, and $7.8508 \times 10^{-5}$). Since our variational functions contain general three-body correlations, the accuracy of the wave functions and energies for the trimers can be improved without any apparent limit other than the machine precision. During optimization of the wave functions for the trimers, we typically start with the ground-state wave function which has a prefactor degree of 5 or 6. For the trimers we chose not to vary the polynomial coefficients in the exponent and simply used the fixed terms required by the boundary conditions. The quality of the wave functions may be improved by varying polynomial coefficients in the exponent, and for larger clusters it becomes important to include such polynomials.

For the optimization we used samples consisting of 4000 configurations, and we gradually increased the prefactor degree to improve the quality of the trial functions. For Ne trimers we performed diffusion Monte Carlo [18] calculations using optimized wave functions with prefactor degrees up to 14. Although, in principle, for trimers nothing should preclude further improvement, the observed changes were statistically insignificant in the 12 to 14 degree range. Table I contains results for degree 12.

There was no statistically significant difference between time steps $\tau = 0.4$, 0.2, and 0.1, and thus no noticeable time-step error. In the diffusion Monte Carlo calculations we used 1.3 million Monte Carlo steps. For Ar, Kr, and Xe trimers we found that the quality of the wave functions

TABLE I. Energy levels $E_k$ of the rare gas trimers; the errors are estimated to be a few units in the least significant digit.

| $k$ | Ne$_3$ | Ar$_3$ | Kr$_3$ | Xe$_3$ |
|---|---|---|---|---|
| 1 | $-1.719\,560$ | $-2.553\,289\,43$ | $-2.760\,555\,34$ | $-2.845\,241\,50$ |
| 2 | $-1.222\,83$ | $-2.250\,185\,5$ | $-2.581\,239\,0$ | $-2.724\,955\,8$ |
| 3 | $-1.142\,0$ | $-2.126\,361$ | $-2.506\,946\,8$ | $-2.675\,064\,8$ |
| 4 | $-1.038$ | $-1.996\,43$ | $-2.412\,444$ | $-2.608\,615$ |
| 5 | $-0.890$ | $-1.946\,7$ | $-2.387\,973$ | $-2.592\,226$ |

does not improve beyond the prefactor degree 10. The results in Table I for the three more massive noble gas atoms were obtained using trial wave functions with such polynomials. Convergence with respect to the time step was established by comparing $\tau = 0.8$, 0.6, and 0.4. The number of Monte Carlo steps is the same as for Ne. Except for the energy of the fifth level of Ne, which is 0.009 too high, the results in Table I agree with, and in some cases improve upon, those of Leitner *et al.* [15].

In Table II we present results for the energies of the first five levels of Ar clusters of sizes four through seven. Our method allows one to go beyond seven atom clusters, but as one can see from Table II the statistical error increases with system size. To obtain more accurate results for larger clusters it would probably be helpful to include higher order correlations in the wave function. In the calculations for four through seven atom clusters we used a 10 degree prefactor and an exponent of degree 3. Again, 1.3 million step diffusion Monte Carlo results were compared for $\tau = 0.8$, 0.6, and 0.4.

As to the performance of our method as the mass $m$ decreases and the atoms become more weakly bound, we find that both the optimization and the projection methods begin to fail, because the elementary basis functions lack the required variational freedom. This breakdown is illustrated in Fig. 1, which contains three energy levels as a function of mass for a 4 atom cluster. The results are plotted using variables chosen so that there is linear dependence both for large masses and for energies close to zero [14]. As the energy of the levels approaches zero, the scatter in the data points increases, and ultimately the method fails to produce reliable results. Again, the use of trial wave functions with four-body correlations is likely to make it possible to continue to smaller masses.

TABLE II. Energy levels of Ar clusters of up to seven atoms; the errors are estimated to be a few units in the least significant digit.

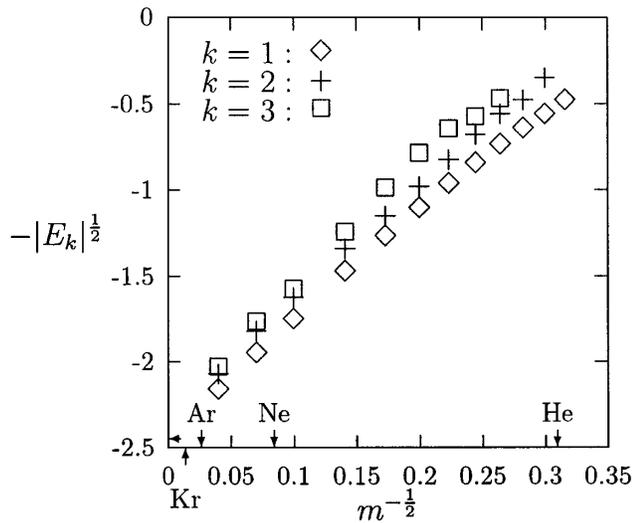| $k$ | Ar$_4$ | Ar$_5$ | Ar$_6$ | Ar$_7$ |
|---|---|---|---|---|
| 1 | $-5.118\,11$ | $-7.785\,1$ | $-10.887\,9$ | $-14.191$ |
| 2 | $-4.785$ | $-7.567$ | $-10.561$ | $-13.969$ |
| 3 | $-4.674$ | $-7.501$ | $-10.51$ | $-13.80$ |
| 4 | $-4.530$ | $-7.39$ | $-10.46$ | $-13.74$ |
| 5 | $-4.39$ | $-7.36$ | $-10.35$ | $-13.71$ |

FIG. 1.   $-\sqrt{-E_k}$ of lowest three levels ($k = 1, 2, 3$) for 4 atom clusters vs $m^{-1/2}$. The estimated errors for most energies are smaller than the plot symbols, and increase for decreasing mass. Missing data points indicate that no reliable estimates were obtained. The vertical arrows indicate Kr, Ar, Ne, and He; the horizontal arrow indicates the classical value $-\sqrt{6}$.

As mentioned earlier, exponential scaling with dimensionality limits the applicability of the DVR method. We can only speculate how the Monte Carlo method discussed here scales, since the accuracy of the results is determined mostly by the elusive quality of the basis functions. It is precisely the degradation of the trial functions which is responsible for the big differences in accuracy among the results we presented. Clearly, more highly excited states have more structure, but the harmonic approximation suggests that the corresponding increase in complexity scales with a low power of the excitation level. It is plausible that it suffices to include in the elementary basis functions $n$-body correlations up to some finite order $n$ only, which suggest polynomial complexity for the computation of the basis functions. Much evidence suggests that the projection stage of the calculations also scales with a small power of system size. We performed the projection part of the calculations by a variant of pure diffusion Monte Carlo [2,19]. The statistical noise of this approach for large systems increases exponentially, but there are alternatives to avoid this, and there it is likely that the usual power law behavior of diffusion Monte Carlo methods [20] can be recovered. Let it suffice to say that the longer runs typically took a few hours, on a Four Processor SGI Origin 200, to produce the results presented here.

[1] D. H. Zhang, Q. Wu, J. Z. H. Zhang, M. von Dirke, and Z. Bačić, J. Chem. Phys. **102**, 2315 (1995).

[2] D. M. Ceperley and B. Bernu, J. Chem. Phys. **89**, 6316 (1988).

[3] B. Bernu, D. M. Ceperley, and W. A. Lester, Jr., J. Chem. Phys. **93**, 552 (1990); W. R. Brown, W. A. Glauser, and W. A. Lester, Jr., J. Chem. Phys. **103**, 9721 (1995).

[4] D. Blume, M. Lewerenz, P. Niyaz, and K. B. Whaley, Phys. Rev. E **55**, 3664 (1997); D. Blume and K. B. Whaley, J. Chem. Phys. **112**, 2218 (2000), and references therein.

[5] M. P. Nightingale and H. W. J. Blöte, Phys. Rev. Lett. **80**, 1007 (1998).

[6] M. P. Nightingale and H. W. J. Blöte, Phys. Rev. B **62**, 1089 (2000).

[7] J. de Boer, Physica (Utrecht) **14**, 139 (1948).

[8] J. K. L. MacDonald, Phys. Rev. **43**, 830 (1933).

[9] G. H. Golub and C. F. van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1989), 2nd ed., Chap. 5.5.

[10] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1992), Sec. 2.6.

[11] We ignored all singular values $\sigma_k$ with $\sigma_k < 10^3 \sigma_1 \epsilon_{dbl}$, where $\epsilon_{dbl}$ is the double precision machine accuracy.

[12] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988); C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, in *Computer Simulation Studies in Condensed Matter Physics, Recent Developments,* edited by D. P. Landau, K. K. Mon, and H. Schüttler, Springer Proceedings in Physics (Springer, Berlin, 1988).

[13] Andrei Mushinski and M. P. Nightingale, J. Chem. Phys. **101**, 8831 (1994).

[14] M. Meierovich, A. Mushinski, and M. P. Nightingale, J. Chem. Phys. **105**, 6498 (1996).

[15] D. M. Leitner, J. D. Doll, and R. M. Whitnell, J. Chem. Phys. **94**, 6644 (1991).

[16] V. Melik-Alaverdian and M. P. Nightingale, Int. J. Mod. Phys. C **10**, 1409 (2000).

[17] J. H. Hetherington, Phys. Rev. A **30**, 2713 (1984).

[18] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993), and references therein. We used a single, weighted walker and an accept reject step with $\tau_{eff} = \tau$ (0) for accepted (rejected) moves.

[19] M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988); **88**, 1100 (1988).

[20] M. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).