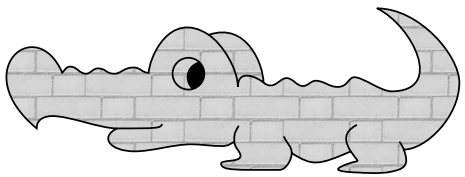
VMC sampling efficiency

Pablo López Ríos

TCM group. Cavendish Laboratory. University of Cambridge.

June 2, 2010



The VMC algorithm

- In VMC we sample configurations $\{\mathbf{R}_1, \ldots, \mathbf{R}_M\}$ distributed according to $|\Psi(\mathbf{R})|^2$
- We evaluate the variational energy as $E_{\mathrm{VMC}} = \frac{1}{M}\sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)$
- This energy has an uncertainty given by $\Delta = \frac{\sigma}{\sqrt{M/n_{\mathrm{corr}}}}$
  - σ^2 is the variance of the sample of local energies, which depends on Ψ
  - n_{corr} is the (integrated) correlation length of the sample of local energies, which depends on how we sample configurations
- A VMC calculation is more efficient the less time it takes to achieve a target errorbar: $\mathscr{E} = \left(\Delta^2 M T_{\mathrm{iter}}\right)^{-1} = \left(\sigma^2 n_{\mathrm{corr}} T_{\mathrm{iter}}\right)^{-1}$
- It is inefficient to attempt to maximize this directly with respect to any parameter due to the multiple evaluations of n_{corr} that this would require

The VMC algorithm

- In VMC we sample configurations $\{\mathbf{R}_1, \ldots, \mathbf{R}_M\}$ distributed according to $|\Psi(\mathbf{R})|^2$
- We evaluate the variational energy as $E_{\mathrm{VMC}} = \frac{1}{M}\sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)$
- This energy has an uncertainty given by $\Delta = \frac{\sigma}{\sqrt{M/n_{\mathrm{corr}}}}$
  - σ^2 is the variance of the sample of local energies, which depends on Ψ
  - n_{corr} is the (integrated) correlation length of the sample of local energies, which depends on how we sample configurations
- A VMC calculation is more efficient the less time it takes to achieve a target errorbar: $\mathscr{E} = \left(\Delta^2 M T_{\mathrm{iter}}\right)^{-1} = \left(\sigma^2 n_{\mathrm{corr}} T_{\mathrm{iter}}\right)^{-1}$
- It is inefficient to attempt to maximize this directly with respect to any parameter due to the multiple evaluations of n_{corr} that this would require

The VMC algorithm

- In VMC we sample configurations $\{\mathbf{R}_1, \ldots, \mathbf{R}_M\}$ distributed according to $|\Psi(\mathbf{R})|^2$
- We evaluate the variational energy as $E_{\mathrm{VMC}} = \frac{1}{M} \sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)$
- This energy has an uncertainty given by $\Delta = \frac{\sigma}{\sqrt{M/n_{\mathrm{corr}}}}$
  - σ^2 is the variance of the sample of local energies, which depends on Ψ
  - n_{corr} is the (integrated) correlation length of the sample of local energies, which depends on how we sample configurations
- A VMC calculation is more efficient the less time it takes to achieve a target errorbar: $\mathscr{E} = \left(\Delta^2 M T_{\mathrm{iter}}\right)^{-1} = \left(\sigma^2 n_{\mathrm{corr}} T_{\mathrm{iter}}\right)^{-1}$
- It is inefficient to attempt to maximize this directly with respect to any parameter due to the multiple evaluations of n_{corr} that this would require

The VMC algorithm

- In VMC we sample configurations $\{\mathbf{R}_1, \ldots, \mathbf{R}_M\}$ distributed according to $|\Psi(\mathbf{R})|^2$
- We evaluate the variational energy as $E_{\mathrm{VMC}} = \frac{1}{M} \sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)$
- This energy has an uncertainty given by $\Delta = \frac{\sigma}{\sqrt{M/n_{\mathrm{corr}}}}$
  - σ^2 is the variance of the sample of local energies, which depends on Ψ
  - n_{corr} is the (integrated) correlation length of the sample of local energies, which depends on how we sample configurations
- A VMC calculation is more efficient the less time it takes to achieve a target errorbar: $\mathscr{E} = \left(\Delta^2 M T_{\mathrm{iter}}\right)^{-1} = \left(\sigma^2 n_{\mathrm{corr}} T_{\mathrm{iter}}\right)^{-1}$
- It is inefficient to attempt to maximize this directly with respect to any parameter due to the multiple evaluations of n_{corr} that this would require

The VMC algorithm

- In VMC we sample configurations $\{\mathbf{R}_1, \ldots, \mathbf{R}_M\}$ distributed according to $|\Psi(\mathbf{R})|^2$
- We evaluate the variational energy as $E_{\mathrm{VMC}} = \frac{1}{M} \sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)$
- This energy has an uncertainty given by $\Delta = \frac{\sigma}{\sqrt{M/n_{\mathrm{corr}}}}$
  - σ^2 is the variance of the sample of local energies, which depends on Ψ
  - n_{corr} is the (integrated) correlation length of the sample of local energies, which depends on how we sample configurations
- A VMC calculation is more efficient the less time it takes to achieve a target errorbar: $\mathscr{E} = \left(\Delta^2 M T_{\mathrm{iter}}\right)^{-1} = \left(\sigma^2 n_{\mathrm{corr}} T_{\mathrm{iter}}\right)^{-1}$
- It is inefficient to attempt to maximize this directly with respect to any parameter due to the multiple evaluations of n_{corr} that this would require

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{|\Psi(\mathbf{R}'_m)|^2}{|\Psi(\mathbf{R}_m)|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{|\Psi(\mathbf{R}'_m)|^2}{|\Psi(\mathbf{R}_m)|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{\left|\Psi(\mathbf{R}'_m)\right|^2}{\left|\Psi(\mathbf{R}_m)\right|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{\left|\Psi(\mathbf{R}'_m)\right|^2}{\left|\Psi(\mathbf{R}_m)\right|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{\left|\Psi(\mathbf{R}'_m)\right|^2}{\left|\Psi(\mathbf{R}_m)\right|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

VMC sampling

- $\{\mathbf{R}_m\}_{m=1,\ldots,M}$ are generated using the Metropolis algorithm:
  - Propose move from $\mathbf{R}_m$ to $\mathbf{R}'_m$ with probability $T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$
  - Compute $A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) = \min\left(1, \frac{T(\mathbf{R}_m \leftarrow \mathbf{R}'_m)}{T(\mathbf{R}'_m \leftarrow \mathbf{R}_m)} \frac{\left|\Psi(\mathbf{R}'_m)\right|^2}{\left|\Psi(\mathbf{R}_m)\right|^2}\right)$
  - Draw random number $0 < \zeta < 1$ from a uniform distribution, and
    - If $\zeta < A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)$, make $\mathbf{R}_{m+1} = \mathbf{R}'_m$ (accept move)
    - Otherwise, set $\mathbf{R}_{m+1} = \mathbf{R}_m$ (reject move)
- To achieve reasonable acceptance ratios, proposed configurations are the original plus a normally-distributed random displacement of variance τ
- This causes serial correlation ($n_{\text{corr}} > 1$)

Electron-by-electron sampling

- It is possible to use a variation of the Metropolis algorithm where one proposes single-electron moves and accepts or rejects them individually
- Advantage: larger steps can be taken with high acceptance ratios, thus reducing n_{corr}
- Disadvantage: the evaluation of N single-electron wave-function ratios is more expensive than that of one all-electron wave function ratio, and especially for complicated functional forms (e.g., Slater determinants with backflow transformations), which increases T_{iter}

Electron-by-electron sampling

- It is possible to use a variation of the Metropolis algorithm where one proposes single-electron moves and accepts or rejects them individually
- Advantage: larger steps can be taken with high acceptance ratios, thus reducing n_{corr}
- Disadvantage: the evaluation of N single-electron wave-function ratios is more expensive than that of one all-electron wave function ratio, and especially for complicated functional forms (e.g., Slater determinants with backflow transformations), which increases T_{iter}

Electron-by-electron sampling

- It is possible to use a variation of the Metropolis algorithm where one proposes single-electron moves and accepts or rejects them individually
- Advantage: larger steps can be taken with high acceptance ratios, thus reducing n_{corr}
- Disadvantage: the evaluation of N single-electron wave-function ratios is more expensive than that of one all-electron wave function ratio, and especially for complicated functional forms (e.g., Slater determinants with backflow transformations), which increases T_{iter}

Decorrelation loops

- One can perform $p > 1$ Metropolis steps between evaluations of the local energy
- Advantage: n_{corr} decreases
- Disadvantage: the extra moves increase T_{iter}

Decorrelation loops

- One can perform $p > 1$ Metropolis steps between evaluations of the local energy
- Advantage: n_{corr} decreases
- Disadvantage: the extra moves increase T_{iter}

Decorrelation loops

- One can perform $p > 1$ Metropolis steps between evaluations of the local energy
- Advantage: n_{corr} decreases
- Disadvantage: the extra moves increase T_{iter}

Averaging successive local energies

- The mth local energy can be replaced by the average $[1 - A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)]\, E_L(\mathrm{R}_m) + A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) E_L(\mathrm{R}'_m)$
- Advantage: more statistics, especially important at low acceptance ratios, potentially reducing n_{corr}
- Disadvantage: needs more energy evaluations, increasing T_{iter}
- This has proved inefficient in electron-by-electron sampling, so will only test in configuration-by-configuration sampling

Averaging successive local energies

- The mth local energy can be replaced by the average $[1 - A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)]\, E_L(\mathrm{R}_m) + A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) E_L(\mathrm{R}'_m)$
- Advantage: more statistics, especially important at low acceptance ratios, potentially reducing n_{corr}
- Disadvantage: needs more energy evaluations, increasing T_{iter}
- This has proved inefficient in electron-by-electron sampling, so will only test in configuration-by-configuration sampling

Averaging successive local energies

- The mth local energy can be replaced by the average $[1 - A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)]\, E_L(\mathrm{R}_m) + A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) E_L(\mathrm{R}'_m)$
- Advantage: more statistics, especially important at low acceptance ratios, potentially reducing n_{corr}
- Disadvantage: needs more energy evaluations, increasing T_{iter}
- This has proved inefficient in electron-by-electron sampling, so will only test in configuration-by-configuration sampling

Averaging successive local energies

- The mth local energy can be replaced by the average $[1 - A(\mathbf{R}'_m \leftarrow \mathbf{R}_m)]\, E_L(\mathrm{R}_m) + A(\mathbf{R}'_m \leftarrow \mathbf{R}_m) E_L(\mathrm{R}'_m)$
- Advantage: more statistics, especially important at low acceptance ratios, potentially reducing n_{corr}
- Disadvantage: needs more energy evaluations, increasing T_{iter}
- This has proved inefficient in electron-by-electron sampling, so will only test in configuration-by-configuration sampling

Things to look into

- Optimal value of τ?
- Electron-by-electron versus configuration-by-configuration - which to use when?
- Decorrelation loops - optimal length?
- Is averaging energies over proposed configurations useful?

Things to look into

- Optimal value of τ?
- Electron-by-electron versus configuration-by-configuration - which to use when?
- Decorrelation loops - optimal length?
- Is averaging energies over proposed configurations useful?

Things to look into

- Optimal value of τ?
- Electron-by-electron versus configuration-by-configuration - which to use when?
- Decorrelation loops - optimal length?
- Is averaging energies over proposed configurations useful?

Things to look into

- Optimal value of τ?
- Electron-by-electron versus configuration-by-configuration - which to use when?
- Decorrelation loops - optimal length?
- Is averaging energies over proposed configurations useful?

Methodology

- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

Methodology

- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

Methodology

- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

Methodology

- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

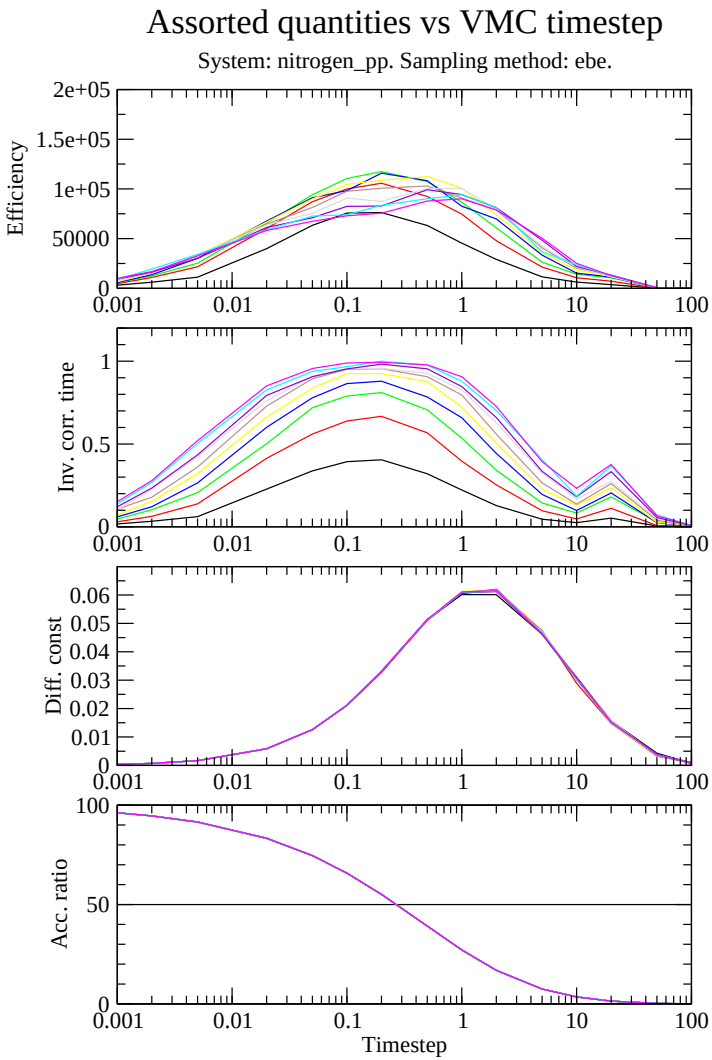
Methodology

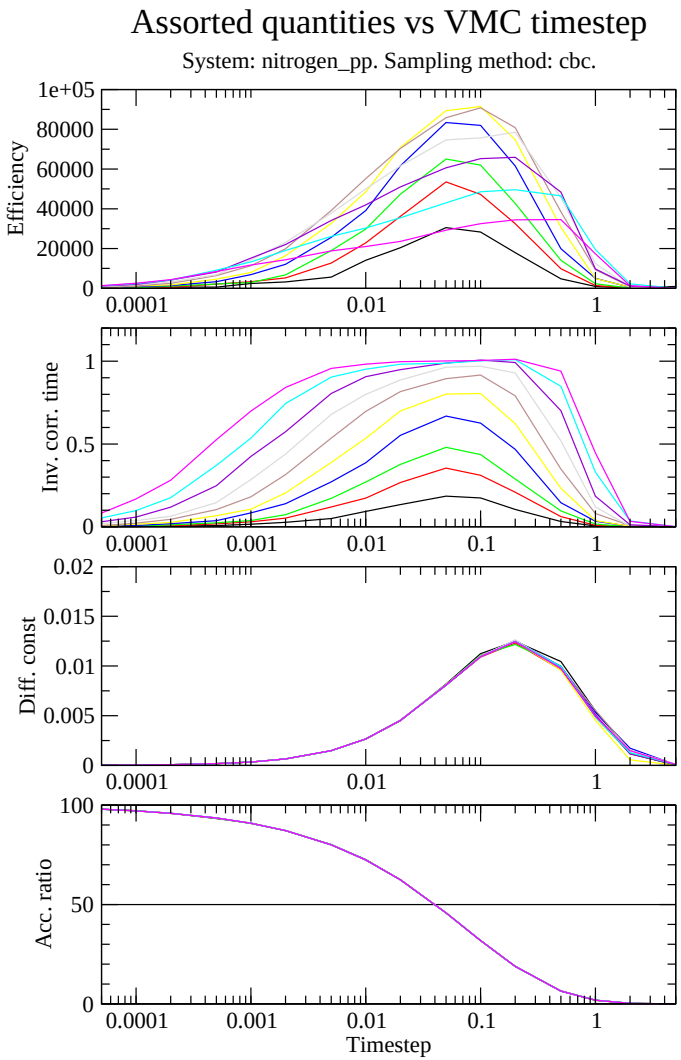
- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

Methodology

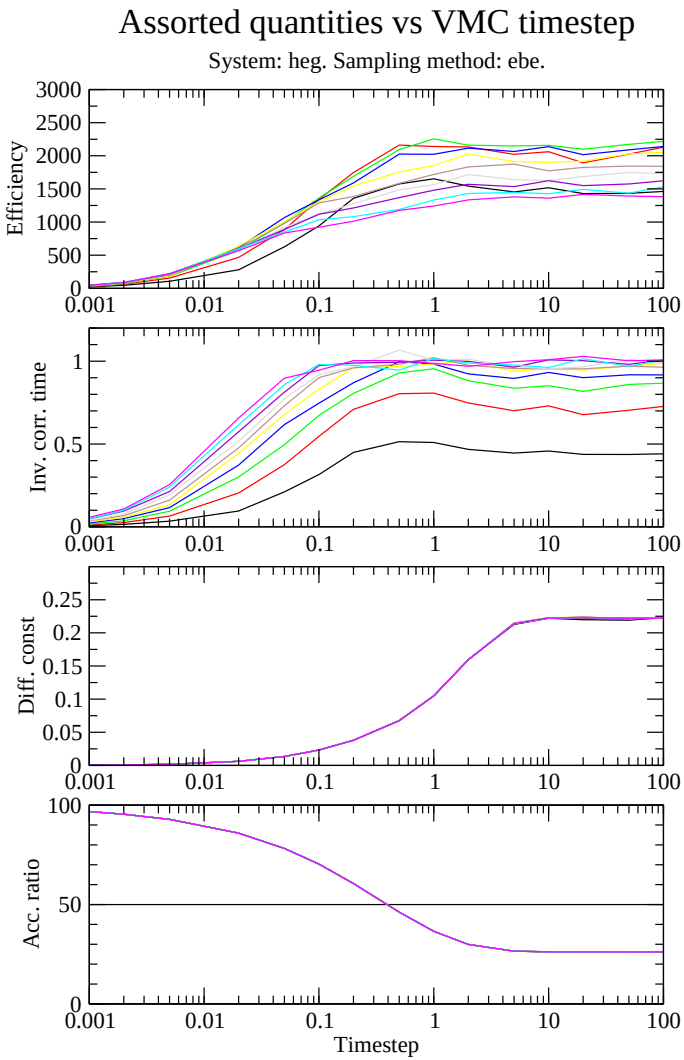
- Choose 6 relevant systems of different sizes
- Run short (but significant) VMC calculations spanning 16 values of τ and 10 values of p
- Run electron-by-electron and configuration-by-configuration versions of the above, the latter with and without averaging over successive energies
- Use Slater-Jastrow and Slater-Jastrow-backflow wave function forms
- Total: 5760 runs
- Use the data to locate maximum efficiency for each case, compare, analyze, etc

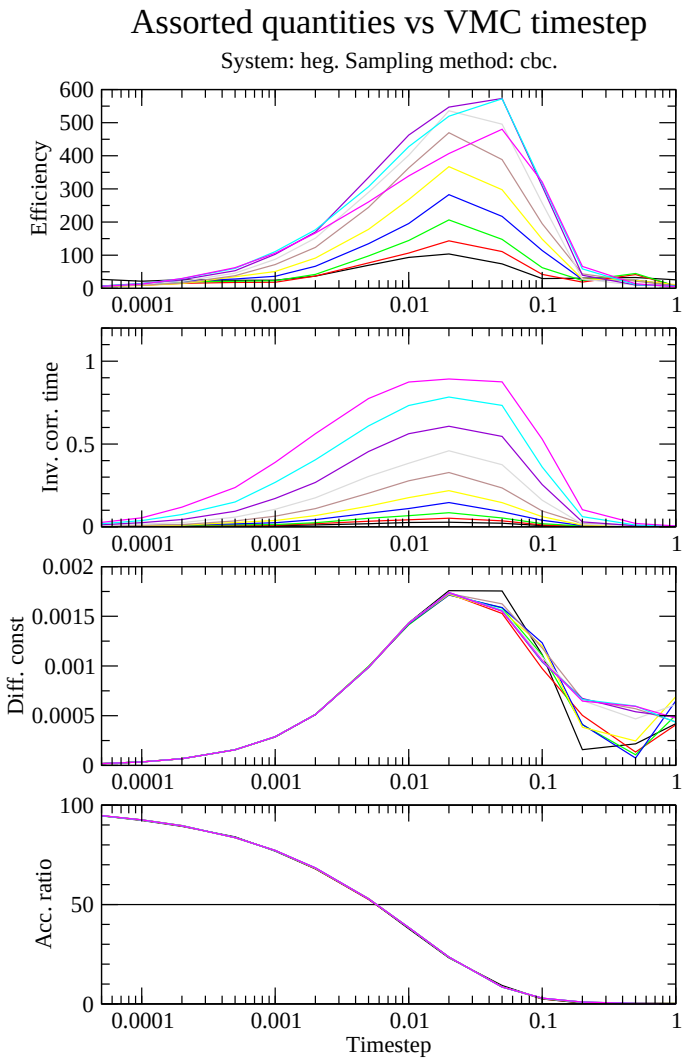
Pseudo Nitrogen atom, Slater-Jastrow, EBES vs CBCS



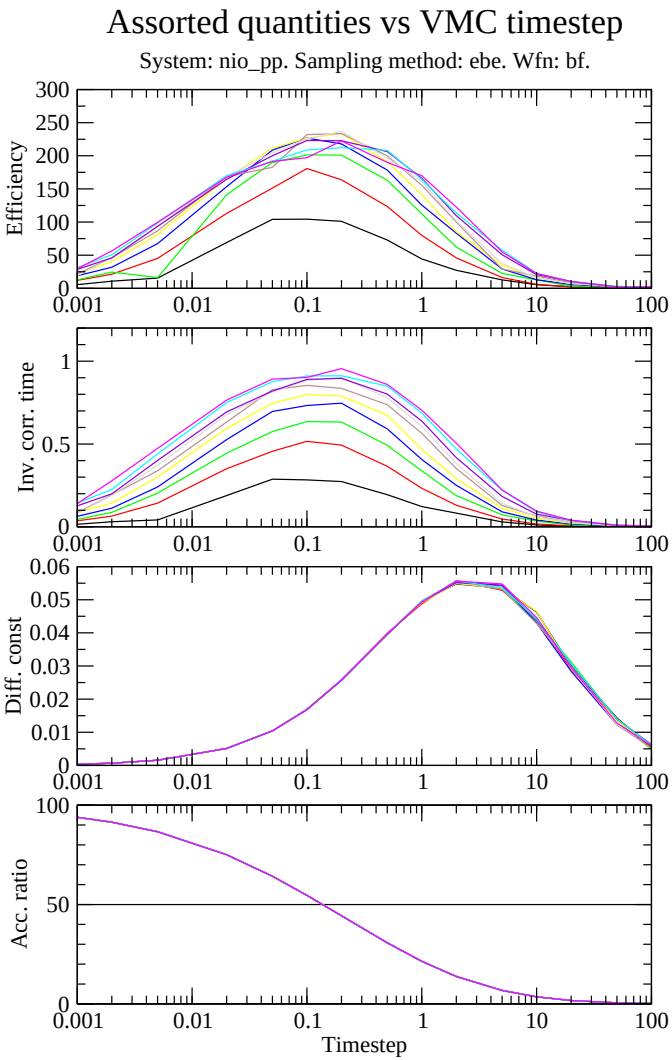


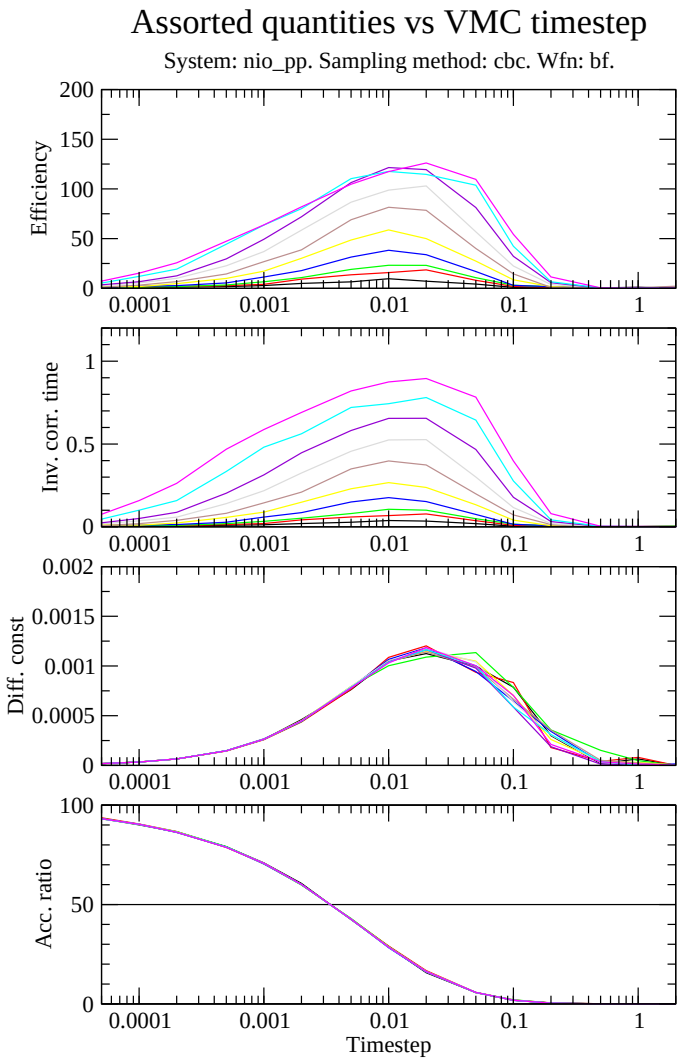
HEG, Slater-Jastrow, EBES vs CBCS



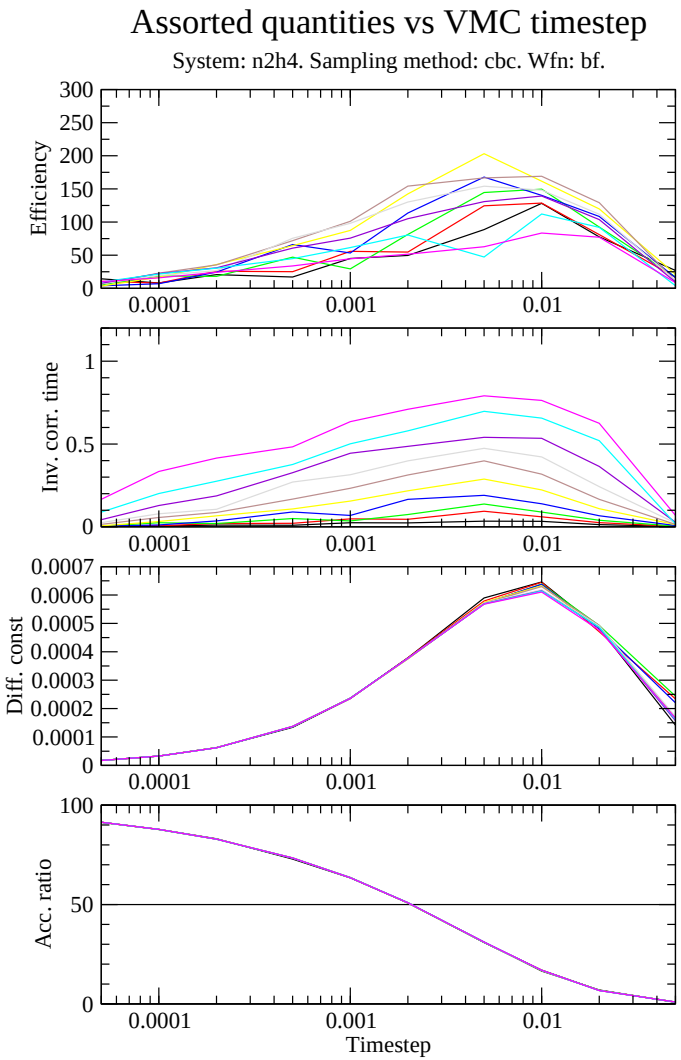


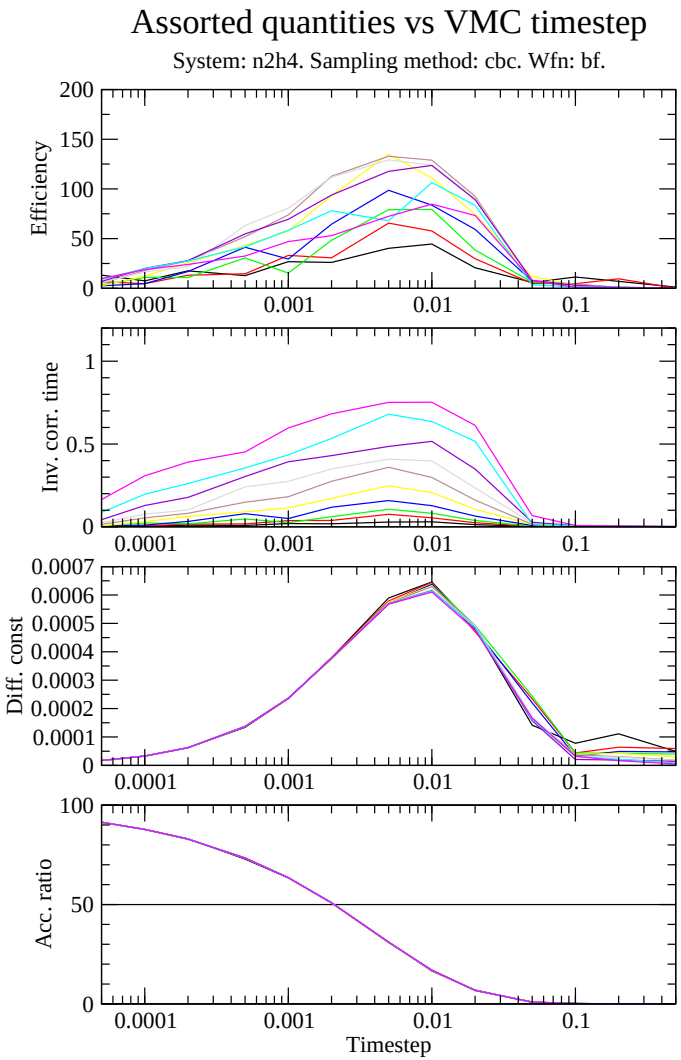
Pseudo NiO molecule, backflow, EBES vs CBCS





All-electron N_2H_4, backflow, CBCS vs CBCS2





Functional form of $\mathscr{E}(p)$

- Cost of one energy evaluation: $T_{\text{iter}}(p) = pT_{\text{move}} + T_{\text{energy}}$
- Assuming $M \to \infty$, and that the autocorrelation of the local energies is dominated by a single exponential, $n_{\text{corr}}(p) = 1 + 2\frac{(n_{\text{corr}}-1)^p}{(n_{\text{corr}}+1)^p - (n_{\text{corr}}-1)^p}$
- One can minimize $T_{\text{iter}}(p)n_{\text{corr}}(p)$ numerically if n_{corr} and $T_{\text{energy}}/T_{\text{move}}$ are know.

Functional form of $\mathscr{E}(p)$

- Cost of one energy evaluation: $T_{\text{iter}}(p) = pT_{\text{move}} + T_{\text{energy}}$
- Assuming $M \to \infty$, and that the autocorrelation of the local energies is dominated by a single exponential, $n_{\text{corr}}(p) = 1 + 2\frac{(n_{\text{corr}}-1)^p}{(n_{\text{corr}}+1)^p - (n_{\text{corr}}-1)^p}$
- One can minimize $T_{\text{iter}}(p)n_{\text{corr}}(p)$ numerically if n_{corr} and $T_{\text{energy}}/T_{\text{move}}$ are know.

Functional form of $\mathscr{E}(p)$

- Cost of one energy evaluation: $T_{\text{iter}}(p) = pT_{\text{move}} + T_{\text{energy}}$
- Assuming $M \to \infty$, and that the autocorrelation of the local energies is dominated by a single exponential, $n_{\text{corr}}(p) = 1 + 2\frac{(n_{\text{corr}}-1)^p}{(n_{\text{corr}}+1)^p - (n_{\text{corr}}-1)^p}$
- One can minimize $T_{\text{iter}}(p)n_{\text{corr}}(p)$ numerically if n_{corr} and $T_{\text{energy}}/T_{\text{move}}$ are know.

Summary and recommendations

- Use electron-by-electron sampling
- Optimize τ so as to achieve a 50% acceptance ratio
- Set p to 3-5, or compute n_{corr} from a short run and maximize $\mathscr{E}$ numerically
- Do not average over successive energies
- We've been doing it right all along!

Summary and recommendations

- Use electron-by-electron sampling
- Optimize τ so as to achieve a 50% acceptance ratio
- Set p to 3-5, or compute n_{corr} from a short run and maximize $\mathscr{E}$ numerically
- Do not average over successive energies
- We've been doing it right all along!