A Variational Monte Carlo Analysis of a Bose-Einstein Condensate

Dahl, Jon Kristian,[1] Hardersen, Alida,[1] and Sønderland, Per-Dimitri[1]

[1] *University of Oslo*

(Dated: April 4, 2021)

The ground state energy of a Bose-Einstein condensate is calculated using Variational Monte Carlo methods. We focus first on a system of non-interacting particles in a spherical potential trap before proceeding to a full system of interacting particles in an elliptical potential. Both systems are calculated with the Gross-Pitaevskii equation in the Hartree-Fock approximation. Two applications of the Metropolis-Hastings sampling method are used. One is the so-called brute-force Metropolis-Hastings method, while the other method is importance sampling based on the Langevin and Fokker-Planck equations. We find that while the brute-force method is faster for 2^{20} Monte Carlo cycles, the added importance sampling yields more precise results when performing a statistical analysis of the data using the blocking technique. In the case of non-interacting particles we find that the variational parameter α corresponding to the ground state energy is $\alpha = 0.5$. We are unsuccessful in our attempt to evaluate the ground state energy in the interacting system. An unknown error causes the ground state energy minimum to not coincide with the minimum ground state energy variance where we expect the true energy minimum to be located. The calculations of the onebody density for interacting system does show expected behaviour where an increase in the hard-core diameter of the bosons results in a larger spread of the particles.

I. INTRODUCTION

The Bose-Einstein condensate (BEC) is a state of matter first theorized in the early 1900's by Satyendra Nath Bose and Albert Einstein, and was experimentally verified in 1995 - 1998 by Cornell and Wiemann [3]. In a BEC, matter stops behaving like individual particles and instead behaves as a single quantum state described by a single wave function. A BEC is formed when a low density gas of bosons is cooled to temperatures close to absolute zero. A particularity of this state of matter is that certain quantum phenomena start to manifest on a macroscopic scale, making the study of BECs very attractive and indeed a very active field in physics today. Examples of such quantum phenomena are superfluidity and superconductivity [7].

Similar to all other fields of study in physics, numerical simulations are of vital importance in the study of BECs due to the practicality of computer simulations with regards to time, cost, and ease of use. When it comes to many-body systems, which BECs are examples of, Monte Carlo simulations are often the go-to method due to the simplicity of the implementation, and since Monte Carlo calculations makes possible simulations which are computationally too heavy to perform with any other tool.

In this project we use variational Monte Carlo (VMC) to calculate the ground state energy of a Bose-Einstein condensate. We use the Gross-Pitaevskii equation which describes the ground state of a quantum mechanical system comprised of identical bosons. In the VMC calculation we employ the Metropolis-Hastings algorithm to sample particle positions, aided by a technique called importance sampling for improving said sampling. We study the BEC with varying degrees of particle interaction by adjusting the interaction parameter a (the *hard-core diameter* of the bosons) in the interaction term of the wave function. We implement a gradient descent algorithm to calculate the variational parameter, α, on which the ground state energy of the system is minimized, and we apply statistical analysis on the calculated data with the blocking method.

We begin this report with a theory section where we give the necessary background information on the variational principle, Monte Carlo and the setup of the quantum mechanical system. Following is the methods section where we introduce the different algorithms we use in our calculations. In the results and discussion section we present our data, discuss its implications, and compare it to the underlying theory. We summarize our findings in the conclusions section, and end the report with an appendix. The analytical calculations of the equations in the theory section are located in the appendix, along with additional figures which are more of a supplement to the results rather than results themselves.

All source code is publicly available at the GitHub repository: https://github.com/JonKDahl/FYS4411_V21/tree/main/project1.

II. THEORY

A. The variational principle

The variational principle gives us a method of approximating the ground state energy E_{gs} of a system with a Hamiltonian $\hat{H}$ when we are not able to solve the Schrödinger equation. For any normalized trial wave function ψ_T we have that

$$E_{\text{gs}} \leq \langle \psi_T | \hat{H} \mid \psi_T \rangle, \tag{1}$$

where T denotes 'trial'. The expectation value of the Hamiltonian in the state ψ_T will almost always overestimate the ground state energy of the true state of the

system. To show that this is true, we start with the eigenfunctions $\{\psi_1, \psi_2, ..., \psi_n\}$ of $\hat{H}$. Since $\hat{H}$ is self adjoint, its eigenvectors span its Hilbert space (see spectral theorem [10]). We can therefore express the trial wave function as a linear combination of the eigenfunctions of $\hat{H}$,

$$\psi_T = \sum_{i=1}^{n} c_i \psi_i. \tag{2}$$

Assuming that ψ_T is normalized,

$$\begin{aligned} 1 &= \langle \psi_T | \psi_T \rangle \\ &= \left\langle \sum_{i=1}^{n} c_i \psi_i \middle| \sum_{j=1}^{n} c_j \psi_j \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i^* c_j \langle \psi_i | \psi_j \rangle \\ &= \sum_{i=1}^{n} |c_i|^2. \end{aligned} \tag{3}$$

Assuming that the eigenfunctions are orthonormalized,

$$\begin{aligned} \langle \psi_T \mid \hat{H} \mid \psi_T \rangle &= \left\langle \sum_{i=1}^{n} c_i \psi_i \middle| \hat{H} \sum_{j=1}^{n} c_j \psi_j \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i^* E_j c_j \langle \psi_i | \psi_j \rangle \\ &= \sum_{i=1}^{n} E_i |c_i|^2. \end{aligned} \tag{4}$$

Now, the ground state energy is per definition the smallest of the energy eigenvalues, therefore

$$\sum_{i=1}^{n} E_{\text{gs}} |c_i|^2 = E_{\text{gs}} \leq \sum_{i=1}^{n} E_i |c_i|^2 = \langle \psi_T | \hat{H} | \psi_T \rangle. \tag{5}$$

See [6] for details.

According to equation (5) we now have an upper bound for the ground state energy. By minimizing the r.h.s. of equation (5) we can approximate E_{gs}. To perform the minimization, we have to introduce a *variational parameter* α on which we minimize the trial wave function:

$$\frac{d}{d\alpha} \Big(\langle \psi_T | \hat{H} | \psi_T \rangle \Big) = 0. \tag{6}$$

In theory, any choice of wave function can be used for this purpose, but to converge on a good result as fast as possible we should choose a good trial wave function. According to [6] a good choice of wave function is the natural exponential function, for example

$$\psi_T(x; \alpha) = A e^{-\alpha x^2}, \tag{7}$$

where we have also introduced the variational parameter, α.

B. Monte Carlo

We want to calculate the energy expectation value that is associated with the hamiltonian $\hat{H}$ for every given trial wave function. This energy expectation value can be written as,

$$E = \langle H \rangle = \frac{\langle \psi_T | \hat{H} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \frac{\int d\boldsymbol{r}\ \psi_T^*(\boldsymbol{r}) H \psi_T(\boldsymbol{r})}{\int d\boldsymbol{r}\ |\psi_T|^2} \tag{8}$$

Instead of finding an analytical solution to the integral we will use Monte Carlo (MC) integration where we approximate the integral of the expectation value with a sum. One of the advantages of MC integration is that it does not impose any limitations on the form of the chosen wave function as long as the original expression is an expectation value, i.e. it converges, and that you *can* sample from the probability distribution, i.e. the wave function, in the integrand [2]. For a chosen trial wave function ψ_T the normalized probability density function (pdf) $p(\boldsymbol{r})$ is expressed as,

$$p(\boldsymbol{r}) \equiv \frac{|\psi_T(\boldsymbol{r})|^2}{\int d\boldsymbol{r}\ |\psi_T(\boldsymbol{r})|^2}. \tag{9}$$

We can also define the local energy E_L,

$$E_L \equiv \frac{1}{\psi_T} H \psi_T, \tag{10}$$

which allows us to rewrite the variational energy (8) to be on the form

$$E = \int d\boldsymbol{r}\ E_L(\boldsymbol{r}) p(\boldsymbol{r}) \approx \frac{1}{N} \sum_{i=1}^{N} E_L(\boldsymbol{r}_i). \tag{11}$$

Thus, the variational energy can be estimated as the average value of the local energy over a sample of N points (the number of MC cycles) $\boldsymbol{r}_i$ which are sampled from the probability density $p(\boldsymbol{r}_i)$. To see if we have reached a minimum, we calculate the variance of the expectation of the local energy,

$$\sigma_{E_L}^2 = \langle E_L^2 \rangle - \langle E_L \rangle^2, \tag{12}$$

where $\sigma_{E_L}^2 = 0$ means we have found a true minimum.

C. Bose-Einstein condensate

Bose-Einstein condensate (BEC) is a state of matter where the matter stops behaving like individual particles and instead behaves like a single quantum state described by a single wave function. A BEC is formed when a low density gas of bosons is cooled to temperatures close to absolute zero. In such a substance, the total system wave function can be approximated by a product of individual particle wave functions. This approximation is called the Hartree-Fock approximation.

D. Quantum mechanical setup

The quantum system at hand is a hard-sphere boson gas contained by a magnetic trap modelled by a harmonic oscillator potential, $V_{ext}(\boldsymbol{r})$. This potential can either be spherical (S) or elliptical (E) and is given by,

$$V_{ext} = \begin{cases} \frac{1}{2}m\omega_{ho}^2 r^2 & (S) \\ \frac{1}{2}m[\omega_{ho}^2(x^2+y^2)+\omega_z^2 z^2] & (E) \end{cases} \tag{13}$$

where ω_{ho} is the potential strength and ω_z is the potential strength in the z direction for the elliptical trap. The system is also influenced by an internal potential V_{int} given by,

$$V_{int}(|\boldsymbol{r_i}-\boldsymbol{r_j}|) = \begin{cases} \infty & \text{if} \quad |\boldsymbol{r_i}-\boldsymbol{r_j}| \leq a \\ 0 & \text{if} \quad |\boldsymbol{r_i}-\boldsymbol{r_j}| > a \end{cases} \tag{14}$$

where a is the hard-core diameter of the bosons. If the bosons are separated by a distance $|\boldsymbol{r_i}-\boldsymbol{r_j}| > a$ the internal potential is zero, while it is infinity if they attempt to come closer. The full two-body Hamiltonian is given by,

$$\hat{H} = \sum_{i=1}^{N}\left(\frac{-\hbar^2}{2m}\nabla_i^2 + V_{ext}(\boldsymbol{r}_i)\right) + \sum_{i<j}^{N} V_{int}(|\boldsymbol{r}_i-\boldsymbol{r}_j|). \tag{15}$$

The trial wave function we use to describe the ground state for N particles is given by

$$\psi_T(\boldsymbol{r}) = \prod_{i=1}^{N} g(\alpha,\beta,\boldsymbol{r}_i)\prod_{j=1}^{N}\prod_{k=j+1}^{N} f(a,|\boldsymbol{r}_j-\boldsymbol{r}_k|) \tag{16}$$

where $g(\alpha,\beta,\boldsymbol{r}_i)$ is the single-particle wave function and $f(a,|\boldsymbol{r_i}-\boldsymbol{r_j}|)$ is the pair Jastrow function [4]. The trial wave function represents the wave function for the total system of particles and is a product of single particle wave functions by the Hartree-Fock approximation. The parameters α and β are the variational parameters. The single particle wave function is proportional to the harmonic oscillator function for the ground state,

$$g(\alpha,\beta,\boldsymbol{r}_i) = e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} \equiv \phi(\boldsymbol{r}_i) \tag{17}$$

and the Jastrow factor is,

$$f(a,|\boldsymbol{r_i}-\boldsymbol{r_j}|) = \begin{cases} 0 & \text{if} \quad |\boldsymbol{r_i}-\boldsymbol{r_j}| \leq a \\ 1-\frac{a}{|\boldsymbol{r_i}-\boldsymbol{r_j}|} & \text{if} \quad |\boldsymbol{r_i}-\boldsymbol{r_j}| > a \end{cases} \tag{18}$$

We can rewrite the trial wavefunction by defining $r_{ij} = |\boldsymbol{r_i}-\boldsymbol{r_j}|$ and rewriting the Jastrow factor as $f(a,r_{ij}) = \exp u(r_{ij})$ with $u(r_{ij}) = \ln f(a,r_{ij})$, giving the rewritten trial wave function

$$\psi_T(\boldsymbol{r}) = \prod_{i=1}^{N}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j=1}^{N}\sum_{k=j+1}^{N} u(r_{jk})\right). \tag{19}$$

Our implementation of importance sampling also requires knowledge of the quantum force $\mathbf{F}(\boldsymbol{r})$ of the system, which is given by

$$\mathbf{F}(\boldsymbol{r}) = \frac{2\nabla\psi_T(\boldsymbol{r})}{\psi_T(\boldsymbol{r})}. \tag{20}$$

1. *Non-interacting particles*

For non-interacting particles we can set $a = 0$ making the Jastrow factor equal to one, meaning there is no correlation between the particles. The trial wave function is then,

$$\psi_T(\boldsymbol{r}) = \prod_{i=1}^{N}\phi(\boldsymbol{r}_i) = \prod_{i=1}^{N} e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} \tag{21}$$

Because there is no interaction, the internal potential V_{int} given in (14) is zero and the interaction term of the Hamiltonian disappears. We can then find an expression for the local energy (10)(see Appendix B),

$$E_L = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{m}\alpha\left[(x_k^2+y_k^2+\beta^2 z_k^2)-2-\beta\right] + V_{ext}(\boldsymbol{r}_k)\right) \tag{22}$$

The parameter β determines the elliptical shape of the distribution, for a spherical distribution $\beta = 1$ and $V_{ext}(r_k) = \frac{1}{2}m\omega_{ho}^2 r_k^2$. The local energy for N particles in D dimentions can then be written as,

$$E_L = \frac{\hbar^2\alpha N D}{m} - \left(\frac{2\hbar^2\alpha^2}{m} - \frac{1}{2}m\omega_{ho}^2\right)\sum_{k=1}^{N}\boldsymbol{r}_k^2 \tag{23}$$

where $\boldsymbol{r}_k^2 = x_k^2$, $x_k^2+y_k^2$ or $x_k^2+y_k^2+z_k^2$ for $D = 1$, 2 or 3 dimensions respectively. Furthermore, for non-interactig bosons, $\alpha = 1/2a_{ho}^2$ where $a_{ho} \equiv (\hbar/m\omega_{ho})^{1/2}$ is the characteristic length of the trap. Inserting this into (23) yields the simple expression,

$$E_L = \frac{1}{2}\omega_{ho}ND = \frac{1}{2}ND \tag{24}$$

where N is the number of particles and D is the number of dimensions. In the last term we have used natural units ($\hbar = m = \omega = c = 1$). An analytical expression for the quantum force of particle k as defined in (20) is,

$$\mathbf{F} = -4\alpha\boldsymbol{r}_k. \tag{25}$$

2. *Interacting particles*

For the system of interacting hard-sphere bosons we will only consider the elliptical trap with a repulsive potential given in eq. (13). The trap frequency in the perpendicular xy-plane is $\omega_\perp = \omega_{ho}$ while in the z-direction

it is ω_z. Using length in units of a_{ho} and energy in units of $\hbar\omega_{ho}$ (see Appendix D) the external potential can be written as,

$$V_{ext} = \frac{1}{2}(x^2 + y^2 + \gamma^2 z^2) \quad \text{where} \quad \gamma = \frac{\omega_z}{\omega_{ho}}. \tag{26}$$

The full Hamiltonian of the system in units of $\hbar\omega_{ho}$ is then,

$$H = \sum_{i=1}^{N} \frac{1}{2}\left(-\nabla_i^2 + x_i^2 + y_i^2 + \gamma^2 z_i^2\right) + \sum_{i<j}^{N} V_{int}(|\boldsymbol{r}_i - \boldsymbol{r}_j|) \tag{27}$$

and the local energy (10) is calculated using the expression for the second derivative (see appendix A 2 and B 2),

$$\begin{aligned}\frac{\nabla_i^2\psi_T}{\psi_T} =& 2\alpha(2\alpha\boldsymbol{r}_i^2 - 2 - \beta) \\ &- 4\alpha\boldsymbol{r}_i \sum_{l\neq i} \frac{a(\boldsymbol{r}_i - \boldsymbol{r}_l)}{r_{il}^2(r_{il} - a)} \\ &+ \sum_{l\neq i}\sum_{j\neq i} \frac{a^2(\boldsymbol{r}_i - \boldsymbol{r}_l)(\boldsymbol{r}_i - \boldsymbol{r}_j)}{r_{il}^2 r_{ij}^2 (r_{il} - a)(r_{ij} - a)} \\ &+ \sum_{l\neq i} \frac{a}{r_{il}^2(r_{il} - a)}\left[2 + \frac{a - 2r_{il}}{r_{il} - a}\right]\end{aligned} \tag{28}$$

3. *One-body density*

The one-body density for a normalized wave function is defined as the integral over all coordinates except one,

$$\rho(\boldsymbol{r}_i) \equiv \int |\psi(\boldsymbol{r})|^2 d\boldsymbol{r}_1 d\boldsymbol{r}_2 ... d\boldsymbol{r}_{i-1} d\boldsymbol{r}_{i+1} d\boldsymbol{r}_N \tag{29}$$

for a system of N particles. This represents the likelihood of finding particle i in the volume element $d\boldsymbol{r}_i$. This integral can be approximated using Monte Carlo integration. We count the number of particles that are in a sphere with radius $d\boldsymbol{r}_i$

III. METHODS

A. The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm uses a stochastic process called a Markov chain where the probability for transitioning to a new point $\boldsymbol{r}_i$ is only dependent on the previous point $\boldsymbol{r}_{i-1}$. The probability $P(\boldsymbol{r}_j|\boldsymbol{r}_i)$ of transitioning between point $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ is called the transition probability and is given by

$$P(\boldsymbol{r}_i|\boldsymbol{r}_{i-1}, ..., \boldsymbol{r_1}) = P(\boldsymbol{r}_i|\boldsymbol{r}_{i-1}). \tag{30}$$

The Markov chain is realized using a random walk. Starting from an initial point $\boldsymbol{r_1}$ we can sample the second point $\boldsymbol{r_2}$ from the transition probability $P(\boldsymbol{r}_2|\boldsymbol{r}_1)$. Then a third point is sampled from the transition probability $P(\boldsymbol{r}_3|\boldsymbol{r}_2)$ and so on. We can sample the target probability distribution $p(\boldsymbol{r})$ by constructing a Markov chain that converges to $p(\boldsymbol{r})$. A necessary condition is that $p(\boldsymbol{r})$ is an eigenvector of $P(\boldsymbol{r}_j|\boldsymbol{r}_i)$ with eigenvalue 1,

$$\int d\boldsymbol{r}_i P(\boldsymbol{r}_j|\boldsymbol{r}_i)p(\boldsymbol{r}_i) = p(\boldsymbol{r}_j) = \int d\boldsymbol{r}_j P(\boldsymbol{r}_i|\boldsymbol{r}_j)p(\boldsymbol{r}_j)$$

this is called the stationary condition[15]. Another condition is the detailed balance requirement given by

$$P(\boldsymbol{r}_j|\boldsymbol{r}_i)p(\boldsymbol{r}_i) = P(\boldsymbol{r}_i|\boldsymbol{r}_j)p(\boldsymbol{r}_j). \tag{31}$$

which is the requirement used to determine if a proposed new point is accepted in the Metropolis algorithm.

Starting from the initial point $\boldsymbol{r}_i$ we can propose a new point $\boldsymbol{r}'_j$ with a probability $P_{\text{prop}}(\boldsymbol{r}'_j|\boldsymbol{r}_i) = T(\boldsymbol{r}|\boldsymbol{r}_i)$. This new point is then either accepted or rejected:

- Accepted: $\boldsymbol{r}_j = \boldsymbol{r}'_j$ with $P_{\text{acc}}(\boldsymbol{r}'_j|\boldsymbol{r}_i) = A(\boldsymbol{r}_j|\boldsymbol{r}_i)$.
- Rejected: $\boldsymbol{r}_j = \boldsymbol{r}_i$ with $P_{\text{rej}}(\boldsymbol{r}'_j|\boldsymbol{r}_i) = 1 - A(\boldsymbol{r}_j|\boldsymbol{r}_i)$.

The corresponding transition probability can then be written as,

$$P(\boldsymbol{r}_j|\boldsymbol{r}_i) = \begin{cases} A(\boldsymbol{r}_j|\boldsymbol{r}_i)T(\boldsymbol{r}_j|\boldsymbol{r}_i) & \text{for } \boldsymbol{r}_i \neq \boldsymbol{r}_j \\ 1 - \int d\boldsymbol{r}'_j A(\boldsymbol{r}'_j|\boldsymbol{r}_i)T(\boldsymbol{r}'_j|\boldsymbol{r}_i) & \text{for } \boldsymbol{r}_i = \boldsymbol{r}_j \end{cases}$$

We then choose the acceptance probability in such a way that it satisfies the detailed balance condition in (31). For $\boldsymbol{r}_i \neq \boldsymbol{r}_j$ we have,

$$q(\boldsymbol{r}_j, \boldsymbol{r}_i) \equiv \frac{A(\boldsymbol{r}_j|\boldsymbol{r}_i)}{A(\boldsymbol{r}_i|\boldsymbol{r}_j)} = \frac{T(\boldsymbol{r}_i|\boldsymbol{r}_j)p(\boldsymbol{r}_j)}{T(\boldsymbol{r}_j|\boldsymbol{r}_i)p(\boldsymbol{r}_i)} \tag{32}$$

The metropolis choice [11] is then to maximize the acceptance probabilities,

$$A(\boldsymbol{r}_j|\boldsymbol{r}_i) = \min\left[1, q(\boldsymbol{r}_j, \boldsymbol{r}_i)\right]. \tag{33}$$

The proposed new state is automatically accepted if the ratio $q > 1$, but if $q \leq 1$ we compare it to a random number $x \in [0, 1]$. If $q > x$ the move is accepted. We allow the Markov chain to move in the wrong direction in parameter space, according to a given probability criterion. The advantage of this is that it reduces the chance of getting stuck in a local minimum. It allows the random walker to "move up" a bit in order to find a better way "down".

1. *Brute-Force Metropolis-Hastings Sampling*

In brute-force Metropolis-Hastings sampling, we make the assumption that $T(\boldsymbol{r}_j|\boldsymbol{r}_i) = T(\boldsymbol{r}_i|\boldsymbol{r}_j)$. The proposal

probabilities in the ratio in (32) will therefore cancel, leaving the acceptance criteria to only be dependent on the probability densities $p(\boldsymbol{r}_i)$ and $p(\boldsymbol{r}_j)$. Using the expression in (9) we have,

$$q(\boldsymbol{r}_j, \boldsymbol{r}_i) = \frac{p(\boldsymbol{r}_j)}{p(\boldsymbol{r}_i)} = \frac{|\psi_T(\boldsymbol{r}_j)|^2}{|\psi_T(\boldsymbol{r}_i)|^2} \tag{34}$$

since the normalization's cancel. If we count the number of accepted transitions and divide by the number of MC cycles we get the acceptance rate which is useful when we want to determine the optimal step length. The proposed position $\boldsymbol{r}_j$ is chosen as

$$\boldsymbol{r}_j = \boldsymbol{r}_i + \eta \Delta r \tag{35}$$

where η is a random number $\eta \in [0, 1]$ and Δr is the step size.

B. Importance Sampling

While effective in many cases, the brute-force Metropolis-Hastings algorithm suffer some downsides. It only accepts approximately half of the proposed moves, meaning that when a move is rejected we will end up sampling the same state several times which is an inefficient use of calculation time and resources. We can make an improvement to our algorithm by introducing importance sampling where we include the influence of the quantum force (20) on the system. The gradient of the wave function will indicate in which direction the particles are moving, so introducing this force will help us to sample more relevant points, instead of sampling random points.

The method is based on a combination of the Fokker-Planck equation (36) and the Langevin equation (39). The Fokker-Planck equation describes the time evolution of the probability density of a particle that is under the influence of drift forces. Consider the expression

$$\frac{\partial P(\boldsymbol{r}, t)}{\partial t} = D \nabla \cdot (\nabla - \boldsymbol{F}) P(\boldsymbol{r}, t) \tag{36}$$

where D represents the diffusion coefficient, $P(\boldsymbol{r}, t)$ is the time-dependent probability density and $\boldsymbol{F}$ is the drift term given by the quantum force (20). This will push the particle towards the regions where the trial wave function is large. Using the Fokker-Planck equation we get a new expression for the probability $T(\boldsymbol{r}_j|\boldsymbol{r}_i)$ in (32), now given by Greens function,

$$G(\boldsymbol{r}_j, \boldsymbol{r}_i, \Delta t) = \frac{1}{(4\pi D \Delta t)^{3N/2}} \times \exp\left[\frac{-(\boldsymbol{r}_j - \boldsymbol{r}_i - D\Delta t \boldsymbol{F}(\boldsymbol{r}_i))^2}{4D\Delta t}\right]. \tag{37}$$

Inserting this probability in the ratio for the acceptance criteria of the Metropolis algorithm (32) and using the expression for $p(\boldsymbol{r})$ in (9) we get the new criteria

$$q(\boldsymbol{r}_j, \boldsymbol{r}_i) = \frac{G(\boldsymbol{r}_i, \boldsymbol{r}_j, \Delta t)|\psi_t(\boldsymbol{r}_j)|^2}{G(\boldsymbol{r}_j, \boldsymbol{r}_i, \Delta t)|\psi_t(\boldsymbol{r}_i)|^2}. \tag{38}$$

The Langevin equation is a stocastic differential equation which describes the time evolution of a subset of the degrees of freedom in a system. We can use the Langevin equation to find the new particle positions in coordinate space,

$$\frac{\partial \boldsymbol{r}(t)}{\partial t} = D\boldsymbol{F}(\boldsymbol{r}(t)) + \eta. \tag{39}$$

We solve the above Langevin equation using Eulers method to find the new proposed positions r',

$$\boldsymbol{r}_j = \boldsymbol{r}_i + D\boldsymbol{F}(\boldsymbol{r}_i)\Delta t + \eta\sqrt{\Delta t}, \tag{40}$$

where η is a stochastic variable drawn from a normal distribution and Δt is a chosen time step.

C. Gradient Descent

In the search of an optimal variational parameter, α, we employ a gradient descent algorithm. While importance sampling makes sure that we sample the wave function at positions of greater probability than that of the random sampling by the brute-force metropolis algorithm, gradient descent makes sure that we sample variational parameters closer and closer to the variational parameter at which the local energy minimum is located. An initial variational parameter is chosen, and all subsequent variational parameters are decided by the gradient descent algorithm.

As the name implies, a derivative is calculated to decide the direction (sign) and magnitude of the next variational parameter. The derivative is the expectation value of the local energy differentiated with respect to the variational parameter, since it is the expectation value of the local energy we wish to minimize. The expression reads

$$\frac{d\langle E_L(\alpha)\rangle}{d\alpha} = 2\left(\left\langle \frac{E_L(\alpha)}{\psi_T(\alpha)}\frac{d\psi_T(\alpha)}{d\alpha}\right\rangle - \langle E_L(\alpha)\rangle \left\langle \frac{1}{\psi_T(\alpha)}\frac{d\psi_T(\alpha)}{d\alpha}\right\rangle\right). \tag{41}$$

To calculate the next variational parameter, we simply use

$$\alpha_{i+1} = \alpha_i - \lambda \frac{d\langle E_L(\alpha)\rangle}{d\alpha} \tag{42}$$

where λ is a step size we can adjust according to the problem at hand. Due to the wide use of gradient descent in machine learning, λ (often named η) is often called *learning rate*.

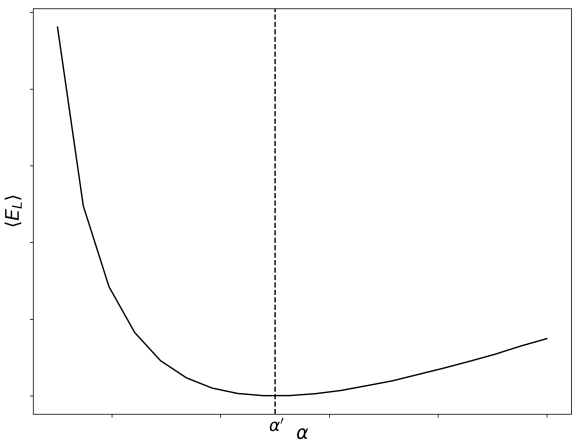


FIG. 1: This is what the expectation value of the local energy as a function of the variational parameter, α, might look like. The function has its minimum at $\alpha = \alpha'$.

Equation (42) is understood as follows. $\langle E_L(\alpha)\rangle$ has a minimum at $\alpha = \alpha'$ as seen in figure 1. If we pick an initial variational parameter $\alpha < \alpha'$, we see that the derivative in equation (41) is negative. Our next pick for α should be larger than the initial pick, and since the derivative is negative at α_i, we must subtract the derivative in equation (41) from our initial α pick. The same logic applies if we choose $\alpha > \alpha'$. This process assumes a well behaved local energy function with a single minimum. Should we encounter a function with several local minima, a stochastic version of gradient descent should be employed.

D. Statistical analysis

We can treat our Monte Carlo simulations as individual experiments, and as with any experiment there are both systematic and statistical errors. Systematic errors are method specific and must be treated differently for each case. These can often be related to the step length or come as a consequence of using an approximate wave function. The statistical errors can be analyzed using the same tools as in any other physics experiment.

1. *Error Estimation for a Stationary Time Series*

We collect a sample of N stochastic variables $\{x_{t_i}\}_{i=1}^N$ which are measurements of the local energy E_L calculated using the analytical expressions previously derived. The measurements are created with a Monte Carlo simulation using Metropolis-Hastings sampling so each point is drawn from a PDF that is conditional on the previous point [15], making the measurements serially correlated. Our sample therefore correspond to a stationary time series and the mean function is constant $\mu_t = \mu$ [14]. We don't know the exact value of this true mean, but we can estimate it using the sample mean

$$m = \frac{1}{N}\sum_{t=1}^{N} x_t. \tag{43}$$

To see how accurate m is as an estimate of μ we need to find the error $\sigma(m)$. This is found by calculating the variance of the sample mean,

$$\begin{aligned}\sigma^2(m) &= \mathrm{Var}(m) = \langle m^2\rangle - \langle m\rangle^2 \\ &= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{k=1}^{N}\left(\langle x_t x_k\rangle - \langle x_t\rangle\langle x_k\rangle\right).\end{aligned} \tag{44}$$

We recognize the covariance matrix in the last term,

$$\mathrm{Cov}(x_i, x_j) = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle \tag{45}$$

where the diagonal elements are the familiar variances $\mathrm{Cov}(x_i, x_i) = \mathrm{Var}(x_i)$. In the special case where our sample consists of independent identically distributed measurements with finite variance σ^2, the two terms in eq. (45) are equal, $\langle x_i x_j\rangle = \langle x_i\rangle\langle x_j\rangle$ so all off-diagonal elements of the matrix would be zero. We could then calculate $\sigma^2(m)$ as,

$$\sigma^2(m) = \frac{1}{N^2}\sum_{t=1}^{N}\mathrm{Var}(x_t) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} \tag{46}$$

But this is clearly not the case for our sample as our data are correlated. Continuing from equation (44), we see that we can simplify the expression further by separating the terms that describe the uncorrelated and correlated parts of the covariance matrix,

$$\sigma^2(m) = \sum_{t=1}^{N}\mathrm{Var}(x_t) + \sum_{t=1}^{N}\sum_{k=t+1}^{N}\mathrm{Cov}(x_t, x_k). \tag{47}$$

Using the identity for a stationary time series, the autocovariance function $\gamma(t,k) = \mathrm{Cov}(x_t, x_k)$ only depends on t and k through the difference $h = |t-k|$. If we write $k = t + h$ then,

$$\begin{aligned}\gamma(t,k) &= \gamma(t, t+h) = \mathrm{Cov}(x_t, x_{t+h}) \\ &= \mathrm{Cov}(x_0, x_h) = \gamma(0,h) \equiv \gamma(h,0) = \gamma(h)\end{aligned} \tag{48}$$

because the time difference between $t + h$ and h is the same as the time difference between h and 0. We can now use this to write the error estimate squared in terms

of the autocovariance,

$$\begin{aligned}\sigma^2(m) &= \frac{1}{N^2}\sum_{t=1}^{N}\mathrm{Var}(x_t) + \frac{2}{N^2}\sum_{h=1}^{N-1}\sum_{t=1}^{N-h}\mathrm{Cov}(x_t, x_{t+h}) \\ &= \frac{1}{N}\gamma(0) + \frac{2}{N^2}\sum_{h=1}^{N-1}(n-h)\gamma(h) \\ &= \frac{\gamma(0)}{N} + \underbrace{\frac{2}{N}\sum_{h=1}^{N-1}\left(1-\frac{h}{N}\right)\gamma(h)}_{e} \end{aligned} \tag{49}$$

where we used that the variance $\mathrm{Var}(x_t) = \gamma(0)$. We name the variable e the truncation error which is the error that comes from the correlation terms. We can also write equation (49) in terms of the autocorrelation function $\kappa(h)$ which is defined as the autocovariance divided by the variance, $\kappa(h) = \gamma(h)/\gamma(0)$. The error estimate squared then becomes,

$$\begin{aligned}\sigma^2(m) &= \frac{\gamma(0)}{N} + \frac{2}{N}\gamma(0)\sum_{h=1}^{N-1}\left(1-\frac{h}{N}\right)\frac{\gamma(h)}{\gamma(0)} \\ &= \frac{\gamma(0)}{N}\underbrace{\left[1 + 2\sum_{h=1}^{N-1}\left(1-\frac{h}{N}\right)\kappa(h)\right]}_{\tau} \end{aligned} \tag{50}$$

where τ is the autocorrelation time which, roughly speaking, can be viewed as an estimate of the average time for the data points to de-correlate. If $\gamma(0) = \sigma^2$ the expressions in equations (49) and (50) becomes

$$\sigma^2(m) = \frac{\sigma^2}{N} + e \quad \text{and} \quad \sigma^2(m) = \frac{\tau}{N}\sigma^2. \tag{51}$$

We immediately notice that for uncorrelated measurements, $\tau = 1$ and $e = 0$, so $\sigma^2(m)$ is simply the standard error of the mean as found in equation (46). This shows that when ignoring the correlation and calculating the error using equation (46), $\sigma(m) = \sigma/\sqrt{N}$, the estimate will be much smaller than the true error. This is not surprising seeing as correlated measurements are often less spread out due to the influence that the previous measurement has on the current one. However, including calculations of covariance or auto-correlation time for millions of datapoints would be computationally expensive. We will therefore use a resampling technique called blocking.

2. *The Blocking Method*

The blocking method is used to reduce the correlation in the data by applying blocking transformations until the correlation of the measurements no longer influence the error estimate. In a sense, we de-correlate the data. The general idea is that we will pre-average our initial data into blocks of increasing size, calculate the error based on these blocks, and use the block size as an estimate of the autocorrelation time.

A blocking transformation involves transforming a dataset $\{x_i, x_2, ..., x_n\}$ into a new dataset $\{x'_1, x'_2, ..., x'_{n'}\}$ that is half as large, $n' = \frac{1}{2}n$, and where each element x'_i is calculated as the average of subsequent pairs of elements in the previous dataset,

$$x'_i = \frac{1}{2}(x_{2i-1} - x_{2i}) \tag{52}$$

The mean m' of the new dataset will be the the same as the mean of the old dataset m, and this is also true for the variance $\sigma^2(m') = \sigma^2(m)$ [5]. From equation (51) we know that $\sigma^2(m) \leq \gamma(0)/n$, and because the size of the dataset is halved in every blocking transformation, the value $\gamma(0)/n$ will increase for each transformation. Therefore, if we continue with repeated blocking transformations, eventually the elements of x' will be uncorrelated and the variance will become constant, i.e, the truncation error $e_k \to 0$ as the number of blocking transformations, k, becomes large.

The blocking method used in this analysis is thoroughly described in [8] and includes a hypothesis test to determine at which blocking transformation k the error becomes constant. Here we will give a short summary of the method, but we refer the readers to [8] for a more detailed description and proof. We start with our original sample with $N = 2^d$ measurements[1]. For each new sample X_i, with $i \in \{0, 1, .., d-1\}$ created using the blocking transformation in equation (52) we calculate the sample variance $\hat{\sigma}_i^2$ and first order autocovariance $\hat{\gamma}_i(1)$[2] as

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{t=1}^{n_i}(x_t - m)^2, \tag{53}$$

$$\hat{\gamma}_i(1) = \frac{1}{n_i}\sum_{t=1}^{n-1}(x_t - m)(x_{t+1} - m). \tag{54}$$

We then compute the test-statistic M_j which is asymptotic χ^2_{d-j} distributed under the hypothesis $\gamma_i(1)$ for all $i \leq j$, so all values of M_j that are larger than $q_{d-j}(1-\alpha)$ are rejected. We choose to use $\alpha = 0.05$, and values for $q_{d-j}(1-\alpha)$ can be found in [1]. We define the test statistic M_j as

$$M_j = \sum_{i=j}^{d-1}\frac{n_i[(n_i-1)\hat{\sigma}_i^2/n_i^2 + \hat{\gamma}_i(1)]^2}{\hat{\sigma}_i^4} \tag{55}$$

for all $1 \leq j \leq d-1$. We then compare the value of the test statistic M_k to q_{d-k} for all k, and choose the

[1] Because the sample size is halved for each blocking tranform it is convenient to use 2^d

[2] The theoretical auto-covariance can be estimated by the sample covariance $\hat{\gamma}(h) = \frac{1}{N}\sum_{t=1}^{n-h}(x_t - m)(x_{t+h} - m)$[14]

smallest k for which $M_k \leq q_{d-k}(1-\alpha)$. This means that the correct final estimate for the variance $\sigma^2(m)$ is found in blocking transformation $i = k$, and the error can be calculated as,

$$\sigma_b(m) = \frac{\hat{\sigma}_k}{\sqrt{n_k}}. \tag{56}$$

E. Numerical Differentiation With autodiff

While analytical differentiation is well and good, there are some situations where analytical differentiation is impractical. If an expression is particularly large or complex, an analytical differentiation may be both hard and may introduce human calculation errors. We have implemented a numerical differentiation library called autodiff [9] for calculating the kinetic energy term of the local energy. We use the numerical differentiation for comparison with the analytical differentiation in regards of numerical precision and computation time. We will also be using the `Armadillo` library [12][13].

IV. RESULTS AND DISCUSSION

A. Non-interacting Particles

We begin by looking at our results obtained for a system of non-interacting hard-sphere bosons as described in section II D 1.

1. *Brute Force Metropolis-Hastings*

Using the brute-force metropolis algorithm we run the simulation with ten equally spaced values for the step size $\Delta r \in [0.1, 1.0]$ and 2^{20} MC-cycles. We can see from figure 2, which shows the acceptance rate as a function of step size, that the acceptance rate decreases slowly as the step size Δr increases. If we choose Δr too high, fewer moves will be accepted and the energy will rarely be updated. If Δr is chosen to be very small, the acceptance rate will be ~ 1 and the proposed moves will almost always be accepted, leading to many very small updates of the energy. Both of these situations require a large number of MC-cycles to give a good approximation of the local energy. We will proceed using the step size $\Delta r = 0.2$ in all simulations where the brute-force metropolis algorithm is used.

The expected local energy for ten different values of the variational parameter $\alpha \in [0.1, 1]$ is then calculated using both the analytic expression for local energy (23), and a numerical calculation of the derivative using the Autodiff library. Both simulations are done for systems of 10, 100, and 500 particles, and for the analytical method in 1, 2 and 3 dimensions, while the numerical approach is only used for 3 dimensions. The expected exact local energy calculated using eq. (24) for $N = 10$, 100 and 500 particles in 1, 2 and 3 dimensions is given in table I, these are the values we will be comparing our simulations to.

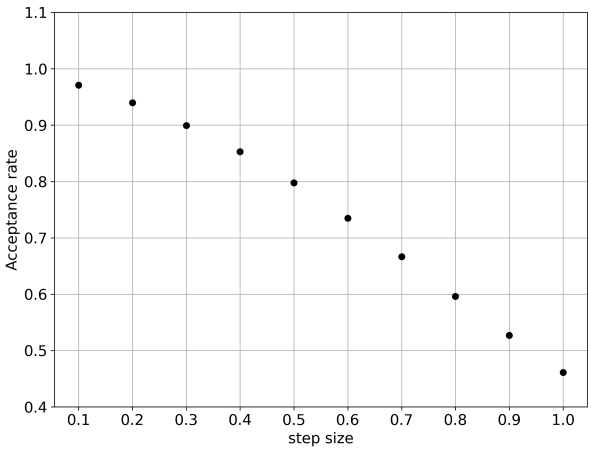


FIG. 2: Acceptance rate as a function of step size for the brute-force metropolis method applied to a system of 10 non-interacting particles in three dimensions using 2^{20} MC cycles for each variation.

The local energy as a function of variational parameter α is illustrated in Figure 3 for $N = 500$ particles. The gray bands show the standard deviation calculated as $\sqrt{\sigma}$ with σ given in (12). As expected, the ground state energy is found with $\alpha = 0.5$, where the local energies are $E_L = 250$, 500 and 750 for 1, 2 and 3 dimensions respectively. In all these points the variance is exactly zero. Similar results for $N = 10$ and 100 particles can be found in E.

Table II shows a comparison between the numerical and analytical approaches. We can see that the two give exactly the same local energy results, but the numeric method is significantly slower for 100 and 500 particles. Notice that the local energy values are accompanied by the *blocking* errors σ_b (56) and not the error calculated as $\sqrt{\sigma^2/N_{MC}}$. The evolution of the blocking error as the number of blocking transformations grow is shown in Figure 4 for $\alpha = 0.4$. At zero iterations the error is equal to the original uncorrelated error $\sqrt{\sigma^2/N_{MC}}$ and as the iterations continue the error increases, illustrating that calculating the uncorrelated error with correlated data leads to an estimate that is far lower than the actual true error. After ~ 10 blocking transformations the error starts to have small fluctuations and after 18 iterations it decreases again. The fluctuations may arise as a consequence of the sample size becoming very small.

2. *Importance sampling*

We will now introduce importance sampling to the metropolis algorithm. Table III shows the energy and

TABLE I: The exact values of local energy calculated as $E_L = \frac{ND}{2}$ (24).

	1D			2D			3D		
N	10	100	500	10	100	500	10	100	500
E_L	5	50	250	10	100	500	15	150	750

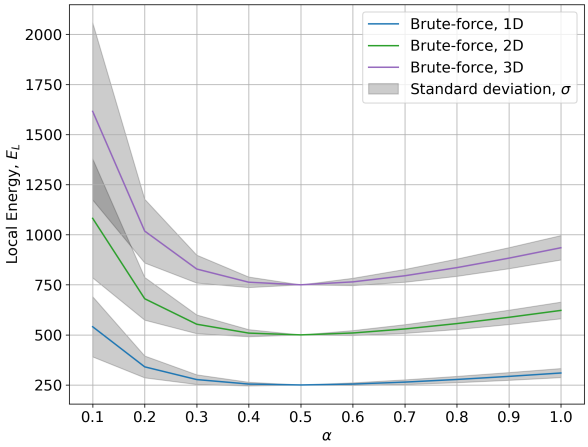


FIG. 3: The expected energy as a function of variational parameter α for the brute-force metropolis method on a system of 500 particles. The step size used is $\Delta r = 0.2$ and the number of MC cycles is 2^{20}.

CPU time for the system of $N = 10, 100, 500$ particles in three dimensions, both using analytical and numerical approaches to the local energy. If we compare the results to those of the brute-force method in table II we can see that the importance sampling is a little bit slower, which is to be expected due to the added calculation of Greens function and the quantum force.

Figure 5 shows the acceptance rate as a function of time step $\Delta t \in [0, 0.5]$. Here we can see that the acceptance rate decreases quite rapidly as the time step increases. We choose to use $\Delta t = 0.01$ in our simulations as this gives an acceptance rate of approximately 99% for $N = 10$, 100 and 500 particles.

We again find that $\alpha = 0.5$ gives the correct values for the local energy as can be seen in figure 6 which shows local energies as a function of alpha. Comparing these to the result obtained for brute-force metropolis sampling (figure 3), we can see that the standard deviation (grey band) is smaller for importance sampling than it is for brute force. This can also be seen in the error estimates shown in table III. Studying this further by comparing the blocking errors of the two methods for an increasing number of MC-cycles (figure 7) we can see that for $N_{mc} < 2^8$ the brute-force metropolis method gives a lower error then metropolis with importance sampling added, but for $N_{mc} > 2^8$ the error estimate from importance sampling is consistently lower. This is because the sampling is less random and therefore the energies will fluctuate less.

TABLE II: Results obtained using the brute-force metropolis algorithm with the analytic and numeric calculation of local energy E_L for $N = 10$, 100, and 500 particles in 3 dimensions. We use 2^{20} MC-cycles for each variation $\alpha \in [0.1, 1]$, and the brute-force step size is $\Delta r = 0.2$. The error estimates σ_b are obtained using the blocking method, and the MC sampling is done using parallelized code with a maximum of 256 threads.

		Analytic		Numeric	
N	α	$E_L \pm \sigma_b$	CPU [s]	$E_L \pm \sigma_b$	CPU [s]
10	0.1	32.21 ± 0.26	0.83	32.21 ± 0.26	0.88
	0.2	20.313 ± 0.097	0.73	20.313 ± 0.097	0.77
	0.3	16.552 ± 0.042	0.67	16.552 ± 0.042	0.71
	0.4	15.251 ± 0.016	0.64	15.251 ± 0.016	0.67
	0.5	15.0 ± 0.0	0.61	15.0 ± 0.0	0.67
	0.6	15.291 ± 0.012	0.61	15.291 ± 0.012	0.65
	0.7	15.891 ± 0.021	0.57	15.891 ± 0.021	0.62
	0.8	16.692 ± 0.028	0.56	16.692 ± 0.028	0.59
	0.9	17.631 ± 0.034	0.55	17.631 ± 0.034	0.61
	1.0	18.66 ± 0.04	0.54	18.66 ± 0.04	0.58
100	0.1	323.05 ± 0.81	8.44	323.05 ± 0.81	13.84
	0.2	203.40 ± 0.30	7.56	203.40 ± 0.30	12.55
	0.3	165.64 ± 0.13	7.038	165.64 ± 0.13	12.22
	0.4	152.55 ± 0.05	6.88	152.553 ± 0.050	12.079
	0.5	150.0 ± 0.0	6.45	150.0 ± 0.0	11.91
	0.6	152.926 ± 0.034	6.29	152.926 ± 0.034	11.84
	0.7	159.034 ± 0.061	6.16	159.034 ± 0.061	12.23
	0.8	167.133 ± 0.081	6.0028	167.133 ± 0.081	11.83
	0.9	176.610 ± 0.099	5.88	176.610 ± 0.099	11.69
	1.0	186.99 ± 0.12	5.77	186.99 ± 0.11	11.64
500	0.1	1615.81 ± 1.99	59.64	1615.81 ± 1.98	276.60
	0.2	1018.22 ± 0.75	56.29	1018.22 ± 0.75	273.73
	0.3	828.73 ± 0.34	56.82	828.73 ± 0.34	270.85
	0.4	762.86 ± 0.13	55.91	762.86 ± 0.13	269.64
	0.5	750.0 ± 0.0	54.97	750.0 ± 0.0	268.68
	0.6	764.56 ± 0.087	53.76	764.565 ± 0.087	266.84
	0.7	795.08 ± 0.15	53.36	795.08 ± 0.15	264.35
	0.8	835.66 ± 0.19	54.27	835.66 ± 0.19	262.91
	0.9	882.96 ± 0.24	53.49	882.96 ± 0.24	262.37
	1.0	934.95 ± 0.28	53.18	934.95 ± 0.28	260.32

3. *Gradient descent*

Using the gradient descent method, we can search for the optimal variational parameter for the ground state energy. Starting at an initial $\alpha = 0.2$ we run the search using 2^{20} MC-cycles for each variational parameter. Figure 8 shows the local energy as a function of alpha for a system of 10 particles in three dimensions and figure

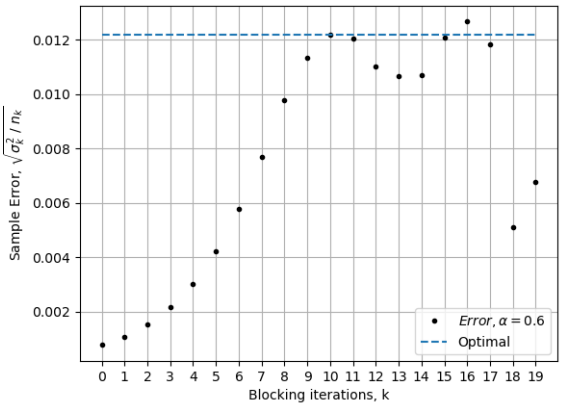


FIG. 4: Evolution of the blocking error as a function of iterations. The data is simulated using the brute force metropolis algorithm with a step size of $\Delta r = 0.2$, 2^{20} MC cycles and analytic expressions for energy. We get similar results with the numeric method.

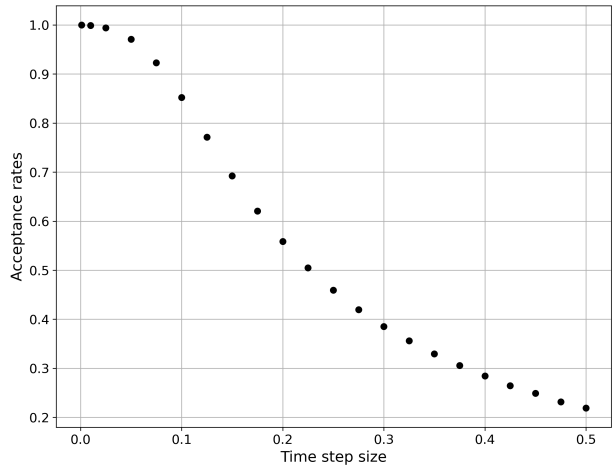


FIG. 5: Acceptance rate as a function oftime step for the metropolis method with importance sampling applied to a system of 10 non-interacting particles in three dimensions using 2^{20} MC cycles for each variation.

9 shows the corresponding error $\sqrt{\sigma^2/N_{mc}}$ with σ given in (12). We can see that the algorithm starts by taking large steps while it is far away from the minimum, and as it approaches the minimum the steps become smaller and smaller. It eventually finds a minimum at $\alpha = 0.5$ which corresponds well with our results from the brute-force and importance sampling methods.

B. Interacting Particles

We will now include interactions with an elliptic potential in our system as described in section II D 2.

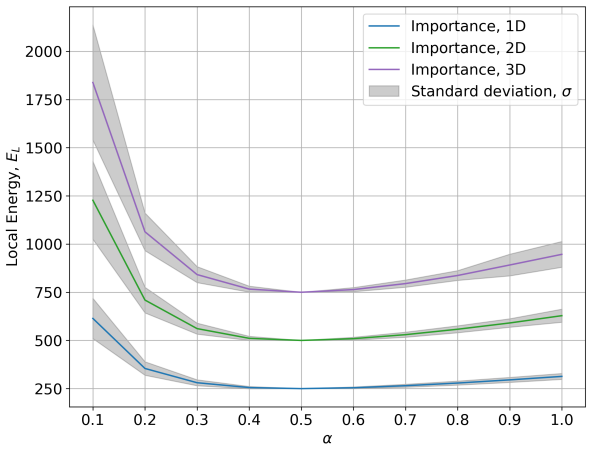


FIG. 6: The expected energy as a function of variational parameter α for the importance sampling metropolis method on a system of 500 particles. The step size used is $\Delta t = 0.01$ and the number of MC cycles is 2^{20}.

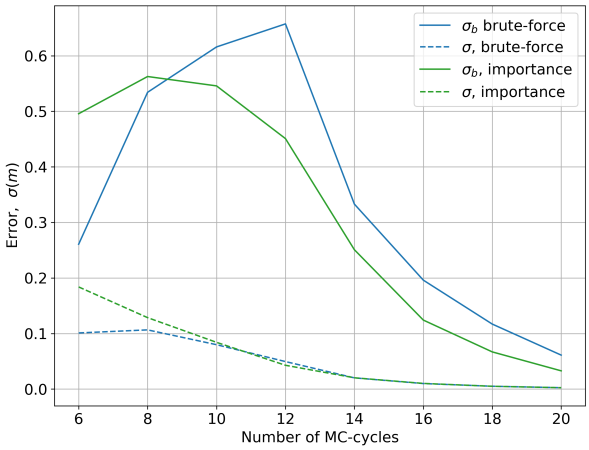


FIG. 7: Both the blocking errors σ_b(full lines) and the initial errors σ(dashed lines) as a function of number of MC cycles (2^N). The simulations are done on a system of 10 particles in 3 dimensions both using the brute-force method (blue) with a step size $\Delta r = 0.2$ and importance sampling(green) with a time step $\Delta t = 0.01$

1. *Importance Sampling with Interacting Particles*

The results from importance sampling Metropolis-Hastings for the system of interacting particles in 3 dimensions are listed in table IV. For 10 variational parameters $\alpha \in [0.1, 1]$ we find that none of the α's give a $\sigma^2 = 0$. As explained in section II B, we know that the minimum of the expectation value of the local energy is located where the variance is zero. In practice, we don't necessarily expect the variance to exactly equal zero, but

TABLE III: Results obtained using the importance sampling metropolis algorithm for the analytic and numeric calculation of local energy E_L for $N = 10$, 100, and 500 particles in 3 dimensions. The results are obtained using 2^{20} MC-cycles for each variation $\alpha \in [0.1, 1]$, and the time step is $\Delta t = 0.01$. The error estimates σ_b are obtained using the blocking method, and the MC sampling is done using parallelized code with a maximum of 256 cores.

		Analytic		Numeric	
N	α	$E_L \pm \sigma_b$	CPU [s]	$E_L \pm \sigma_b$	CPU [s]
10	0.1	36.86 ± 0.18	1.63	37.20 ± 0.20	2.31
	0.2	21.380 ± 0.062	1.32	21.297 ± 0.061	2.09
	0.3	16.843 ± 0.026	1.28	16.834 ± 0.025	2.08
	0.4	15.3229 ± 0.0095	1.29	15.3330 ± 0.0093	2.07
	0.5	15.0 ± 0.0	1.29	15.0 ± 0.0	2.11
	0.6	15.2830 ± 0.0062	1.30	15.2675 ± 0.0062	2.08
	0.7	15.902 ± 0.011	1.29	15.893 ± 0.010	2.07
	0.8	16.713 ± 0.015	1.28	16.738 ± 0.014	2.07
	0.9	17.716 ± 0.017	1.29	17.693 ± 0.017	2.08
	1.0	18.803 ± 0.019	1.29	18.809 ± 0.019	2.07
100	0.1	368.37 ± 0.58	10.18	367.42 ± 0.56	19.80
	0.2	213.07 ± 0.19	9.95	213.16 ± 0.18	19.92
	0.3	168.297 ± 0.083	9.98	168.326 ± 0.086	19.69
	0.4	153.285 ± 0.027	10.06	153.302 ± 0.031	19.74
	0.5	150.0 ± 0.0	9.91	150.0 ± 0.0	20.23
	0.6	152.781 ± 0.021	9.96	152.781 ± 0.019	19.79
	0.7	159.049 ± 0.034	9.93	159.018 ± 0.034	20.17
	0.8	167.397 ± 0.045	9.91	167.406 ± 0.044	19.75
	0.9	177.234 ± 0.052	9.98	177.231 ± 0.051	19.90
	1.0	188.133 ± 0.064	9.96	188.108 ± 0.067	19.62
500	0.1	1839.33 ± 1.29	104.51	1840.11 ± 1.29	322.48
	0.2	1064.27 ± 0.45	101.32	1064.18 ± 0.42	324.17
	0.3	841.88 ± 0.18	106.11	842.25 ± 0.18	322.75
	0.4	766.214 ± 0.065	98.95	766.273 ± 0.067	323.78
	0.5	750.0 ± 0.0	100.99	750.0 ± 0.0	323.83
	0.6	763.765 ± 0.043	100.43	763.810 ± 0.045	326.30
	0.7	794.890 ± 0.076	98.90	794.818 ± 0.075	324.19
	0.8	836.89 ± 0.10	99.79	836.79 ± 0.10	324.83
	0.9	889.87 ± 2.44	98.66	885.85 ± 0.12	323.51
	1.0	942.490 ± 2.024	98.71	944.67 ± 2.95	320.06

that is what we see in the non-interacting case in tables II and III. Due to the increased complexity of the interacting wave function, quantum force and local energy, combined with the stochastic nature of the Monte Carlo calculations, we expect the variance to be close to, but not equal to zero.

From the data in table IV we see that the variance has a minimum at approximately $\alpha = 0.5$ for both 10 and 50 particles, but the calculation of the local energy is clearly lower at $\alpha = \{0.2, 0.3\}$. We suspect that there is

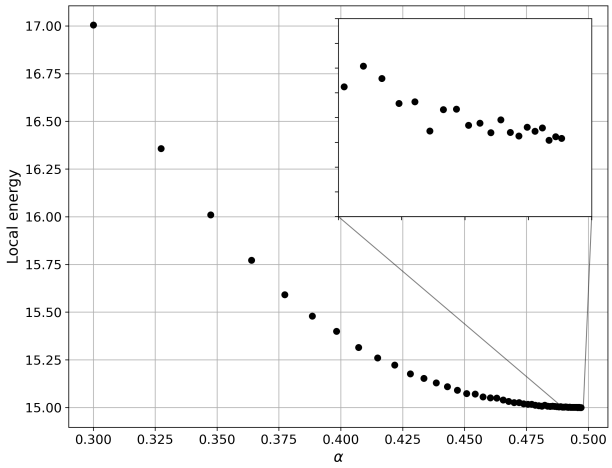


FIG. 8: The local energy of the gradient descent search for the best variational parameter α using importance sampling with a time step $\Delta t = 0.01$. The simulation is done for a system of 10 particles in 3 dimensions, the learning rate is 10^{-4} and the number of MC-cycles is 2^{20}.

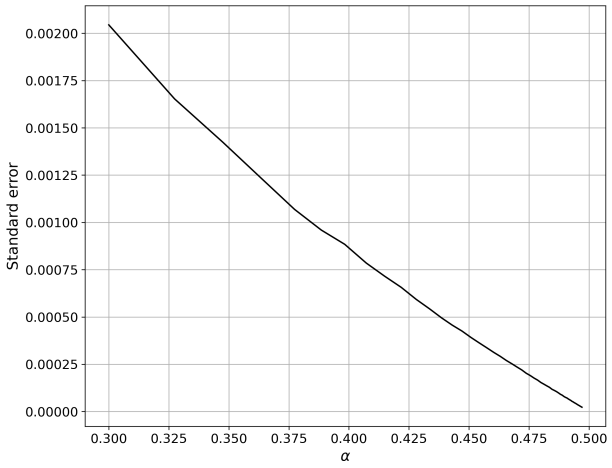


FIG. 9: The error calculated as $\sqrt{\sigma^2/N_{mc}}$ for the gradient descent search in figure 8

something wrong with our calculation of the local energy. The non-interacting minimum is located at $\alpha = 0.5$ and we expect the interaction to slightly shift this value to somewhere in the vicinity of $\alpha = 0.5$, but not all the way down to $\alpha = \{0.2, 0.3\}$. The variance supports that the minimum is located somewhere around $\alpha = 0.5$, but we find it hard to rely on the data from the simulation of the interacting system when the variance and energy data do not match each other. Since the non-interaction data seems good, we believe that the error is not located in the Monte Carlo, Metropolis-Hastings or importance sampling implementations, but rather in either the wave function, local energy, or quantum force implementation

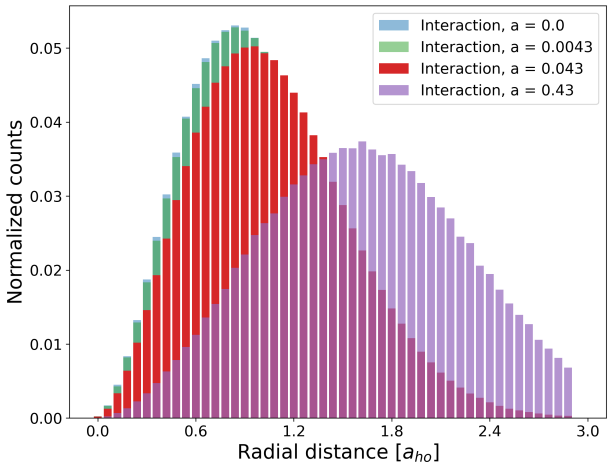


FIG. 10: Particle density as a function of radial distance. The y axis is normalized so that the area of each individual set of bins are 1. The data is generated with importance sampling with a step size of 0.01, 2^{20} MC-cycles, 10 particles, and a variational parameter of 0.5. 50 bins are used on the interval [0, 3] a_{ho}.

for the interacting system. Calculations with gradient descent on the interacting system show a convergence towards a value α slightly larger than 0.5, but shows the same disconnection from the local energy values which are lowest at $\alpha = \{0.2, 0.3\}$.

Table IV also lists the acceptance rate for each variational parameter. The acceptance rate is normalized to the interval $A \in [0, 1]$ where 0 is no accepted moves and 1 is only accepted moves. We can see that the acceptance rate starts at 0.97 for both 10 and 50 particles and decreases as α increases. Though the acceptance rate decreases by approximately 0.1 for both 10 and 50 particles, it stays above 0.9 at the critical points of $\alpha = \{0.2, 0.3, 0.5\}$ where the energy and variance minima are located. The mismatch of the two minima does not seem to be rooted in a low acceptance rate, and we observe from our calculations that the two minima no not move when we adjust the importance sampling step size which is a major influencer of the acceptance rate.

As of now, we have not been able to find the source of the error and we must make do with what we have at hand now.

2. *Onebody density*

Figure 10 displays the particle density as a function of radial distance. The figure shows the interacting case for $a = 0, 0.0043, 0.043, 0.43$. The shapes of the distributions both resemble normal distributions with exponential tails. We see from the data that the particles get pushed away from the origin, and away from each other, for larger values of a. Since a decides the strength of

TABLE IV: Results for the MC simulations using importance sampling with interaction included for $N = 10$, 50 and 100 particles. The time step used is $\Delta t = 0.01$ and each variation uses 2^{20} MC-cycles, the hard-core diameter of the particles is set to $a = 0 - 0043$. The local energy E_L is calculated using analytic expressions, σ denotes the variance as calculated in (12), A denotes the acceptance rate and the last column shows the CPU time in seconds. The MC sampling is done using parallelized code with a maximum of 256 cores.

N	α	E_L	σ^2	A	CPU [s]
10	0.1	52.57	396.21	0.97	0.71
	0.2	42.04	77.60	0.96	0.59
	0.3	42.04	21.19	0.94	0.58
	0.4	45.24	6.32	0.93	0.61
	0.5	50.19	4.18	0.92	0.57
	0.6	55.91	7.18	0.91	0.57
	0.7	62.09	13.52	0.89	0.59
	0.8	68.71	22.86	0.88	0.57
	0.9	75.51	31.80	0.88	0.57
	1.0	82.42	44.21	0.86	0.58
50	0.1	269.05	5874.28	0.97	210.67
	0.2	216.18	996.53	0.95	207.05
	0.3	214.49	290.1	0.94	202.63
	0.4	231.57	127.78	0.93	202.74
	0.5	256.46	125.64	0.91	199.09
	0.6	285.69	180.51	0.90	198.75
	0.7	317.27	266.03	0.88	191.29
	0.8	350.71	369.54	0.86	186.67
	0.9	385.57	488.82	0.84	183.19
	1.0	420.43	703.24	0.82	178.66

the interaction term of the wave function, also called the Jastrow factor, we expect it to have a significant effect on the distribution of particles. From equation (18) we see that at $a = 0$ equates to the non-interacting wave function. As a increases, the Jastrow factor sets the wave function to zero should any two particles be closer than a, forcing the particles to be spread farther away from each other. This is exactly what we see in figure 10. Increasing a widens the distribution of particles which is the same as saying that all particles are farther spaced from one another.

We also observe that as a increases, the peak of the particle distribution gets pushed farther away from the origin. Since figure 10 represents density per *radial* distance, each bin represents a spherical shell of volume. Imagine a regular onion. It is comprised of layers, wrapped around and on top of each other until a sphere of onion is achieved. It is easy to see that the outermost layer needs to cover a larger spherical area than the innermost layer, meaning that the outermost layer, assuming constant layer thickness, has a larger volume of

onion than the innermost layer. The same can be said for the particle distribution in figure 10. Each bin indeed represents a spherical layer of particles and by the onion analogy, the outermost layer has a larger volume and can thus contain more particles. When we increase the interaction parameter a, the particles are forced to consume a larger volume, and the outer layers have more volume than the inner ones.

Even though there is more volume for the particles at greater radial distances, the particles are still confined to approximately $r = [0, 3a_{\text{ho}}]$ due to the shape of the wave function. The probability of finding a particle at position $r \to \infty$ decreases exponentially, giving rise to the exponential tail of the particle distributions in figure 10.

V. CONCLUSION

In this project we have used Variational Monte Carlo method to find the ground state energy of a Bose-Einstein condensate of non-interacting hard-sphere bosons, and begun the process of doing the same for interacting bosons. We found that while using the metropolis sampling method combined with importance sampling is slower than the brute-force approach, the error estimates proved to be smaller and thus the energy calculation more precise. In the non-interacting case we find that the ground state energy corresponds to a variational parameter $\alpha = 0.5$ as is expected from theory. Using a numeric method to calculate the local energy yields good results, but does require more processing power then simply using the analytical expressions. While we were not able to resolve the bug in our calculation of local energy for interacting particles within the time frame of this project, the calculations of one-body density for the interacting system does show the expected behaviour where an increase in the hard-core diameter of the bosons results in a larger spread of the particles.

[1] Behnke, O., Kröninger, K., Schott, G., & Schörner-Sadenius, T. 2013, Data analysis in high energy physics: a practical guide to statistical methods (Weinheim: Wiley-VCH), doi: 10.1002/9783527653416

[2] Bishop, C. 2013, Pattern Recognition and Machine Learning: All "just the Facts 101" Material, Information science and statistics (Springer (India) Private Limited). https://books.google.no/books?id=HL4HrgEACAAJ

[3] Cornell, E., & Wieman, C. 1998, Scientific American, 278, 40

[4] DuBois, J. L., & Glyde, H. R. 2001, Phys. Rev. A, 63, 023602, doi: 10.1103/PhysRevA.63.023602

[5] Flyvbjerg, H., & Petersen, H. G. 1989, The Journal of Chemical Physics, 91, 461, doi: 10.1063/1.457480

[6] Griffiths, D. J. 2004, Introduction to Quantum Mechanics (2nd Edition), 2nd edn. (Pearson Prentice Hall)

[7] Hashimoto, T., Ota, Y., Tsuzuki, A., et al. 2020, Science Advances, 6, doi: 10.1126/sciadv.abb9052

[8] Jonsson, M. 2018, Phys. Rev. E, 98, 043304, doi: 10.1103/PhysRevE.98.043304

[9] Leal, A. 2021, autodiff, https://github.com/autodiff/autodiff/, GitHub

[10] Lindstrøm, T., & Hveberg, K. 2016, Flervariabel Analyse med Lineær Algebra (Gyldendal Akademisk)

[11] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087, doi: 10.1063/1.1699114

[12] Sanderson, C., & Curtin, R. 2016, Journal of Open Source Software, 1, 26, doi: 10.21105/joss.00026

[13] —. 2018, Lecture Notes in Computer Science, 422–430, doi: 10.1007/978-3-319-96418-8_50

[14] Shumway, R. H., & Stoffer, D. S. 2017, Time Series Analysis and Its Applications, 4th edn. (Springer International Publishing), doi: 10.1007/978-3-319-52452-8

[15] Toulouse, J., Assaraf, R., & Umrigar, C. J. 2015, Introduction to the variational and diffusion Monte Carlo methods. https://arxiv.org/abs/1508.02989

Appendix A: Derivative of the Trial Wavefunction

1. First derivative

The trial wavefunction for the ground state with N atoms is given by

$$\psi_T(\boldsymbol{r}) = \psi_T(\boldsymbol{r}_1, ..., \boldsymbol{r}_N, \alpha\beta) = \prod_i \phi(\boldsymbol{r}_i) \exp\left(\sum_{j<k} u(r_{jk})\right) = \phi(\boldsymbol{r})U(\boldsymbol{r}) \tag{A1}$$

where we have defined $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ and $u(r_{ij}) = \ln f(r_{ij})$. The first derivative for particle k is,

$$\nabla_k \psi_T(\boldsymbol{r}) = [\nabla_k \phi(\boldsymbol{r})]\, U(\boldsymbol{r}) + \phi(\boldsymbol{r}) \nabla_k U(\boldsymbol{r}) \tag{A2}$$

where the gradients are,

$$\nabla_k \phi(\boldsymbol{r}) = \nabla_k \prod_i \phi(\boldsymbol{r}_i) = \nabla_k \phi(\boldsymbol{r}_k) \prod_{i \neq k} \phi(\boldsymbol{r}_i) \tag{A3}$$

$$\nabla_k U(\boldsymbol{r}) = \nabla_k \exp\left(\sum_{j<m} u(r_{jm})\right) = \exp\left(\sum_{j<m} u(r_{jm})\right) \sum_{j<m} \nabla_k u(r_{jm}) \tag{A4}$$

Because the r_{jm} is symmetric under the interchange of variables, we can rewrite the last sum in (A4) as,

$$\sum_{j<m} \nabla_k u(r_{jm}) = \sum_{j=1} \sum_{m=j+1} \nabla_k u(r_{jm}) = \sum_{j=1}^{k-1} \nabla_k u(r_{jk}) + \sum_{l=k+1}^{N} \nabla_k u(r_{kl}) = \sum_{l \neq k} \nabla_k u(r_{kl})$$

Inserting these back into the expression for the gradient of the trial wavefunction (A2) we get,

$$\nabla_k \psi_T(\boldsymbol{r}) = \nabla_k \phi(\boldsymbol{r}_k) \left[\prod_{i \neq k} \phi(\boldsymbol{r}_i)\right] \exp\left(\sum_{j<k} u(r_{jk})\right) + \left[\prod_i \phi(\boldsymbol{r}_i)\right] \exp\left(\sum_{j<m} u(r_{jm})\right) \sum_{l \neq k} \nabla_k u(r_{kl}) \tag{A5}$$

2. Second derivative

We can now re-write the differential operator ∇_k using variable substitution $r_{kl} = |\boldsymbol{r}_k - \boldsymbol{r}_l|$,

$$\nabla_k = \nabla_k \frac{\partial r_{kl}}{\partial r_{kl}} = \nabla_k r_{kl} \frac{\partial}{\partial r_{kl}} = \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}} \frac{\partial}{\partial r_{kl}} \tag{A6}$$

The gradient of $u(r_{kl})$ can then be written as,

$$\nabla_k u(r_{kl}) = \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}} \frac{\partial u'(r_{kl})}{\partial r_{kl}} = \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}} u'(r_{kl}) \tag{A7}$$

The second derivative is,

$$\begin{aligned}
\nabla_k^2 u(r_{kl}) &= \nabla_k \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}} u'(r_{kl}) = \left(\nabla_k \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}}\right) u'(r_{kl}) + \frac{\boldsymbol{r}_k - \boldsymbol{r}_l}{r_{kl}} \nabla_k u'_{r_{kl}} \\
&= \frac{\nabla_k(\boldsymbol{r}_k - \boldsymbol{r}_l) r_{kl} - (\boldsymbol{r}_k - \boldsymbol{r}_l) \nabla_k r_{kl}}{r_{kl}^2} \frac{\partial u(r_{kl})}{\partial r_{kl}} + \frac{(\boldsymbol{r}_k - \boldsymbol{r}_l)}{r_{kl}} \frac{(\boldsymbol{r}_k - \boldsymbol{r}_l)}{r_{kl}} \frac{\partial}{\partial r_{kl}} \frac{\partial u(r_{kl})}{\partial r_{kl}} \\
&= \frac{d-1}{r_{kl}} u'(r_{kl}) + u''(r_{kl})
\end{aligned}$$

where d is the number of dimensions of $\boldsymbol{r}_i$. In our case it will be $d = 3$, so the second derivative becomes,

$$\nabla_k u(r_{kl}) = \frac{2}{r_{kl}} u'(r_{kl}) + u''(r_{kl}) \tag{A8}$$

The laplacian of the wavefunction can be found using the result from eq. (A5),

$$\begin{aligned}\nabla_k^2\psi_T(\boldsymbol{r}) &= \nabla_k(\nabla_k\psi_T(\boldsymbol{r})) \\ &= \nabla_k\left[\nabla_k\phi(\boldsymbol{r}_k)\left[\prod_{i\neq k}\phi(\boldsymbol{r}_i)\right]\exp\left(\sum_{j<k}u(r_{jk})\right)\right] + \nabla_k\left[\left[\prod_i\phi(\boldsymbol{r}_i)\right]\exp\left(\sum_{j<m}u(r_{jm})\right)\sum_{l\neq k}\nabla_k u(r_{kl})\right] \\ &= (I)+(II) \end{aligned} \tag{A9}$$

We can simplify the terms using the expressions for the gradient in (A7) and (A8). Starting with term (I), we get,

$$\begin{aligned}(I) &= \nabla_k^2\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right) + \nabla_k\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\nabla_k\exp\left(\sum_{j<k}u(r_{jk})\right) \\ &= \nabla_k^2\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right) + \nabla_k\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k u(r_{kl}) \\ &= \nabla_k^2\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right) + \nabla_k\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\frac{\boldsymbol{r}_k-\boldsymbol{r}_l}{r_{kl}}u'(r_{kl})\end{aligned}$$

And for term (II) we have,

$$\begin{aligned}(II) =& \nabla_k\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k u(r_{kl}) \\ &+\prod_i\phi(\boldsymbol{r}_i)\nabla_k\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k u(kl) \\ &+\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k^2 u(kl) \\ =& \nabla_k\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k u(r_{kl}) \\ &+\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k u(r_{kl})\sum_{i\neq k}\nabla_k u(r_{ki}) \\ &+\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\nabla_k^2 u(r_{kl}) \\ =& \nabla_k\phi(\boldsymbol{r}_k)\prod_{i\neq k}\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\frac{\boldsymbol{r}_k-\boldsymbol{r}_l}{r_{kl}}u'(r_{kl}) \\ &+\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\sum_{i\neq k}\frac{(\boldsymbol{r}_k-\boldsymbol{r}_l)}{r_{kl}}\frac{(\boldsymbol{r}_k-\boldsymbol{r}_i)}{r_{ki}}u'(r_{kl})u'(r_{ki}) \\ &+\prod_i\phi(\boldsymbol{r}_i)\exp\left(\sum_{j<k}u(r_{jk})\right)\sum_{l\neq k}\left(\frac{2}{r_{kl}}u'(r_{kl})+u''(r_{kl})\right)\end{aligned}$$

Notice that the first expression in (II) and the last expression in (I) are equal. We can now combine the two and divide by the trial-wavefunction in eq. (A1), the second derivative becomes

$$\begin{aligned}\frac{1}{\psi_T(\boldsymbol{r})}\nabla_k^2\psi_T(\boldsymbol{r}) =& \frac{\nabla_k^2\phi(\boldsymbol{r}_k)}{\phi(\boldsymbol{r}_k)} + \frac{2\nabla_k\phi(\boldsymbol{r}_k)}{\phi(\boldsymbol{r}_k)}\sum_{l\neq k}\frac{\boldsymbol{r}_k-\boldsymbol{r}_l}{r_{kl}}u'(r_{kl}) \\ &+ \sum_{l\neq k}\sum_{i\neq k}\frac{(\boldsymbol{r}_k-\boldsymbol{r}_l)}{r_{kl}}\frac{(\boldsymbol{r}_k-\boldsymbol{r}_i)}{r_{ki}}u'(r_{kl})u'(r_{ki}) \\ &+ \sum_{l\neq k}\left(\frac{2}{r_{kl}}u'(r_{kl}) + u''(r_{kl})\right) \end{aligned} \tag{A10}$$

Appendix B: Local Energy

1. Non-interacting particles

For non-interacting particles we can set $a = 0$ making the Jastrow factor equal to one, menaing there is no correlation between the particles. The trial wavefunction and the Hamiltonian is then,

$$\psi_T(\boldsymbol{r}) = \prod_i \phi(\boldsymbol{r}_i) = \prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} \qquad \text{and} \qquad H = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{2m}\nabla_k^2 + V_{ext}(\boldsymbol{r}_k)\right)$$

where $V_{ext}(\boldsymbol{r}_k)$ is given in (13). The local energy of the trial wavefunction is then,

$$E_L = \frac{1}{\psi_T(\boldsymbol{r})}H\psi_T(\boldsymbol{r}) = \frac{1}{\psi_T(\boldsymbol{r})}\sum_{k=1}^{N}\left(\frac{-\hbar^2}{2m}\nabla_k^2 + V_{ext}(\boldsymbol{r}_k)\right)\psi_T(\boldsymbol{r}) = \frac{1}{\psi_T(\boldsymbol{r})}\sum_{k=1}^{N}\left(\frac{-\hbar^2}{2m}\nabla_k^2\psi_T(\boldsymbol{r}) + V_{ext}(\boldsymbol{r}_k)\psi_T(\boldsymbol{r})\right) \tag{B1}$$

We will first caclutate the product $\nabla_k^2\psi_T(\boldsymbol{r})$ where $\nabla_k^2 = \left(\frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2}\right)$, calculating the derivatives separately

$$\begin{aligned}
\frac{\partial^2}{\partial x_k^2}\psi_T(\boldsymbol{r}) &= \frac{\partial^2}{\partial x_k^2}\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} = \frac{\partial}{\partial x_k}\left(-2\alpha x_k\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)}\right) = 2\alpha(2\alpha x_k^2 - 1)\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} \\
\frac{\partial^2}{\partial y_k^2}\psi_T(\boldsymbol{r}) &= \frac{\partial^2}{\partial y_k^2}\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} = \frac{\partial}{\partial y_k}\left(-2\alpha y_k\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)}\right) = 2\alpha(2\alpha y_k^2 - 1)\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} \\
\frac{\partial^2}{\partial z_k^2}\psi_T(\boldsymbol{r}) &= \frac{\partial^2}{\partial z_k^2}\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)} = \frac{\partial}{\partial z_k}\left(-2\alpha\beta z_k\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)}\right) = 2\alpha\beta(2\alpha\beta z_k^2 - 1)\prod_i e^{-\alpha(x_i^2+y_i^2+\beta z_i^2)}
\end{aligned}$$

inserting these into $\nabla_k^2\psi_T(\boldsymbol{r})$ we have

$$\begin{aligned}
\nabla_k^2\psi_T(\boldsymbol{r}) &= \frac{\partial^2}{\partial x_k^2}\psi_T(\boldsymbol{r}) + \frac{\partial^2}{\partial y_k^2}\psi_T(\boldsymbol{r}) + \frac{\partial^2}{\partial z_k^2}\psi_T(\boldsymbol{r}) = 2\alpha\left[(2\alpha x_k^2 - 1) + (2\alpha y_k^2 - 1) + \beta(2\alpha\beta z_k^2 - 1)\right]\psi_T(\boldsymbol{r}) \\
&= 2\alpha\left[2\alpha(x_k^2 + y_k^2 + \beta^2 z_k^2) - 2 - \beta\right]\psi_T(\boldsymbol{r})
\end{aligned}$$

We can now find the expression for the local energy in (22),

$$E_L = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{m}\alpha\left[(x_k^2 + y_k^2 + \beta^2 z_k^2) - 2 - \beta\right] + V_{ext}(\boldsymbol{r}_k)\right)$$

With a spherical potential trap $V_{ext}(\boldsymbol{r}) = \frac{1}{2}m\omega_{ho}^2\boldsymbol{r}^2$, and $\beta = 1$ we can use the same procedure as above to find the local energy for N particles in one, two and three dimentions,

$$\text{1D:} \qquad E_L = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{m}\alpha(2\alpha x_k^2 - 1) + \frac{1}{2}m\omega_{ho}^2 x_k^2\right) \tag{B2}$$

$$\text{2D:} \qquad E_L = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{m}\alpha(2\alpha(x_k^2 + y_k^2) - 2) + \frac{1}{2}m\omega_{ho}^2(x_k^2 + y_k^2)\right) \tag{B3}$$

$$\text{3D:} \qquad E_L = \sum_{k=1}^{N}\left(\frac{-\hbar^2}{m}\alpha(2\alpha(x_k^2 + y_k^2 + z_k^2) - 3) + \frac{1}{2}m\omega_{ho}^2(x_k^2 + y_k^2 + z_k^2)\right) \tag{B4}$$

Generalized in D dimensions we have,

$$\begin{aligned} E_L &= \sum_{k=1}^{N} \left(\frac{\hbar^2}{m} \alpha (2\alpha \boldsymbol{r}_k^2 - D) + \frac{1}{2} m \omega_{ho}^2 \boldsymbol{r}_k^2 \right) \\ &= \frac{-\hbar^2}{m} \alpha \left(2\alpha \sum_{k=1}^{N} \boldsymbol{r}_k^2 - \sum_{k=1}^{N} D \right) + \frac{1}{2} m \omega_h o^2 \sum_{k=1}^{N} \boldsymbol{r}_k^2 \\ &= \frac{\hbar^2 \alpha N D}{m} - \left(\frac{2\hbar^2 \alpha^2}{m} - \frac{1}{2} m \omega_{ho}^2 \right) \sum_{k=1}^{N} \boldsymbol{r}_k^2 \end{aligned} \tag{B5}$$

where $\boldsymbol{r}_k^2 = x_k^2, x_k^2 + y_k^2, x_k^2 + y_k^2 + z_k^2$ for $D = 1, 2, 3$ dimensions respectively.

2. Interacting particles

For interactive particles

$$E_L = \frac{1}{\psi_T} H \psi_T = \sum_{i=1}^{N} \frac{1}{2} \left(-\frac{\nabla_i^2 \psi_T}{\psi_T} + x_i^2 + y_i^2 + \gamma^2 z_i^2 \right) \tag{B6}$$

we can use the second derivative in equation (A10) with

$$\begin{aligned} \frac{\nabla_k \phi(\boldsymbol{r}_k)}{\phi(\boldsymbol{r}_k)} &= -2\alpha (x_k, y_k, \beta z_k) \\ \frac{\nabla_k^2 \phi(\boldsymbol{r}_k)}{\phi(\boldsymbol{r}_k)} &= 2\alpha [2\alpha (x_k^2 + y_k^2 + \beta^2 z_k^2) - 2 - \beta] \\ u'(r_{kj}) &= \frac{\partial}{\partial r_{kj}} \left(1 - \frac{a}{r_{kj}} \right) = \left(1 - \frac{a}{r_{kj}} \right)^{-1} \frac{a}{r_{kj}^2} = \frac{a}{r_{kj}(r_{kj} - a)} \\ u''(r_{kj}) &= \frac{\partial}{\partial r_{kj}} \left(\frac{a}{r_{kj}(r_{kj} - a)} \right) = \frac{a^2 - 2 r_{kj} a}{r_{kj}^2 (r_{kj} - a)^2} \end{aligned}$$

so

$$\frac{1}{\psi_T} \nabla_i^2 \psi_T = 2\alpha (2\alpha \boldsymbol{r}_i^2 - 2 - \beta) - 4\alpha \boldsymbol{r}_i \sum_{l \neq i} \frac{a(\boldsymbol{r}_i - \boldsymbol{r}_l)}{r_{il}^2 (r_{il} - a)} \tag{B7}$$

$$+ \sum_{l \neq i} \sum_{j \neq i} \frac{a^2 (\boldsymbol{r}_i - \boldsymbol{r}_l)(\boldsymbol{r}_i - \boldsymbol{r}_j)}{r_{il}^2 r_{ij}^2 (r_{il} - a)(r_{ij} - a)} + \sum_{l \neq i} \frac{a}{r_{il}^2 (r_{il} - a)} \left[2 + \frac{a - 2 r_{il}}{r_{il} - a} \right] \tag{B8}$$

Appendix C: Simplifying Greens Ratio

We can simplify the ratio in equation (38) by writing Greens function in the following way,

$$G(\boldsymbol{r}_j, \boldsymbol{r}_i, \Delta t) = \frac{1}{(4\pi D\Delta t)^{3N/2}} \exp\left[-\frac{\kappa_i^2}{4D\Delta t}\right] \quad \text{where} \quad \kappa_i = \boldsymbol{r}_j - \boldsymbol{r}_i - D\Delta t F(\boldsymbol{r}_i)$$

The ratio then becomes,

$$\frac{G(\boldsymbol{r}_i, \boldsymbol{r}_j, \Delta t)}{G(\boldsymbol{r}_j, \boldsymbol{r}_i, \Delta t)} = \frac{\exp\left[-\frac{\kappa_j^2}{4D\Delta t}\right]}{\exp\left[-\frac{\kappa_i^2}{4D\Delta t}\right]} = \exp\left[\frac{1}{4D\Delta t}(\kappa_i^2 - \kappa_j^2)\right] = \exp\left[\frac{1}{4D\Delta t}(\kappa_i - \kappa_j)(\kappa_i + \kappa_j)\right]$$

we can calculate

$$\kappa_i - \kappa_j = \boldsymbol{r}_j - \boldsymbol{r}_i - D\Delta t F(\boldsymbol{r}_i) - (\boldsymbol{r}_i - \boldsymbol{r}_j - D\Delta t F(\boldsymbol{r}_j)) = 2(\boldsymbol{r}_j - \boldsymbol{r}_i) - D\Delta t(F(\boldsymbol{r}_i) - F(\boldsymbol{r}_j))$$
$$\kappa_i + \kappa_j = \boldsymbol{r}_j - \boldsymbol{r}_i - D\Delta t F(\boldsymbol{r}_i) + (\boldsymbol{r}_i - \boldsymbol{r}_j - D\Delta t F(\boldsymbol{r}_j)) = -D\Delta t(F(\boldsymbol{r}_i) + F(\boldsymbol{r}_j))$$

So the exponent becomes,

$$\begin{aligned}\frac{1}{4D\Delta t}(\kappa_i - \kappa_j)(\kappa_i + \kappa_j) &= -\frac{1}{4}(F(\boldsymbol{r}_i) + F(\boldsymbol{r}_j))\,(2(\boldsymbol{r}_j - \boldsymbol{r}_i) - D\Delta t(F(\boldsymbol{r}_i) - F(\boldsymbol{r}_j))) \\ &= \frac{1}{2}\left(F(\boldsymbol{r}_i) + F(\boldsymbol{r}_j)\right)\left(\frac{1}{2}D\Delta t(F(\boldsymbol{r}_i) - F(\boldsymbol{r}_j)) - \boldsymbol{r}_j + \boldsymbol{r}_i\right)\end{aligned}$$

And the ratio of Greens functions reduces to,

$$\frac{G(\boldsymbol{r}_i, \boldsymbol{r}_j, \Delta t)}{G(\boldsymbol{r}_j, \boldsymbol{r}_i, \Delta t)} = \exp\left[\frac{1}{2}(F(\boldsymbol{r}_i) + F(\boldsymbol{r}_j))\left(\frac{1}{2}D\Delta t(F(\boldsymbol{r}_i) - F(\boldsymbol{r}_j)) - \boldsymbol{r}_j + \boldsymbol{r}_i\right)\right] \tag{C1}$$

Appendix D: Scaling the Hamiltonian

Length in units of a_{ho}, $r \to r/a_{ho}$, and energy in units of $\hbar\omega_{ho}$

$$V_{ext} = \frac{1}{2}m\left[\omega_{ho}^2(x^2a_{ho}^2 + y^2a_{ho}^2) + \omega_z^2z^2a_{ho}^2\right] = \frac{1}{2}ma_{ho}^2\left[\omega_{ho}^2(x^2+y^2) + \omega_z^2z^2\right] = \frac{1}{2}\frac{\hbar}{\omega_{ho}}\left[\omega_{ho}^2(x^2+y^2) + \omega_z^2z^2\right]$$
$$= \frac{1}{2}\left[\hbar\omega_{ho}(x^2+y^2) + \frac{\hbar\omega_{ho}\omega_z^2}{\omega_{ho}^2}z^2\right] = \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right]\hbar\omega_{ho}$$

where $\gamma = \omega_z/\omega_{ho}$. In units of energy we therefore have,

$$V_{ext} = \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right] \tag{D1}$$

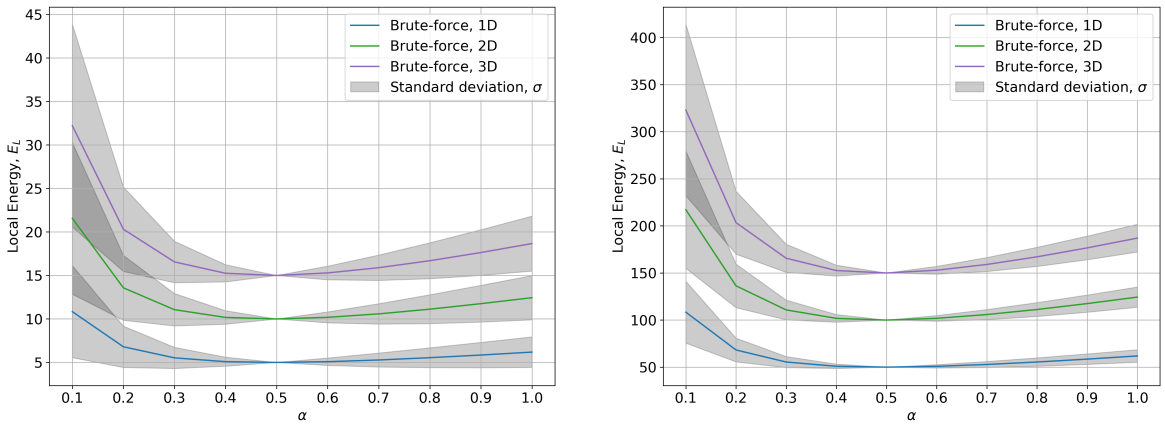
The Laplacian transforms as $\nabla^2 \to \frac{1}{a_{ho}}\nabla^2 = \frac{m\omega_{ho}}{\hbar}$ so the one-body contribution of the hamiltonian becomes,

$$H = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\frac{m\omega_{ho}}{\hbar}\nabla_i^2 + \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right]\hbar\omega_{ho}\right) \tag{D2}$$
$$= \sum_{i=1}^{N}\left(-\frac{1}{2}\hbar\omega_{ho}\nabla_i^2 + \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right]\hbar\omega_{ho}\right) \tag{D3}$$
$$= \sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2 + \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right]\right)\hbar\omega_{ho} \tag{D4}$$

Measuring in units of $\hbar\omega_{ho}$ we then arrive at,

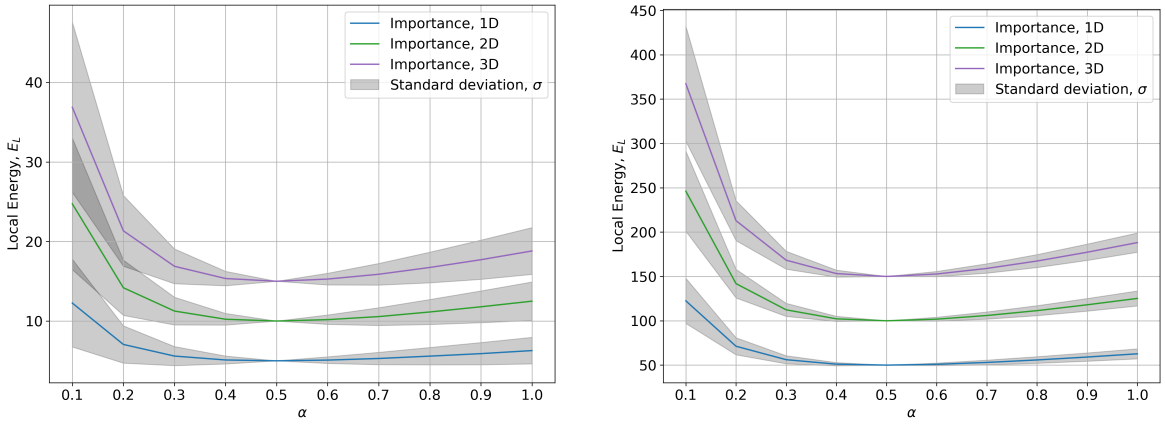
$$H = \sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2 + \frac{1}{2}\left[x^2+y^2+\gamma^2z^2\right]\right) + \sum_{i<j}^{N}V_{int}(|\boldsymbol{r}_i - \boldsymbol{r}_j|) \tag{D5}$$

Appendix E: Additional Figures



(a) $N = 10$ (b) $N = 100$

FIG. 11: The energy as a function of variational parameter α for the brute-force metropolis method on a system of a) 10 particles and b) 100 particles. The step size used is $\Delta r = 0.2$ and the number of MC cycles is 2^{20}.



(a) $N = 10$ (b) $N = 100$

FIG. 12: The energy as a function of variational parameter α for the for the importance sampling method on a system of a) 10 particles and b) 100 particles. The step size used is $\Delta t = 0.01$ and the number of MC cycles is 2^{20}.