4 Optimized Quantum Monte Carlo Wave Functions

Arne Lüchow
RWTH Aachen University
Institute of Physical Chemistry, 52056 Aachen

Contents

E. Pavarini, E. Koch, and S. Zhang (eds.)
Many-Body Methods for Real Materials
Modeling and Simulation Vol. 9
Forschungszentrum Jülich, 2019, ISBN 978-3-95806-400-3
http://www.cond-mat.de/events/correl19

1 Introduction

Electron structure quantum Monte Carlo (QMC) refers to a number of Monte Carlo-based techniques for electron structure calculations. In this lecture, we focus on real-space QMC where the electronic Schrödinger equation in first quantized form is the starting point. Recently, a number of Monte Carlo methods have evolved where the Schrödinger equation is treated in the Fock space, i.e., in second quantized form. While this equation has a number of advantages, its exact solution is defined by the finite basis set, and the physical solution is obtained only after basis set extrapolation. The electronic Schrödinger equation

$$H\,\Psi = E\,\Psi\,, \quad \text{with} \qquad H = T + V_{en} + V_{ee} \tag{1}$$

contains kinetic energy of the electrons T, Coulombic electron-nucleus (or electron-core) attraction V_{en}, and the electron-electron repulsion V_{ee}.
There are two seemingly unrelated Monte Carlo methods for obtaining approximate solutions for Eq. (1). The first option is the evaluation of the energy expectation value $\langle H\rangle = \langle\Psi|H|\Psi\rangle$ for a trial wave function Ψ with Monte Carlo integration. The variational principle allows the optimization of the trial wave function Ψ by minimizing $\langle H\rangle$ with respect to parameters of the wave function. This method is usually called *variational Monte Carlo* (VMC). The wave function depends on all electrons, and depends thus on the real coordinates $\mathbf{r}_1, \ldots \mathbf{r}_n$ of all n electrons and, additionally, on the spin state of each electron. Note that the Monte Carlo-based integration allows arbitrary forms for the trial wave function. Therefore, Ψ can deviate from the standard Slater determinant form and, for instance, compact correlated wave functions can be employed easily. The optimization of trial wave functions is the subject of this lecture.
The other option for solving the Schrödinger equation with Monte Carlo methods is a stochastic projection method where the exact solution of Eq. (1) is projected out of a starting wave function. These methods are known under the name projector Monte Carlo or *diffusion Monte Carlo (DMC)* and are, due to the projection, in general more accurate than the VMC methods. Surprisingly, the DMC and the VMC methods are closely related, and the optimized VMC trial wave function can improve the DMC energies substantially as we demonstrate in section 4.
This lecture is organized as follows. The VMC and DMC methods are briefly described in the next section before the wave function optimization techniques are discussed in detail in the subsequent section. Selected applications demonstrating the optimization of trial wave functions are given in the final section.

2 Real-space quantum Monte Carlo methods

2.1 Variational Monte Carlo

In this lecture, the following notation will be used: $\mathbf{r}_i = (x_i, y_i, z_i)$ denotes the cartesian coordinates of electron i while the spin is indicated by the quantum number $m_{s,i} = \pm\frac{1}{2}$.

$$\mathbf{x}_i = (\mathbf{r}_i, m_{s,i}) = (x_i,\ y_i,\ z_i,\ m_{s,i}) \tag{2}$$

collects all electron variables and $\mathbf{R} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ all electrons of the (finite) system. With this notation, the energy expectation of an unnormalized wave function $\Psi(\mathbf{R})$ is given in real space by

$$E_{\mathrm{VMC}} = \frac{\langle \Psi(\mathbf{R}) | H \, | \Psi(\mathbf{R}) \rangle}{\langle \Psi(\mathbf{R}) | \Psi(\mathbf{R}) \rangle}. \tag{3}$$

This ratio can be rewritten as an integral over a probability density $p(\mathbf{R})$ and the *local energy* $E_L(\mathbf{R})$

$$E_{\mathrm{VMC}} = \int E_L(\mathbf{R})\, p(\mathbf{R})\, d\tau, \qquad E_L(\mathbf{R}) = \frac{H\Psi(\mathbf{R})}{\Psi(\mathbf{R})}, \qquad p(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int |\Psi(\mathbf{R})|^2 \, d\mathbf{R}} \tag{4}$$

where the integration extends over all space. Note that $p(\mathbf{R})$ describes the probability density of positions and spins of all electrons simultaneously.
A sample with the probability density $p(\mathbf{R})$ can be obtained efficiently with the *Metropolis-Hastings algorithm* [1, 2] without the necessity of calculating the normalization integral. The integral in Eq. (4) can then be obtained simply as the sample mean of the local energy over a large sample of $p(\mathbf{R})$

$$E_{\mathrm{VMC}} = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} E_L(\mathbf{R}_k)\,, \quad \text{with} \quad \mathbf{R}_k \sim p(\mathbf{R}) \tag{5}$$

where $\sim p$ means 'distributed according to the probability density p.' It is important to note that this energy evaluation requires only the ability to calculate the Laplacian of the trial wave function with respect to all electrons but no integration other than the Monte Carlo integration itself. This obviously allows substantial freedom in the choice of Ψ, and in particular the use of correlation factors coupling the electron coordinates for optimal description of the electron-electron interaction and correlation. Variational Monte Carlo refers to the energy calculation using Monte Carlo integration and Metropolis-Hastings sampling of the probability density. Further details can be found in several review articles, for instance in [3, 4] and in references therein.

2.2 Diffusion Monte Carlo

The diffusion Monte Carlo method yields more accurate energies than the VMC method because in DMC the exact ground state wave function is projected according to

$$\lim_{t \to \infty} e^{-Ht} \Psi \propto \Psi_0 \tag{6}$$

The projection of the ground state wave function is obtained after constructing the corresponding Green function with a short-time approximation which leads to a stochastic process, called *drift-diffusion process*,

$$\mathbf{R}_{k+1} = \mathbf{R}_k + \mathbf{b}(\mathbf{R}_k)\tau + \Delta W_\tau, \qquad \mathbf{R}_0 \sim p(\mathbf{R}) \tag{7}$$

with a discretized time step $\tau = t/N$ and a *drift vector* $\mathbf{b}$ defined with the trial function Ψ

$$\mathbf{b}(\mathbf{R}) = \frac{\nabla\Psi(\mathbf{R})}{\Psi(\mathbf{R})}\,. \tag{8}$$

ΔW_τ denotes a normal variate with mean $\mu = 0$ and variance $\sigma^2 = \tau$. The drift-diffusion process implements the antisymmetry of the projected solution because the drift is directed away from the nodes of the trial function Ψ and is, at the nodal surface, in fact a normal vector of the surface.
The drift-diffusion process is coupled with a weighting process

$$w_{k+1} = w_k e^{-\left[\frac{1}{2}\left(E_L(\mathbf{R}_k)+E_L(\mathbf{R}_{k+1})\right)-E_{\mathrm{ref}}\right]\tau}, \qquad w_0 = 1 \tag{9}$$

where a reference energy E_{ref} is employed to stabilize the process. The weighted drift-diffusion process is implemented by a sample of *walkers* which are propagated according to Eq. (7). Each walker has a weight attached which changes in each step according to Eq. (9). Note that a walker corresponds to the positions of all electrons simultaneously. The stochastic process is therefore capable of describing the electron correlation in real space. It is usually stabilized by a branching/killing process based on the walker weights. As in molecular dynamics, the exact distribution is obtained in the limit of vanishing time step, but also in the long total time limit. With the final distribution of weighted walkers the DMC energy is obtained as a weighted mean

$$E_{\mathrm{DMC}} = \frac{\sum_{k=1}^{N} w_k E_L(\mathbf{R}_k)}{\sum_{k=1}^{N} w_k} \tag{10}$$

with the weight w_k of the kth walker with the coordinates $\mathbf{R}_k$.
VMC and DMC are closely related as the drift-diffusion process Eq. (7) is usually employed as proposed step in the Metropolis-Hastings algorithm. Both algorithms are Markov chains, and after reaching the equilibrium distribution, the energy or other estimators are calculated as 'time' and sample average where the time is not the physical time. Each step in the Markov chain is referred to as *Monte Carlo step*. Furthermore, the VMC distribution $p(\mathbf{R})$ is usually the starting distribution for the DMC process as indicated in Eq (7). But the main connection results from the trial function Ψ whose nodes, or more precisely nodal hypersurface, $\Psi(\mathbf{R}) = 0$ is imposed on the DMC solution via the drift Eq. (8). The DMC algorithm described above is known as fixed-node DMC and results in the long t and short τ limit in the exact ground state energy of the Schrödinger equation with the nodes of the trial function Ψ as additional boundary condition. Without this boundary condition, the stochastic process would converge toward the mathematical ground state of H which is the nodeless bosonic solution. The error resulting from the fixed node approximation is called *fixed-node error*.
The optimization of the VMC energy requires the minimization of the energy expectation value Eq. (5) with respect to the parameters of Ψ. The optimization of the DMC energy would require the variation of the nodes of Ψ. Unfortunately, no efficient method is known for the direct optimization of the nodes and hence usually DMC energies are optimized indirectly via the minimization of the VMC energy.

3 Stochastic optimization of electronic wave functions

The basis of all stochastic optimizations is a large sample of electron configurations $\{\mathbf{R}_k\},\ k = 1,\ldots,K$ with a distribution of $p(\mathbf{R}) = |\Psi(\mathbf{R})|^2 / \int |\Psi(\mathbf{R})|^2\, d\tau$, see Eq. (4), obtained with the Metropolis-Hastings algorithm. The calculation of the sample is costly due to the serial correlation between successive Monte Carlo steps, and it can be attempted to do a *fixed sample optimization* which is a deterministic minimization [5]. As can be expected, the resulting parameters are biased toward this sample, and in practice a small number of iterations with new, independent samples of electron configurations is required.
The trial wave function $\Psi(\mathbf{R},\mathbf{p}) \equiv \Psi(\mathbf{p})$ and the local energy $E_L(\mathbf{R},\mathbf{p}) \equiv E_L(\mathbf{p})$ both depend on all electron positions and a set of parameters collected in a parameter vector $\mathbf{p}$. The VMC energy as sample mean of the local energy depends on the parameters and is simply

$$E_m(\mathbf{p}) = \frac{1}{K}\sum_{k=1}^{K} E_L(\mathbf{R}_k,\mathbf{p}). \tag{11}$$

As the parameter optimization proceeds from the initial parameters $\mathbf{p}_0$ to $\mathbf{p}$ the fixed sample ($\sim\ |\Psi(\mathbf{p}_0)|^2$) does no longer correspond to the distribution $|\Psi(\mathbf{p})|^2$ of the modified wave function. The change in the distribution can be corrected with a weighted mean

$$E_{wm}(\mathbf{p}) = \frac{\sum_{k=1}^{K} w_k E_L(\mathbf{R}_k,\mathbf{p})}{\sum_{k=1}^{K} w_k}, \qquad w_k = \frac{|\Psi(\mathbf{R}_k,\mathbf{p})|^2}{|\Psi(\mathbf{R}_k,\mathbf{p}_0)|^2}. \tag{12}$$

The variance of the local energy

$$V_m(\mathbf{p}) = \frac{1}{K}\sum_{k=1}^{K} \big(E_L(\mathbf{R}_k,\mathbf{p}) - E_m\big)^2 \tag{13}$$

determines the standard error of the sample mean energy $\sqrt{V_m}/\sqrt{K}$. The more accurate the wave function the smaller the local energy fluctuations and thus the variance. For any exact eigenfunction of H the local energy is constant and thus the sample variance exactly zero. This *zero-variance property* of the sample mean as energy estimator allows the determination of energies with small statistical error bars with reasonable sample sizes provided that accurate (i.e. low variance) wave functions are used. It is important for an efficient optimization to construct similar estimators for instance for gradients with respect to parameters.
For the optimization, the gradients of the wave function and the local energy with respect to the parameters are required. Here, we denote the parameter derivatives as follows

$$\Psi_i = \frac{\partial\Psi(\mathbf{p})}{\partial p_i}, \qquad \Psi_{ij} = \frac{\partial^2\Psi(\mathbf{p})}{\partial p_i \partial p_j}, \qquad E_{L,i} = \frac{\partial E_L(\mathbf{p})}{\partial p_i}\,. \tag{14}$$

Minimization of the VMC energy with respect to the parameters is desired similarly to the Hartree-Fock or configuration interaction (CI) methods. Minimization of the variance is, in principle, equivalent to energy minimization because only exact eigenfunctions of H have a variance of zero. Variance minimization is substantially simpler and more efficient because the

sample variance is a sum of squares, while the local energy is possibly not bounded from below. In earlier QMC work the variance minimization was preferred while in more recent papers the energy minimization dominates because of a higher accuracy of the results.

3.1 Variance and MAD minimization

The sample variance is often simplified with an estimate $E_{\rm ref}$ of the mean energy (which is easily available from a previous DFT calculation)

$$V_r(\mathbf{p}) = \frac{1}{K}\sum_{k=1}^{K}\left(E_L(\mathbf{R}_k,\mathbf{p}) - E_{\rm ref}\right)^2 \tag{15}$$

The variance minimization is a nonlinear least-squares minimization problem that can be solved with standard methods such as the Levenberg-Marquardt method. The minimization requires the Jacobian, the derivatives $E_{L,i}(\mathbf{R}_k)$ of the residuals $E_L(\mathbf{R}_k,\mathbf{p})-E_{\rm ref}$, and has been employed for many years [6]. We have obtained best results with the standard code 'nl2sol' for nonlinear least squares problems [7]. The variance minimization is very stable and suitable even for initial parameters that differ substantially from the optimal ones. It is therefore often used as a first optimization step followed by an energy minimization. Nonlinear least-squares minimizations require a number of iterations. If the fixed sample is used, the variance defined in Eq. (15) ignores the change of the distribution $|\Psi(\mathbf{p})|^2$ with the parameters which can be accounted for by using a weighted mean as in Eq. (12)

$$V_{wm}(\mathbf{p}) = \frac{\sum_{k=1}^{K} w_k (E_L(\mathbf{R}_k,\mathbf{p}) - E_{\rm ref})^2}{\sum_{k=1}^{K} w_k} \tag{16}$$

Unfortunately, the weighted variance tends to be unstable when the weights start to deviate substantially from one. In this case, the effective sample size is substantially reduced. Investigations have shown that the unweighted variance Eq. (13) or Eq. (15) leads to more stable and efficient minimizations [8].
Alternatively, it is possible to optimize the mean absolute deviation (MAD) of the local energy

$$\mathrm{MAD}_r(\mathbf{p}) = \frac{1}{K}\sum_{k=1}^{K}\left|E_L(\mathbf{R}_k,\mathbf{p}) - E_{\rm ref}\right| \tag{17}$$

Since a sum of positive elements is involved the MAD minimization is also stable and found to be advantageous in some cases [9].

3.2 Energy minimization

Energy minimization is preferable to variance minimization when the goal is to calculate energy differences. As mentioned above, energy minimization requires more computational effort meaning substantially larger samples than variance minimization. Due to the stochastic nature of QMC most codes use specifically optimized routines based on standard methods. The main

variants for energy optimization are Newton-like methods, the linear method, and the stochastic reconfiguration method. The methods require different derivatives and matrix elements and differ in efficiency. For the stochastic optimization of wave functions the number of iterations to reach a convergence threshold is only one aspect of the efficiency, another being the computational effort for each step.

The computational effort for each optimization step depends on the variance of the estimators for the required gradient or matrix elements. In many cases, different estimators can be constructed that have the same limit but show different variances. Substantial work has been done in the past in finding low variance estimators, in particular estimators whose variance vanishes as the trial wave function approaches the exact eigenfunction of H.

In this section, we follow the notation used by Toulouse and Umrigar who gave an excellent account of energy minimization methods including a discussion of earlier work [10]. The expectation value over $p(\mathbf{R})$ is denoted

$$\langle A \rangle = \int A(\mathbf{R})\, p(\mathbf{R})\, d\mathbf{R}, \qquad p(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int |\Psi(\mathbf{R})|^2\, d\mathbf{R}}, \tag{18}$$

and the corresponding estimator is the sample mean

$$\langle A \rangle = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} A(\mathbf{R}_k), \qquad \mathbf{R}_k \sim p(\mathbf{R}). \tag{19}$$

3.2.1 Newton methods

When expanding $E_{\mathrm{VMC}}(\mathbf{p})$ around $\mathbf{p}_0$ with $\mathbf{p} = \mathbf{p}_0 + \Delta\mathbf{p}$ in a Taylor series up to second order, the Newton method is obtained leading to the parameter vector change

$$\Delta\mathbf{p} = -\mathbf{h}^{-1}\mathbf{g} \tag{20}$$

with the gradient and Hessian matrix of the VMC energy

$$g_i = \frac{\partial E_{\mathrm{VMC}}}{\partial p_i}, \qquad h_{ij} = \frac{\partial^2 E_{\mathrm{VMC}}}{\partial p_i \partial p_j} \tag{21}$$

It is numerically more stable to solve the linear system $\mathbf{h}\Delta\mathbf{p} = -\mathbf{g}$ rather than inverting the estimator of the Hessian matrix. The Newton method has been first used for energy optimization in VMC by Lin et al. [11] and later improved by Umrigar and Filippi [12] and Sorella [13]. Newton methods are known for their limited convergence radius. One common stabilization technique is the adaptive addition of a positive constant to the diagonal of the Hessian matrix $\mathbf{h}$ which has the effect of switching smoothly with growing constant to steepest descent. For details see reference [10] and references therein.

The greatest challenge for the use of the Newton method for energy minimization is the construction of low variance estimators for the gradient and the Hessian matrix. For the gradient of the VMC energy we first observe that

$$\frac{\partial E_{\mathrm{VMC}}}{\partial p_i} = \frac{\partial \langle E_L \rangle}{\partial p_i} \neq \langle E_{L,i} \rangle \tag{22}$$

because of the dependence of the distribution $p(\mathbf{R}) = p(\mathbf{R}, \mathbf{p})$ on the parameters. An estimator of the gradient can be derived as follows

$$\begin{aligned}
\frac{\partial}{\partial p_i}\langle E_L\rangle &= \frac{\partial}{\partial p_i}\frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} \\
&= \frac{\frac{\partial}{\partial p_i}\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} - \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}\frac{\frac{\partial}{\partial p_i}\langle\Psi|\Psi\rangle}{\langle\Psi|\Psi\rangle} \\
&= \frac{\langle\Psi_i|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} + \frac{\langle\Psi|H|\Psi_i\rangle}{\langle\Psi|\Psi\rangle} - 2\langle E_L\rangle\frac{\langle\Psi_i|\Psi\rangle}{\langle\Psi|\Psi\rangle} \\
&= \left\langle\frac{\Psi_i}{\Psi}E_L\right\rangle + \left\langle\frac{H\Psi_i}{\Psi}\right\rangle - 2\langle E_L\rangle\left\langle\frac{\Psi_i}{\Psi}\right\rangle
\end{aligned} \tag{23}$$

Comparing

$$E_{L,i} = \frac{H\Psi_i}{\Psi} - E_L\frac{\Psi_i}{\Psi} \;\Rightarrow\; \langle E_{L,i}\rangle = \left\langle\frac{H\Psi_i}{\Psi}\right\rangle - \left\langle\frac{\Psi_i}{\Psi}E_L\right\rangle \tag{24}$$

one obtains

$$\frac{\partial}{\partial p_i}\langle E_L\rangle = 2\left[\left\langle\frac{\Psi_i}{\Psi}E_L\right\rangle - \left\langle\frac{\Psi_i}{\Psi}\right\rangle\langle E_L\rangle\right] + \langle E_{L,i}\rangle \tag{25}$$

The last term vanishes due to Hermicity of H

$$\langle E_{L,i}\rangle = \left\langle\frac{H\Psi_i}{\Psi}\right\rangle - \left\langle\frac{\Psi_i}{\Psi}\frac{H\Psi}{\Psi}\right\rangle = \frac{\langle\Psi|H|\Psi_i\rangle}{\langle\Psi|\Psi\rangle} - \frac{\langle\Psi_i|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} = 0 \tag{26}$$

and we obtain finally [11]

$$\frac{\partial E_{\mathrm{VMC}}}{\partial p_i} = 2\left[\left\langle\frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})}E_L(\mathbf{R})\right\rangle - \left\langle\frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})}\right\rangle\langle E_L(\mathbf{R})\rangle\right]. \tag{27}$$

Note that this estimator of the parameter derivative of the VMC energy does not depend on the parameter derivative of the local energy! The wave function derivative Ψ_i arises only in a ratio with the wave function Ψ itself. Normalization constants, which are expensive to calculate in QMC, therefore cancel. Furthermore, this estimator has the zero-variance property because the two terms in the difference become identical as the local energy becomes a constant for the exact eigenfunction of H. The difference is calculated with correlated sampling, using the same sample for both terms such that fluctuations of $\Psi_k(\mathbf{R})/\Psi(\mathbf{R})$ cancel substantially. Additionally, the estimator has the form of a covariance, and it has been observed that the fluctuations of the covariance $\langle xy\rangle - \langle x\rangle\langle y\rangle$ are much smaller than fluctuations of either x or y [12].

In the light of this result, a number of different estimators for the Hessian matrix $\mathbf{h}$ have been proposed and employed that are based on the ratios Ψ_i/Ψ, Ψ_{ij}/Ψ and the covariance form. For the Hessian matrix, no estimator without the parameter derivative of the local energy $E_{L,i}$ is known. The estimators are a bit lengthy and not reproduced here. For a discussion of the various estimators of the Hessian matrix see reference [10].

The implementation of Newton-type methods requires therefore implementations of the following three terms for every type of trial wave function

$$E_{L,i}(\mathbf{R}), \quad \frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})}, \quad \frac{\Psi_{ij}(\mathbf{R})}{\Psi(\mathbf{R})} \tag{28}$$

Below, we discuss the computational effort for these terms for some common wave function and Hamiltonian types.

3.2.2 Linear method

If the wave function depends linearly on the parameters as in CI or Jastrow-CI wave functions, the parameters are obtained non-iteratively by diagonalization of the Hamiltonian matrix. This approach can be extended to nonlinear parameters by expanding the wave function to first order in the parameters

$$\Psi^{(1)}(\mathbf{p}) = \Psi^{(1)}(\mathbf{p}_0 + \Delta\mathbf{p}) = \Psi(\mathbf{p}_0) + \sum_{i=1}^{n_p} \Delta p_i \Psi_i(\mathbf{p}_0) \tag{29}$$

The idea is now to consider $\Psi_i = \Psi_i(\mathbf{p}_0)$, $i = 0, \ldots, n_p$, with $\Psi_0 \equiv \Psi(\mathbf{p}_0)$ and the number of parameters n_p as a basis and obtain the energy minimum by solving the matrix eigenvalue problem

$$\mathbf{H}\Delta\mathbf{p} = E\mathbf{S}\Delta\mathbf{p} \tag{30}$$

with the Hamilton $H_{ij} = \langle\Psi_i|H|\Psi_j\rangle$ and overlap matrices $S_{ij} = \langle\Psi_i|\Psi_j\rangle$. Note that $\Delta p_0 = 1$ in Eq. (29) which defines the normalization of the resulting eigenvector corresponding to the intermediate normalization in standard quantum chemistry. In the case of linear parameters the CI method is recovered. In the case of nonlinear parameters the linear expansion is repeated iteratively until convergence.

Nightingale and Melik-Alaverdian demonstrated that it is advantageous and, in fact, leads to a zero-variance property if the finite sample estimators of the Hamiltonian matrix elements do not make use of the Hermicity of the Hamiltonian [14, 10]. This results for finite samples in *unsymmetric* Hamiltonian matrices in the eigenvalue problem and in parameter changes with substantially reduced variances.

In practice, the dependence of the normalization constant of the wave function on the parameter change is accounted for. The normalized wavefunction

$$\tilde{\Psi}_0(\mathbf{p}) = \frac{1}{\sqrt{\langle\Psi(\mathbf{p})|\Psi(\mathbf{p})\rangle}}\Psi(\mathbf{p}) \tag{31}$$

results in the parameter derivatives

$$\tilde{\Psi}_i(\mathbf{p}) = \frac{1}{\sqrt{\langle\Psi(\mathbf{p})|\Psi(\mathbf{p})\rangle}}\left(\Psi_i(\mathbf{p}) - \left\langle\frac{\Psi_i(\mathbf{p})}{\Psi_0(\mathbf{p})}\right\rangle\Psi(\mathbf{p})\right) \tag{32}$$

with

$$\frac{\langle\Psi(\mathbf{p})|\Psi_i(\mathbf{p})\rangle}{\langle\Psi(\mathbf{p})|\Psi(\mathbf{p})\rangle} = \left\langle\frac{\Psi_i(\mathbf{p})}{\Psi_0(\mathbf{p})}\right\rangle. \tag{33}$$

The first order expansion now reads

$$\tilde{\Psi}^{(1)}(\mathbf{p}_0 + \Delta\mathbf{p}) = \tilde{\Psi}_0(\mathbf{p}_0) + \sum_{i=1}^{n_p} \Delta p_i \tilde{\Psi}_i(\mathbf{p}_0)\,. \tag{34}$$

The finite sample estimators for the matrices are constructed as follows. H_{00} is simply the current sample mean of the local energy $H_{00} = \langle E_L \rangle$. The H_{i0} elements have the estimator

$$H_{i0} = \langle \tilde{\Psi}_i | H | \tilde{\Psi}_0 \rangle = \int \frac{\tilde{\Psi}_i(\mathbf{R})}{\tilde{\Psi}_0(\mathbf{R})} E_L(\mathbf{R})\, p(\mathbf{R})\, d\mathbf{R} = \left\langle \frac{\Psi_i}{\Psi_0} E_L \right\rangle - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \langle E_L \rangle \tag{35}$$

with

$$\frac{\tilde{\Psi}_i}{\tilde{\Psi}_0} = \frac{\Psi_i}{\Psi_0} - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle . \tag{36}$$

For the H_{0i} elements we obtain with Eq. (32) and without making use of Hermicity

$$H_{0i} = \langle \tilde{\Psi}_0 | H | \tilde{\Psi}_i \rangle = \left\langle \left(\frac{1}{\Psi_0} H \Psi_i - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle E_L \right) \right\rangle = \langle E_{L,i} \rangle + \left\langle \frac{\Psi_i}{\Psi_0} E_L \right\rangle - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \langle E_L \rangle \tag{37}$$

using

$$E_{L,i} = \frac{1}{\Psi_0} \left(\frac{\partial H \Psi_0}{\partial p_i} - E_L \frac{\partial \Psi_0}{\partial p_i} \right) . \tag{38}$$

Analogously, one can derive for the remaining matrix elements H_{ij} with $i, j > 0$ [10]

$$\begin{aligned} H_{ij} \;=\; & \left\langle \frac{\Psi_i}{\Psi_0} \frac{\Psi_j}{\Psi_0} E_L \right\rangle - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \left\langle \frac{\Psi_j}{\Psi_0} E_L \right\rangle - \left\langle \frac{\Psi_j}{\Psi_0} \right\rangle \left\langle \frac{\Psi_i}{\Psi_0} E_L \right\rangle \\ & + \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \left\langle \frac{\Psi_j}{\Psi_0} \right\rangle \langle E_L \rangle + \left\langle \frac{\Psi_i}{\Psi_0} E_{L,j} \right\rangle - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \langle E_{L,j} \rangle \end{aligned} \tag{39}$$

and for the overlap matrix

$$S_{ij} = \left\langle \frac{\Psi_i}{\Psi_0} \frac{\Psi_j}{\Psi_0} \right\rangle - \left\langle \frac{\Psi_i}{\Psi_0} \right\rangle \left\langle \frac{\Psi_j}{\Psi_0} \right\rangle , \qquad S_{00} = 1 . \tag{40}$$

Note that the estimator for the Hamilton matrix of the linear method requires the local energy derivatives $E_{L,i}$ like the estimator for the Hessian matrix of the Newton method.

In practice, after the generalized "almost symmetric" eigenvalue problem in Eq. (30) has been solved, the correct right eigenvector has to be identified. For finite samples the correct eigenvector may not correspond to the lowest eigenvalue as spurious solutions may occur. Since the eigenvectors are orthogonal it is not difficult to identify the physical solution. A reliable way is a very short VMC calculation with new parameters arising from all eigenvectors of a few lowest eigenvalues. Alternatively, the overlap of the new wavefunction with the previous can be estimated, see [10, 15] for further discussions. For a large number of parameters the linear method becomes restricted by the necessity to store the Hamilton and overlap matrices. Zhao and Neuscammam have recently presented a blocked linear method that has been shown to be a very good approximation to the original linear method while requiring substantially less memory which is particularly important for the parallel implementation [16]. The linear method, in one of its variants, turns out to be the most efficient energy minimization method in most applications.

3.2.3 Stochastic reconfiguration, energy fluctuation potential, and perturbative optimization methods

Both the linear method and the Newton method require the calculation of the local energy parameter derivative $E_{L,i}$ which is computational expensive if pseudo potentials are used. A possibly cheaper energy optimization method is the *stochastic reconfiguration method* [17] and its variants, the *energy fluctuation potential* method [18] (EFP) and the *perturbative optimization* method [10]. The stochastic reconfiguration method is based on the expansion of the DMC propagator to first order $\exp(-tH) \approx 1 - Ht$ applied to the trial wave function Ψ and projected into the space of derivative functions Ψ_i including $\Psi_0 \equiv \Psi$.

$$\langle \Psi_i | 1 - tH | \Psi \rangle = \sum_{j=0}^{n_p} \alpha_j \langle \Psi_i | \Psi_j \rangle \tag{41}$$

The terms of this linear equation can be sampled analogously to the linear method, and the resulting parameter changes are $\Delta p_i = \alpha_i / \alpha_0$. For the related EFP method we refer to Ref. [18] and references therein. In this paper, Scemama and Filippi also derived an efficient perturbative EFP method that we present here in the notation by Toulouse and Umrigar used so far who discuss the various optimization methods in some detail [10]. The formula for the perturbative method is

$$\Delta p_i = -\frac{1}{\Delta \varepsilon_i} \sum_{j=1}^{n_p} (\mathbf{S}^{-1})_{ij} H_{j0} \tag{42}$$

with

$$\Delta \varepsilon_i = \frac{H_{ii}}{S_{ii}} - H_{00}\,. \tag{43}$$

By comparison of Eq. (35) and (27), H_{j0} is identified as half the gradient $\mathbf{g}$ of the VMC energy. Hence, Eq. (42) has the form of a Newton step with the inverse overlap matrix as approximative Hessian. In practice, $\Delta \mathbf{p}$ is obtained from the corresponding linear equation. The main difference to the expressions obtained from Eq. (41) lies in the factor $1/\Delta\varepsilon_i$ replacing the scalar t in the stochastic reconfiguration method. This adds flexibility to the method and leads to improved convergence [18]. The local energy derivative $E_{L,i}$ is required for the denominator $\Delta\varepsilon_i$ but not for H_{j0}. If $E_{L,i}$ is expensive to calculate one may replace $\Delta\varepsilon_i$ by a constant $\Delta\varepsilon$ as is done in the stochastic reconfiguration method [17, 10]. A less drastic approximation that we have successfully applied is a more approximative calculation of $\Delta\varepsilon_i$ only in the first iteration step and with a small sample. The convergence of the three variants are demonstrated in Figure 1 for the orbital optimization in C_2 with a full valence CAS and Jastrow wave function and Burkatzki pseudo potentials to remove the core electrons [19, 20]. It can be seen that the calculation of all $\Delta\varepsilon_i$ only in the first step with a small sample does not lead to a diminished accuracy or slower convergence while the convergence is substantially slower if all $\Delta\varepsilon_i$ are replaced by a constant [21].

In Figure 2, a comparison of all three optimization techniques is shown for the MO optimization of C_2 in a full valence CAS wave function with a polarized triple zeta STO basis and a Jastrow factor. In this example, all three methods converge extremely fast and lead to a significant

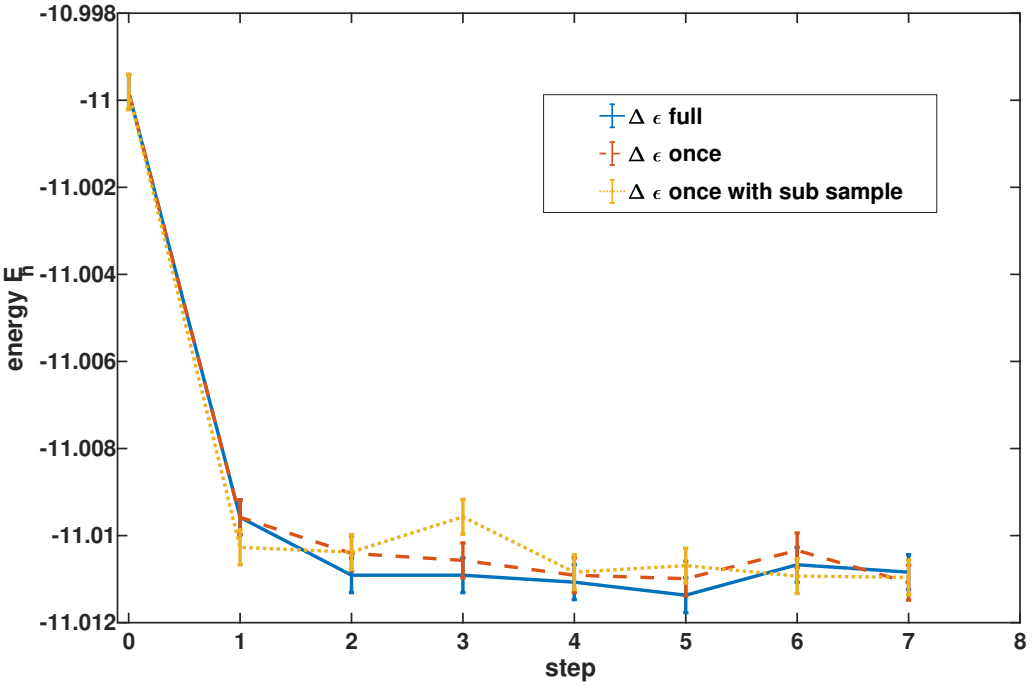

Fig. 1: *Orbital optimization in C_2 with full valence CAS and Jastrow using ECPs with the perturbative method with full $\Delta\varepsilon$ calculation, calculation of $\Delta\varepsilon$ only in the first iteration, and only in the first iteration with a small sample [21].*

drop in energy. While the linear method converges usually very quickly for all parameters the perturbative method does not work well for Jastrow parameters [10].

The stochastic reconfiguration method was computationally accelerated by Neuscamman et al. who observed that the overlap matrix $\langle\Psi_i|\Psi_j\rangle$ does not have to be built explicitly if stochastic estimators for matrix-vector products are used [22]. The replacement of explicit matrices by matrix-vector products is widely employed in numerical methods, often allowing to increase the dimension of systems substantially. The authors show that the calculation of matrix-vector products is only possible when the conjugate-gradient method is used to solve the linear equation Eq. (41). The accelerated stochastic reconfiguration method allowed the optimization of more than hundred thousand parameters in an massively parallel implementation.

3.3 Terms required for parameter optimization

The implementation of the energy and variance minimization requires parameter derivatives of the wave function and the local energy. We discuss here a few important aspects of these parameter derivatives with the restriction to the commonly used Slater-Jastrow wave functions

$$\Psi(\mathbf{c},\boldsymbol{\kappa},\mathbf{q}) = \Phi(\mathbf{c},\boldsymbol{\kappa})\, e^{U(\mathbf{q})} \tag{44}$$

with a Jastrow correlation function U with parameter vector $\mathbf{q}$ and, in general, a linear combination of Slater determinants Φ_d composed of orthogonal molecular orbitals (MO) depending

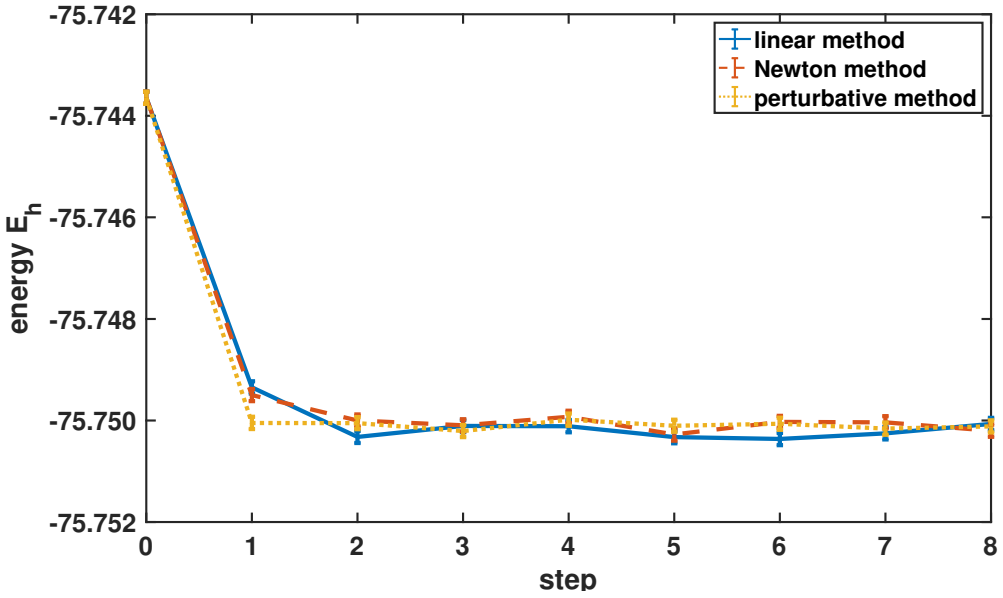

Fig. 2: *Comparison of the linear method with the Newton and the perturbative method for the MO optimization in the C_2 molecule [21].*

on parameters $\boldsymbol{\kappa}$

$$\Phi(\mathbf{c}, \boldsymbol{\kappa}) = \sum_{d=1}^{n_{\mathrm{det}}} c_d \, \Phi^{(d)}(\boldsymbol{\kappa}) \, . \tag{45}$$

In the following, we stick to the above notation and indicate derivatives with respect to parameter p_i with subscript i. In the given formulae, the parameter derivatives arise only in the ratio Ψ_i/Ψ which is then given here. For Jastrow parameters we obtain

$$\frac{\Psi_i}{\Psi} = \frac{1}{\Psi} \frac{\partial \Psi(\mathbf{c}, \boldsymbol{\kappa}, \mathbf{q})}{\partial q_i} = \frac{\partial U}{\partial q_i} =: U_i \tag{46}$$

and for the CI parameters c_i

$$\frac{\Psi_i}{\Psi} = \frac{1}{\Psi} \frac{\partial \Psi(\mathbf{c}, \boldsymbol{\kappa}, \mathbf{q})}{\partial c_i} = \frac{\Phi^{(i)}}{\Phi}. \tag{47}$$

The derivatives with respect to the MO parameters are somewhat more involved. We obtain

$$\frac{\Psi_i}{\Psi} = \frac{1}{\Psi} \frac{\partial \Psi(\mathbf{c}, \boldsymbol{\kappa}, \mathbf{q})}{\partial \kappa_i} = \frac{1}{\Phi} \sum_{d=1}^{n_{\mathrm{det}}} c_d \, \Phi_i^{(d)} \tag{48}$$

where the derivative of a Slater determinant $\Phi_i^{(d)} = \partial \Phi^{(d)} / \partial \kappa_i$ is required.
All Slater determinants are formed by a set of orthogonal MOs which form, as column vectors corresponding to a basis set, a matrix $\boldsymbol{\Phi}$. The discussion here is restricted to real orbitals. Each orbital rotation $\hat{x}_{ij}$ acts on this matrix by mixing MO i and j with the angle x_{ij}

$$\hat{x}_{ij} \boldsymbol{\Phi} = \cos x_{ij} \boldsymbol{\Phi} + \sin x_{ij} \left(\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i \right) \boldsymbol{\Phi}. \tag{49}$$

It is sufficient to calculate derivatives for the current set of orbitals corresponding to $x_{ij} = 0$. Taking the derivative with respect to the rotation angle at $x_{ij} = 0$ results in

$$\frac{\partial}{\partial x_{ij}} \hat{x}_{ij} \boldsymbol{\Phi} \Big|_{x_{ij}=0} = \left(\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i \right) \boldsymbol{\Phi}.$$

With the singlet excitation operator as defined in standard second quantized quantum chemistry [23]

$$\hat{E}_{ij}^- = \hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i$$

we obtain for all Slater determinants $\Phi^{(d)}$ built from this set of MOs

$$\frac{\partial}{\partial x_{ij}} \hat{x}_{ij} \Phi^{(d)} \Big|_{x_{ij}=0} = \hat{E}_{ij}^- \Phi^{(d)}. \tag{50}$$

Note that the operator $\hat{E}_{ij}^-$ has to be applied for each spin orbital. Hence, the derivative of a Slater determinant with respect to an orbital rotation angle (at zero angle) is obtained simply as the value of the corresponding singly excited determinant.
The orbitals are usually partitioned in the inactive, active, and virtual orbitals where inactive orbitals are occupied in all determinants, active orbitals are occupied in some of the determinants, and virtual orbitals are not occupied in any determinant. Orbital rotations between two inactive and between two virtual orbitals do not change the Slater determinant except for possibly a sign change. In complete active space (CAS) wave functions, active-active rotations are also invariant.
The MO parameters $\boldsymbol{\kappa}$ are therefore built from the set of non-invariant orbital rotations x_{ij} forming a real antisymmetric matrix $\mathbf{X}$. The general treatment of orbital rotations makes use of the fact that exponentials of antisymmetric (or skew symmetric) matrices are unitary. Multiplication of the current set of MOs $\boldsymbol{\Phi}$ by $\mathbf{U} = \exp(\mathbf{X})$ results in the new set of MOs. The unitary matrix $\mathbf{U}$ can be obtained with [23]

$$\mathbf{U} = \mathbf{W} \cos \boldsymbol{\tau} \mathbf{W}^T + \mathbf{W} \boldsymbol{\tau}^{-1} \sin \boldsymbol{\tau} \mathbf{W}^T \mathbf{X} \tag{51}$$

where $\mathbf{W}$ and $-\boldsymbol{\tau}^2$ are the eigenvectors and eigenvalues of the symmetric matrix $\mathbf{X}^2$, respectively. For more details on the MO optimization see Ref. [10] and references given therein.

3.3.1 Gradient of the local energy E_L

For variance minimization as well as for some of the energy minimization methods, the parameter derivative of the local energy E_L is required. We obtain straightforwardly for $H = T + V$

$$E_{L,i} := \frac{\partial E_L}{\partial p_i} = \frac{H\Psi_i}{\Psi} - E_L \frac{\Psi_i}{\Psi} = -\frac{1}{2} \frac{\nabla^2 \Psi_i}{\Psi} + \frac{1}{2} \frac{\nabla^2 \Psi}{\Psi} \frac{\Psi_i}{\Psi} + \frac{\partial V}{\partial p_i} \tag{52}$$

where the last term, the parameter derivative of the potential V, vanishes unless a localized pseudo potential is used. Note that the Laplacians $\nabla^2 \Psi$ and $\nabla^2 \Psi_i$ refer to second derivatives with respect to all electron positions, but not to parameter derivatives. The Laplacian of the

wave function derivative $\nabla^2\Psi_i$ is the new term in Eq. (52). Based on the Laplacian of the wave function $\nabla^2\Psi$ which is required already for the local energy evaluation, the parameter derivatives can be established for Jastrow, CI, and MO parameters similarly to the formulae given above.
More interesting is the common case of the presence of a nonlocal pseudo potential $\hat{V}_{\mathrm{nl}}$ in addition to a local potential V_{loc}. In this case we obtain for the local energy

$$E_L = -\frac{1}{2}\frac{\nabla^2\Psi}{\Psi} + V_{\mathrm{loc}} + V_{\mathrm{nl}}, \qquad V_{\mathrm{nl}} = \frac{\hat{V}_{\mathrm{nl}}\Psi}{\Psi} \tag{53}$$

with the *localized nonlocal pseudo potential* V_{nl}.
The parameter derivative becomes

$$E_{L,i} = -\frac{1}{2}\frac{\nabla^2\Psi_i}{\Psi} + \frac{1}{2}\frac{\nabla^2\Psi}{\Psi}\frac{\Psi_i}{\Psi} + \frac{\partial V_{\mathrm{nl}}}{\partial p_i} \tag{54}$$

and $E_{L,i}$ differs from the case without pseudo potential only by the last term

$$\frac{\partial V_{\mathrm{nl}}}{\partial p_i} = \frac{\hat{V}_{\mathrm{nl}}\Psi_i}{\Psi} - \frac{\hat{V}_{\mathrm{nl}}\Psi}{\Psi}\frac{\Psi_i}{\Psi} = \frac{\hat{V}_{\mathrm{nl}}\Psi_i}{\Psi} - V_{\mathrm{nl}}\frac{\Psi_i}{\Psi} \tag{55}$$

The localized pseudo potential is usually calculated numerically on a spherical Lebedev grid [24]. The parameter derivatives $\hat{V}_{\mathrm{nl}}\Psi_i/\Psi$ can be calculated analogously. In spite of efficient update formulae the numerical integration has to be done for all electrons within a cut off radius. This is a computationally expensive step, and more so the calculation of the parameter derivative vector. The Lebedev integration is particularly costly for the MO optimization which is based on singlet excitations $\hat{E}^-_{ij}\Phi^{(d)}$ according to Eq. (50) because a large number of determinants is constructed, all of which require an evaluation at the Lebedev grid points. This step is the reason why the energy minimization for MO parameters is computationally especially expensive with methods requiring $E_{L,i}$ and why the perturbative method is the most efficient choice in this case, provided that $\Delta\varepsilon_i$ in Eq. (43) is calculated only approximately with a small sample and possibly only once.

4 Examples of wave function optimizations

In this section, simultaneous and alternating parameter optimizations are discussed before some examples of trial wave function optimizations are presented. In particular, it will be shown that the VMC energy minimization can lead to substantial and systematic improvements of the nodal surface and thus to more accurate DMC energies. Single- as well as multi-reference approaches are employed. The Jastrow, CI, and MO parameters are both partially and fully optimized in order to assess the effect on the variational energy. Results are given for atomic species, main group, and transition metal dimers. Dissociation energies and spectroscopic constants are presented. The BFD pseudo potentials by Burkatzki, Filippi, and Dolg [19, 20] which are specifically designed for QMC are used with the corresponding triple zeta basis set in all calculations. The Jastrow factors in the examples below contain two-electron and two-electron-nucleus terms as described in [29].

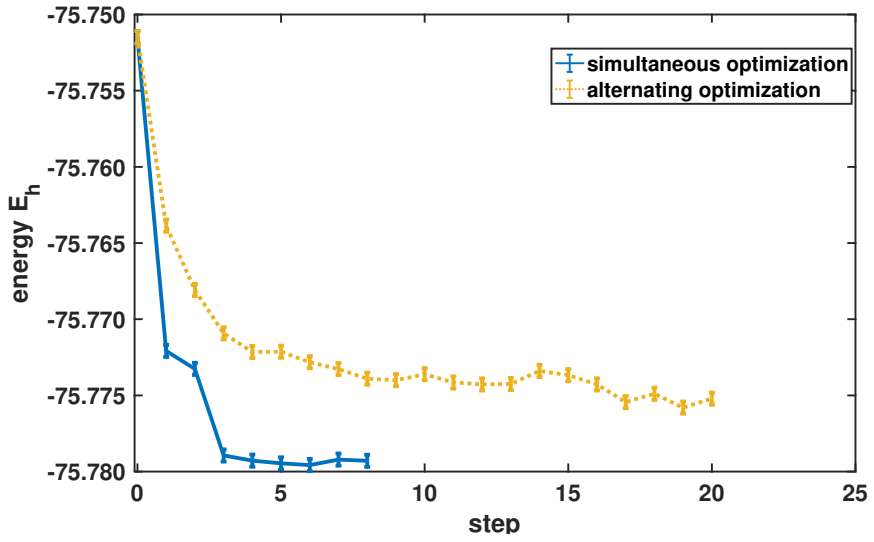

Fig. 3: *Comparison of simultaneous and alternating optimization of Jastrow, MO, and CI parameters in C_2 with an all electron full valence CAS wave function with Jastrow factor using the linear method [21].*

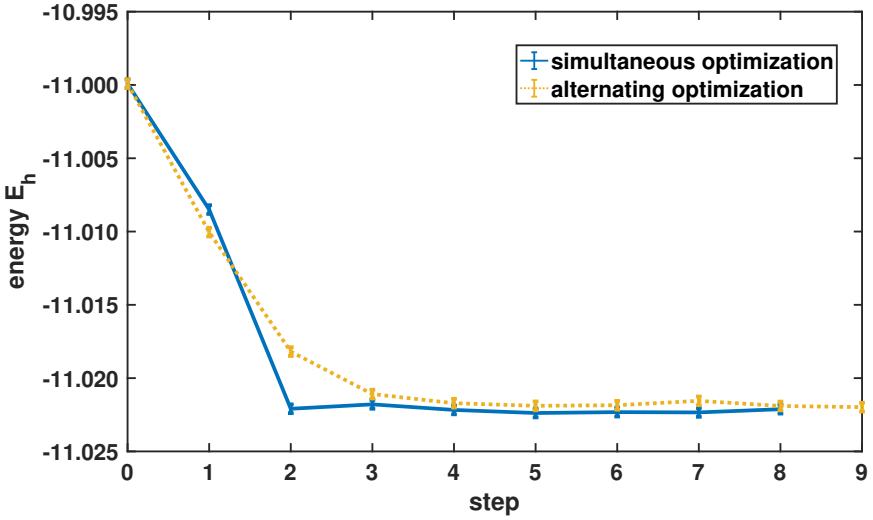

Fig. 4: *Comparison of simultaneous and alternating optimization of Jastrow, MO, and CI parameters in C_2 with pseudo potentials using the linear method [21].*

4.1 Simultaneous and alternating parameter optimizations

Not all optimization methods work equally well for all kinds of parameters. For instance, the linear method is expected to obtain the optimal CI parameters in one step as discussed above. On the other hand, the linear method is computationally more expensive than the perturbative method for the optimization of MO parameters when pseudo potentials are employed. It would therefore desirable to optimize different kinds of parameters alternatingly with possibly different methods. Alternating optimization instead of simultaneous optimization can lead to substantially slower convergence if the parameters are strongly coupled. As an example, we show in Figure 3 the comparison of simultaneous and groupwise alternating optimization of

Jastrow, MO, and CI parameters in C_2 with an all electron full valence CAS wave function with Jastrow factor using the linear method. The slow convergence of the alternating optimizations in comparison to the very fast convergence in the simultaneous case is very obvious. On the other hand, Figure 4 demonstrates that alternating optimization can be very efficient. In this example, the same C_2 molecule is treated, the only difference being the use of pseudo potentials. In general, it has been observed that slow convergence is found in alternating optimizations for all electron calculations while the alternating optimization is efficient when pseudo potentials are used [21].

4.2 Atomic species

4.2.1 Nickel spectrum

The energy gaps between different atomic states is a quantity that can be experimentally determined in a very accurate way, therefore the opportunity to verify the accuracy of the employed method presents itself. The ground state of Nickel, as well as two excited states, are evaluated. Ni is assigned a ground state of ^{3}F with an occupation of $3d^8 4s^2$. The first excited state is a ^{3}D state with an occupation of $3d^9 4s^1$ which is energetically close to the ground state. The third considered state is of ^{1}S symmetry with a $3d^{10}$ occupation.
The VMC and the zero time-step extrapolated DMC energies can be found in Table 1. The calculations were performed with HF and CAS orbitals. All parameters are optimized with respect to the energy using the linear method except for the MO parameters that are optimized with the perturbative method using a small sample only for $\Delta\varepsilon_i$.

Table 1: *Ni VMC and DMC energies in E_h for three states at various optimization levels, using different starting orbitals and BFD-VTZ/sm666.*

State	Wave Function Ansatz	Optimization level	VMC energy	DMC energy
^{3}F	HF	Jas	-170.0563(4)	-170.1146(5)
		Jas+MO	-170.0645(4)	-170.1185(5)
	[10,6]-CAS	Jas+CI	-170.0554(4)	-170.1151(5)
		Jas+MO+CI	-170.0636(4)	-170.1212(5)
^{3}D	HF	Jas	-170.0565(4)	-170.1152(5)
		Jas+MO	-170.0641(4)	-170.1189(5)
	[10,6]-CAS	Jas	-170.0558(4)	-170.1152(5)
		Jas+MO[a]	-170.0656(4)	-170.1199(5)
^{1}S	HF	Jas	-169.9964(4)	-170.0500(5)
		Jas+MO	-169.9972(3)	-170.0506(5)
	[10,6]-CAS	Jas+CI	-169.9991(4)	-170.0525(6)
		Jas+MO+CI	-170.0006(4)	-170.0532(5)

[a] Only one CSF is obtained.

Table 2: *DMC energy gaps in E_h between the ground state and two excited states of Ni at various optimization levels, using different starting orbitals and BFD-VTZ/sm666. The experimental values are taken from the NIST Atomic Spectra Database [25].*

Transition	Method	Optimization level	ΔE
^{3}F - ^{3}D	HF	Jas	-0.016(19)
		Jas+MO	-0.011(19)
	CAS	Jas	-0.003(19)
		Jas+MO	0.035(19)
	NIST		0.02539
^{3}F - ^{1}S	HF	Jas	1.759(21)
		Jas+MO	1.848(21)
	CAS	Jas+CI	1.703(21)
		Jas+MO+CI	1.850(21)
	NIST		1.82614

The MO optimization leads to an improvement of the energy, regardless of choice of initial orbitals. The lowering is however only significant for the ^{3}F and the ^{3}D state. Table 1 shows that the correct ground state of Ni cannot be reproduced at the VMC level. Even after MO optimization, the ^{3}D state exhibits a lower energy than ^{3}F. This may be rectified by choosing a more accurate Jastrow correlation function. For the Ni ^{3}F state, the HF trial wave function yields slightly lower VMC energies than the CAS one. This can be traced back to a loss of symmetry of the HF trial wave function compared to the CAS one. The higher flexibility of the former leads to a lowering of the energy. As for the DMC results, the MO optimization again leads to lower energies which can be explained by the improved nodal surface of the guide functions. The correct ground state can only be portrayed by the CAS guide function at the highest optimization level.

The excitation energies for the different transitions are shown in Table 2. The energy gap is severely underestimated by the HF guide function as well as by the CAS guide function without orbital optimization. The negative energy gaps mirror the fact that these approaches cannot portray the correct ground state for Nickel. One can thus deduce that the single determinant ansatz is not suitable to describe the states of the Ni atom and that the optimization of the orbital parameters is essential. The energy gap, derived from the NIST database, can however be reproduced by the CAS guide function where the orbitals are optimized in the presence of a Jastrow correlation function.

Table 3: *C_2 VMC and DMC energies in E_h at various optimization levels (Jas = Jastrow only), using different starting orbitals and BFD-VTZ/SM-9t. The data are taken from Ref. [21].*

Ansatz	Orbitals	Optimization level	VMC energy	DMC energy
[8,8]-CAS	CAS	Jas	-11.0680(2)	-11.0886(2)
	CAS	Jas+CI	-11.0779(3)	-11.0925(3)
	opt	Jas+MO+CI	-11.0792(2)	-11.0934(2)

Table 4: *C_2 MR-DMC dissociation energies in eV at various optimization levels, using different starting orbitals and BFD-VTZ/SM-9t. The data are taken from Ref. [21].*

Ansatz	Orbitals	Optimization level	D_0
CAS	CAS	Jas	6.351(9)
	CAS	Jas+CI	6.368(9)
	opt	Jas+MO+CI	6.378(7)
	exp.		6.30(2) [27]

4.3 Main group compounds

4.3.1 Carbon dimer C_2

The carbon dimer C_2 is the benchmark compound for static correlation at equilibrium bond distance, the small number of electrons making it easily feasible for multi-reference calculations. The VMC and time-step extrapolated DMC energies are given in Table 3. In contrast to the other calculations, the nine-term SM-9t Jastrow as suggested by Schmidt and Moskowitz [26] is employed for the C_2 calculations. The Jas+CI optimization improves both the VMC and DMC energies considerably. By further optimizing the molecular orbital parameters, a lowering of the energies can be observed, it is however less significant. The close DMC energies at the Jas+CI optimization level and for the fully optimized guide function indicate that the nodes must be similar.

Table 4 shows the computed MR-DMC dissociation energies at different optimization levels. The given dissociation energies D_0 are corrected for zero point energy, the core-valence correlation contribution, and spin-orbit contributions. All multi-reference dissociation energies presented here agree well with experiment, which is rather surprising since the absolute DMC energies differ significantly from one another.

4.4 Transition-metal compounds

4.4.1 Absolute energies

In this section, the VMC and DMC energies of ZnO and the effect that the parameter optimization has on them are evaluated. The VMC and the time-step extrapolated DMC energies are presented in Table 5. The non-optimized parameters are taken from the respective ab initio calculations.

Table 5: *ZnO VMC and DMC energies in E_h at various optimization levels (Jas = Jastrow only), using different starting orbitals and BFD-VTZ/sm666. The data are taken from ref. [28]*

Ansatz	Orbitals	Optimization level	VMC energy	DMC energy
Single det	HF	Jas	-242.8836(3)	-242.9931(5)
	B3LYP	Jas	-242.8944(3)	-243.0022(5)
	opt	Jas+MO	-242.9013(3)	-243.0065(6)
[16,9]-CAS	CAS	Jas	-242.8971(3)	-242.9950(5)
	CAS	Jas+CI	-242.9047(3)	-243.0023(6)
	opt	Jas+MO+CI	-242.9176(3)	-243.0111(5)

Table 5 shows a systematic lowering of the VMC energies when moving from HF over B3LYP KS to CAS orbitals. By only optimizing the Jastrow parameters, similar energies are obtained for the single- and multi-determinant trial wave functions. This shows how crucial the optimization of the antisymmetric part of the wave function is. The vast lowering in VMC energy due to the optimization of the molecular orbitals in the presence of a Jastrow correlation factor shows the substantial impact that the coupling between dynamic and static correlation has on the energies calculated with HF and CAS orbitals. For the B3LYP KS orbitals on the other hand, the MO optimization only has a small effect on the energy of the trial wave function because DFT is able to partly capture this coupling.

The MO optimization significantly improves the nodal surface of the CAS guide function. The improvement is less substantial for the KS guide function which indicates that the nodal surface was already almost optimal for this approach before the optimization. Only the fully optimized guide function is able to yield a DMC energy that is lower than the one calculated for the single-determinant approach with optimized orbitals. This emphasizes the impact that the dynamic correlation has on the quality of the nodal surface. If the molecular orbitals are not optimized, the Jastrow optimization does not change the nodal surface, due to its totally symmetric nature.

4.4.2 Dissociation energies

As example for accurate calculations of dissociation energies of small transition metal compounds we discuss the diatomics FeH, FeO, FeS, and ZnO. For all these molecules accurate experimental dissociation energies are known, while calculated dissociation energies even for accurate methods deviate substantially from the experimental data. We can demonstrate that the nodal surface obtained by VMC energy minimization of Jastrow, CI, and MO parameters in multireference Slater-Jastrow wave functions does lead to accurate dissociation energies with the DMC method. Only the VMC energy minimization for the molecule FeH is discussed here in detail as a typical example for an accurate transition metal wave function. The wave function is constructed from a [9,7]-CAS, built from the $3d$ and $4s$ orbitals on Fe and $1s$ on H, and the sm666 Jastrow factor from Ref. [29]. This results in 69 Jastrow parameters, 159 MO parameters, and 29 CI parameters. The energy minimization steps are shown in Figure 5 after

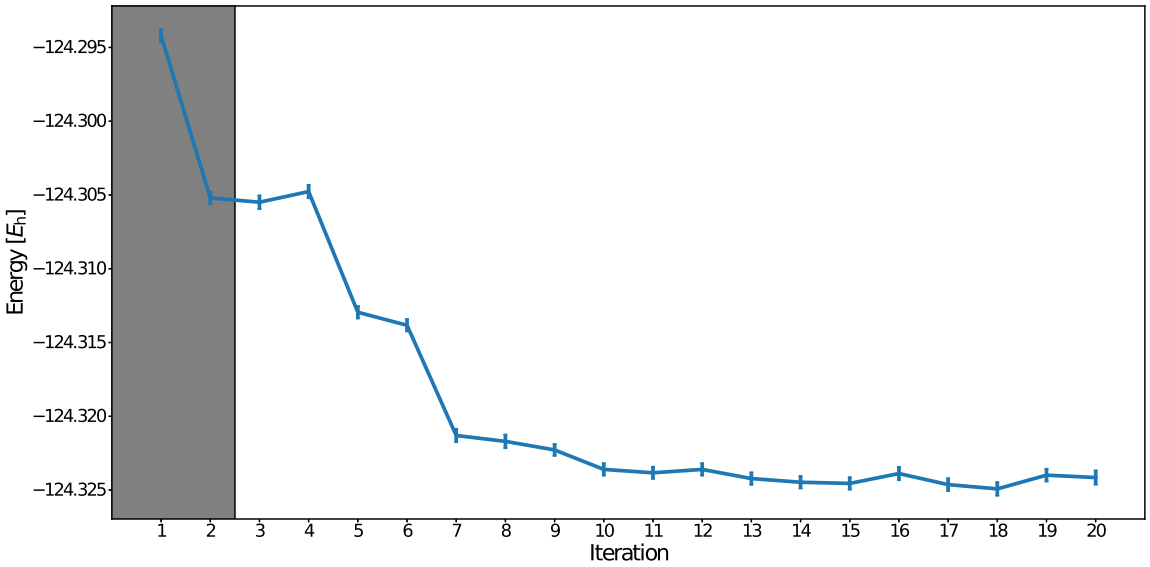

Fig. 5: *VMC energy minimization of the diatomic FeH using alternating optimization of Jastrow, CI, and MO parameters using the linear method for Jastrow and CI, and the perturbative method for the MO parameters. The first two steps shown in grey refer to preoptimization of the Jastrow and MO parameters.*

an initial variance optimization of the Jastrow parameters. The Jastrow and MO parameters are preoptimized with the linear and the perturbative method, respectively (shown in grey). Then the Jastrow, MO, and CI parameters are groupwise alternatingly optimized yielding a smooth and fast convergence.

The dissociation energies of several transition metal compounds at various optimization levels are illustrated in Table 6. The data for ZnO, FeH, and FeO are taken from Ref. [28], while the ones for FeS are presented in Ref. [21]. First of all, a systematic lowering of the deviation from the experimental dissociation energies can observed for all compounds for the different ansätze and optimization levels. The dissociation energies of ZnO and FeO follow similar trends. The single-determinant guide function with optimized orbitals yields more accurate dissociation energies than the CAS nodes at the Jas+CI optimization level. The optimization of the CAS orbitals in the presence of a Jastrow correlation function substantially improves the dissociation energy. For both compounds, an excellent agreement with the experiment is obtained.

For FeH, the single-determinant DMC nodes fail to reproduce the dissociation energy of Schultz and Armentrout [30]. The optimization of the KS orbitals does not change the dissociation energy which implies that the nodal surface was already almost optimal for the single-determinant approach before the optimization. Both the CI and MO optimization significantly improve the dissociation energy of the CAS guide function. The MR-DMC dissociation energy computed for the fully optimized guide function agrees well with the experimental one.

Moving towards a multi-reference guide function is also necessary for the FeS system. The single-determinant approach underestimates the experimental dissociation energy, while the fully optimized MR-DMC ansatz is able to reproduce it.

Table 6: *DMC dissociation energies in eV for several transition metal dimers at various optimization levels, using different starting orbitals and BFD-VTZ/sm666. The energies are corrected for core-valence correlation and spin-orbit effects.*

Compound	Ansatz	Orbitals	Optimization level	D_0
ZnO	Single det	HF	Jas	1.201(19)
		B3LYP	Jas	1.449(19)
		opt	Jas+MO	1.566(21)
	CAS	CAS	Jas	1.253(19)
		CAS	Jas+CI	1.451(21)
		opt	Jas+MO+CI	1.691(19)
	exp.			1.61(4) [31]
FeO	Single det	HF	Jas	2.885(20)
		B3LYP	Jas	3.688(20)
		opt	Jas+MO	3.826(20)
	CAS	CAS	Jas	3.266(20)
		CAS	Jas+CI	3.761(20)
		opt	Jas+MO+CI	4.112(20)
	exp.			4.18(1) [32, 33]
FeH	Single det	HF	Jas	0.814(17)
		B3LYP	Jas	1.020(17)
		opt	Jas+MO	1.020(17)
	CAS	CAS	Jas	1.099(17)
		CAS	Jas+CI	1.369(17)
		opt	Jas+MO+CI	1.791(17)
	exp.			1.63(8) [30]
FeS	Single det	opt	Jas+MO	2.914(15)
	CAS	opt	Jas+MO+CI	3.159(15)
	exp.			3.31(15) [34]

4.4.3 Spectroscopic constants

The evaluation of quantities, such as the equilibrium bond distance, the harmonic frequency, and the anharmonicity allows an assessment of the accuracy of the employed method. Those spectroscopic constants are presented in Table 7 for different compounds. The data for the oxides and for FeH are taken from Ref. [28], the ones of FeS from Ref. [35].

The potential energy curves were computed for the fully optimized MR-DMC guide functions at a fixed time-step and then fitted to a Morse function, from which the spectroscopic constants could be deduced.

Table 7: *Spectroscopic constants for the different transition metal compounds. The equilibrium bond distance is given in* Å*, the harmonic frequency and the anharmonicity in* cm^{-1}.

System	Investigators	Method	r_e	ω_e	$\omega_\mathrm{e}x_\mathrm{e}$
ZnO	This work	MR-DMC	1.709	746(8)	4.4(1)
	Zack et al. [37]	Direct-absorption methods	1.7047(2)	738	4.88
	Fancher et al. [38]	Photoelectron spectrum		805(40)	
FeO	This work	MR-DMC	1.623	866(79)	4.7(7)
	Allen et al.[a]		1.619		
	Drechsler et al. [39]	anion-ZEKE		882	4
FeH	This work	MR-DMC	1.567	1842(27)	38.9(9)
	Philips et al. [40]	Near IR spectrum		1826.86	31.96
	Dulick et al. [41]			1831.8(19)	34.9(9)
FeS	This work	MR-DMC	2.031(7)	499(11)	2.53(11)
	Takano et al. [36]	Microwave spectrum	2.017		
	Wang et al. [42]	Fluorescence spectroscopy		518(5)	1.7(2)

[a] derived from Allen et al. [43]

The MR-DMC bond distances of ZnO and FeO agree well with the experimental ones. The computed equilibrium bond distance of FeS is slightly larger than the one from Takano et al. [36]. The harmonic frequencies as well as the anharmonicities are in good agreement with the experiment for the four dimers. All in all, we showed that MR-DMC can yield promising results for the compounds evaluated here, provided that all sets of parameters of the trial wave functions are optimized. In particular, the difficult energy minimization of the MO parameters does lead to substantially improved accuracy of the VMC and DMC results.

Acknowledgements

Jil Ludovicy and Kaveh Haghighi Mood contributed to this lecture. Computational resources from JARA-HPC grant rwth0278 are gratefully acknowledged.

References

[1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953)

[2] W.K. Hastings, Biometrika **57**, 97 (1970)

[3] A. Lüchow, WIREs Comput. Mol. Sci. **1**, 388 (2011)

[4] B.M. Austin, D.Y. Zubarev, and W.A. Lester, Chem. Rev. **112**, 263 (2011)

[5] Z. Sun, S. Huang, R.N. Barnett, and W.A. Lester, J. Chem. Phys. **93**, 3326 (1990)

[6] C.J. Umrigar, K.G. Wilson, and J.W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988)

[7] J.E. Dennis, Jr., D.M. Gay, and R.E. Welsch, ACM Trans. Math. Softw. **7**, 369 (1981)

[8] P.R.C. Kent, R.J. Needs, and G. Rajagopal, Phys. Rev. B **59**, 12344 (1999)

[9] D. Bressanini, G. Morosi, and M. Mella, J. Chem. Phys. **116**, 5345 (2002)

[10] J. Toulouse and C.J. Umrigar, J. Chem. Phys. **126**, 084102 (2007)

[11] X. Lin, H. Zhang, and A.M. Rappe, J. Chem. Phys. **112**, 2650 (2000)

[12] C.J. Umrigar and C. Filippi, Phys. Rev. Lett. **94**, 150201 (2005)

[13] S. Sorella, Phys. Rev. B **71**, 241103 (2005)

[14] M.P. Nightingale and V. Melik-Alaverdian, Phys. Rev. Lett. **87**, 043401 (2001)

[15] J. Toulouse and C.J. Umrigar, J. Chem. Phys. **128**, 174101 (2008)

[16] L. Zhao and E. Neuscamman, J. Chem. Theory Comput. **13**, 2604 (2017)

[17] S. Sorella, Phys. Rev. B **64**, 024512 (2001)

[18] A. Scemama and C. Filippi, Phys. Rev. B **73**, 241101 (2006)

[19] M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **126**, 234105 (2007)

[20] M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **129**, 164115 (2008)

[21] K. Haghighi Mood: *Full Wave Function Optimization with Quantum Monte Carlo* (Ph.D. thesis, RWTH Aachen University, 2018)

[22] E. Neuscamman, C.J. Umrigar, and G.K.-L. Chan, Phys. Rev. B **85**, 045103 (2012)

[23] T. Helgaker, P. Jorgensen, and J. Olsen: *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2002)

[24] L. Mitáš, E.L. Shirley, and D.M. Ceperley, J. Chem. Phys. **95**, 3467 (1991)

[25] A. Kramida, Y. Ralchenko, J. Reader, and N.A. Team: *NIST Atomic Spectra Database (ver. 5.5)* (`http://physics.nist.gov/asd`, 2017)

[26] K.E. Schmidt and J.W. Moskowitz, J. Chem. Phys. **93**, 4172 (1990)

[27] R.S. Urdahl, Y. Bao, and W.M. Jackson, Chem. Phys. Lett. **178**, 425 (1991)

[28] J. Ludovicy, K. Haghighi Mood, and A. Lüchow, arXiv/1906.10419

[29] A. Lüchow, A. Sturm, C. Schulte, and K. Haghighi Mood, J. Chem. Phys. **142**, 084111 (2015)

[30] R.H. Schultz and P.B. Armentrout, J. Chem. Phys. **94**, 2262 (1991)

[31] D.E. Clemmer, N.F. Dalleska, and P.B. Armentrout, J. Chem. Phys. **95**, 7263 (1991)

[32] D.A. Chestakov, D.H. Parker, and A.V. Baklanov, J. Chem. Phys. **122**, 084302 (2005)

[33] M. Li, S.-R. Liu, and P.B. Armentrout, J. Chem. Phys. **131**, 144310 (2009)

[34] J. Drowart, A. Pattoret, and S. Smoes, in *Proc. Brit. Ceram. Soc.* (1967), Vol. 8, pp. 67–89

[35] K. Haghighi Mood and A. Lüchow, J. Phys. Chem. A **121**, 6165 (2017)

[36] S. Takano, S. Yamamoto, and S. Saito, J. Mol. Spectrosc. **224**, 137 (2004)

[37] L.N. Zack, R.L. Pulliam, and L.M. Ziurys, J. Mol. Spectrosc. **256**, 186 (2009)

[38] C.A. Fancher, H.L. de Clercq, O.C. Thomas, D.W. Robinson, and K.H. Bowen, J. Chem. Phys. **109**, 8426 (1998)

[39] G. Drechsler, U. Boesl, C. Bäßmann, and E.W. Schlag, J. Chem. Phys. **107**, 2284 (1997)

[40] J.G. Phillips, S.P. Davis, B. Lindgren, and W.J. Balfour, Astrophys. J. Suppl. Ser. **65**, 721 (1987)

[41] M. Dulick, C.W. Bauschlicher, Jr., A. Burrows, C.M. Sharp, R.S. Ram, and P. Bernath, Astrophys. J. **594**, 651 (2003)

[42] L. Wang, D.-l. Huang, J.-f. Zhen, Q. Zhang, and Y. Chen, Chin. J. Chem. Phys. **24**, 1 (2011)

[43] M. Allen, L. Ziurys, and J. Brown, Chem. Phys. Lett. **257**, 130 (1996)