Accelerating Full Configuration Interaction Quantum Monte Carlo

William Vigor [1], James Spencer [2,3], Michael Bearpark [1], Alex Thom [1,4]

[1]Department of Chemistry, Imperial College London
[2]Department of Physics, Imperial College London
[3]Department of Materials, Imperial College London
[4]Department of Chemistry, University of Cambridge

Part I

Introduction: The Full Configuration Interaction Quantum Monte Carlo Algorithm

Solving the Schrödinger Equation in a Basis of Slater Determinants

- To estimate the ground state energy of a Molecule with N electrons solve:

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{1}$$

  where $|\Psi\rangle$ is in some finite basis (at fixed geometry).
- If we let $|\Psi\rangle$ be a single determinant of M Molecular Spin Orbitals (MSO), Hartree-Fock lets us compute the variations minimal single determinant wavefunction $|D_{\mathbf{0}}\rangle$.
- Full Configuration Interaction (FCI) adds into $|\Psi\rangle$ all possible determinants $|D_{\mathbf{I}}\rangle$ with N MSO occupied and $M - N$ unoccupied .

Full Configuration Interaction

- Want to solve the eigenvalue problem $H|\Psi\rangle = e|\Psi\rangle$ where $|\Psi\rangle = \sum_I C_{\mathbf{I}}|D_{\mathbf{I}}\rangle$
- Cast in to a matrix problem diagonalise matrix of $\langle D_{\mathbf{I}}|\hat{H}|D_{\mathbf{J}}\rangle$
- Unfortunately this turns out to be impossibly large: $\binom{M}{N}$ by $\binom{M}{N}$.
- **Can use Monte Carlo to sample this space.**

G. H. Booth, A. J. W. Thom, A. Alavi, J. Chem. Phys. **131**, 054106 (2009)

Full Configuration Interaction Quantum Monte Carlo (FCIQMC) vs FCI

- Little communication overhead, can use massively parallel computers.
- Only need to store a stochastic representation of the eigenvector.
- Stochastic, quantifiable error (system and dynamics dependent) can be reduced to FCI accuracy by running for longer.
- **How quickly does this error converge?**
- **Can we find most efficient FCIQMC algorithm? i.e. how can we make the stochastic error converge quickest as a function of computer time?**

Full Configuration Interaction Quantum Monte Carlo: Projector

- Denote the exact solutions (in a finite basis):

$$|\Psi_0\rangle, |\Psi_1\rangle, \ldots |\Psi_N\rangle \tag{2}$$

with energies:

$$E_0, E_1, \ldots, E_N \tag{3}$$

We can expand our wavefunction in this basis:

$$|\psi\rangle = \sum_i^N a_i(\tau)|\Psi_i\rangle \tag{4}$$

If we apply the diffusion equation.

$$\frac{\partial|\psi\rangle}{\partial\tau} = -(\hat{H} - E_r\hat{1})|\psi\rangle \tag{5}$$

and take $\langle\Psi_j|$:

$$\frac{\partial a_j(\tau)}{\partial\tau} = -(E_j - E_r)a_j(\tau) \tag{6}$$

Then:

$$a_j(\tau) = e^{-(E_j-E_r)\tau}a_j(\tau=0) \tag{7}$$

Excited states die as $\tau \to \infty$ and if $E_r < E_0$ the $|\Psi_0\rangle$ contribution grows.

Full Configuration Interaction Quantum Monte Carlo: Discretised Determinant Space

First discretise our determinant space:

- Positive coefficients are represented by a number of positive psips (Ψ particles).
- Negative coefficients are represented by a number of negative psips.

$$|\Psi(\tau)\rangle = \quad +0.25 \quad |D_{\mathbf{0}}\rangle \quad +0.5 \quad |D_{\mathbf{1}}\rangle \quad -0.25 \quad |D_{\mathbf{2}}\rangle$$

J . B. Anderson J. Chem. Phys. **63**, 1499 (1975)

Full Configuration Interaction Quantum Monte Carlo: Discretised Determinant Space

First discretise our determinant space:

- Positive coefficients are represented by a number of positive psips (Ψ particles).
- Negative coefficients are represented by a number of negative psips.

$|\Psi(\tau)\rangle =$  $|D_0\rangle +$  $|D_1\rangle +$  $|D_2\rangle$

J . B. Anderson J. Chem. Phys. **63**, 1499 (1975)

Full Configuration Interaction Quantum Monte Carlo

- Can move τ forwards by $\delta\tau$ by applying:

$$e^{-(\hat{H}-E_r\hat{1})\delta\tau} \approx (\hat{1} - (\hat{H} - E_r\hat{1})\delta\tau) \tag{8}$$

- FCIQMC stochastically applies Eq. 8 in 3 steps, on each iteration.
  1. Spawn
  2. Diagonal Death
  3. Annihilation
- These define the psip population dynamics of FCIQMC.
- Thus the ground state wavefunction comes out if we start with a ψ such that $\langle\psi|\hat{H}|\Psi\rangle \neq 0$ and run for many steps.

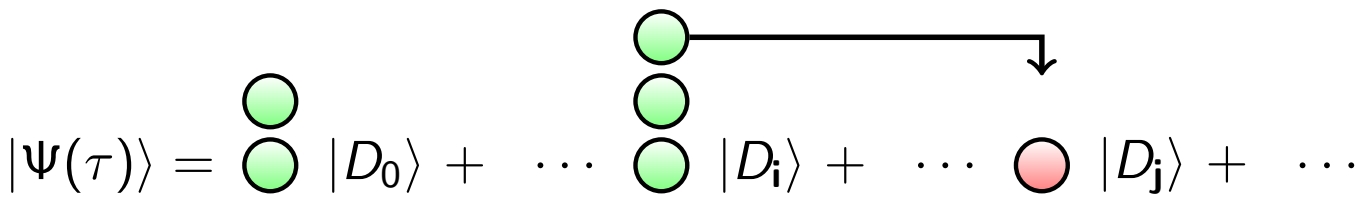
FCIQMC Population Dynamics 1. Spawning

- Each psip attempts to spawn a child psip on a randomly selected determinant.

$$|\Psi(\tau)\rangle = \text{●} |D_0\rangle + \cdots \begin{matrix}\text{●}\\\text{●}\\\text{●}\end{matrix} |D_{\mathbf{i}}\rangle + \cdots \text{●} |D_{\mathbf{j}}\rangle + \cdots$$

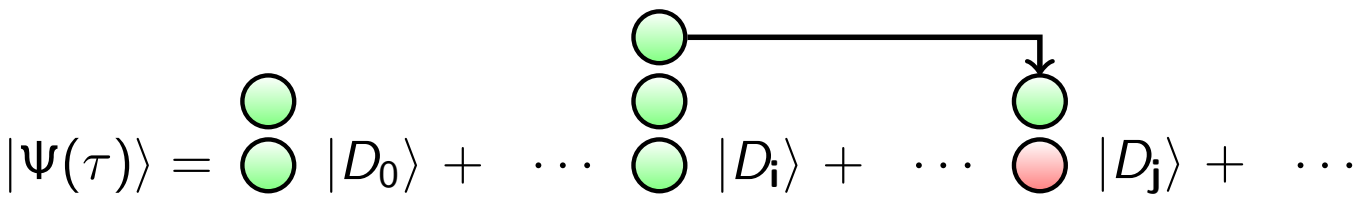
FCIQMC Population Dynamics 1. Spawning

- Each psip attempts to spawn a child psip on a randomly selected determinant.



FCIQMC Population Dynamics 1. Spawning

- Each psip attempts to spawn a child psip on a randomly selected determinant.

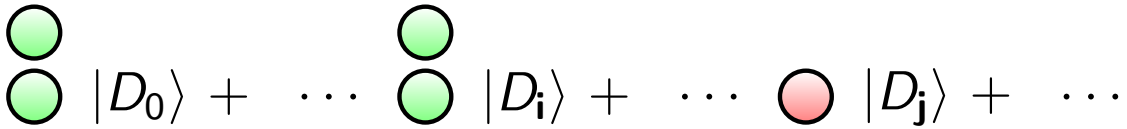


- with probability $(\langle D_\mathbf{i}|H|D_\mathbf{j}\rangle)\delta\tau$
- if $(\langle D_\mathbf{i}|H|D_\mathbf{j}\rangle)\delta\tau < 0$ psip has same sign as parent and vice-versa.

FCIQMC Population Dynamics 2. Diagonal Death

- Each parent psip attempts to die or is cloned.

$$|\Psi(\tau)\rangle = \quad |D_0\rangle + \quad \cdots \quad |D_{\mathbf{i}}\rangle + \quad \cdots \quad |D_{\mathbf{j}}\rangle + \quad \cdots$$



FCIQMC Population Dynamics 2. Diagonal Death

- Each parent psip attempts to die or is cloned.

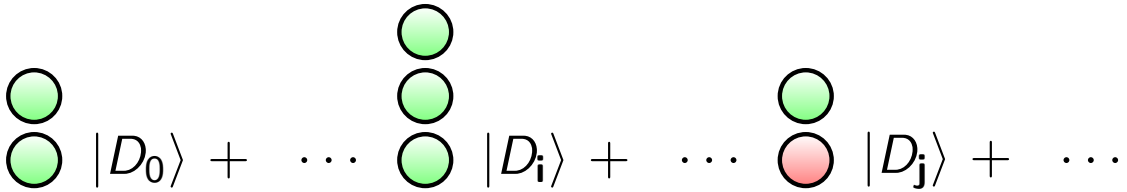
$$|\Psi(\tau)\rangle = \; \bigcirc\!\!\bigcirc \; |D_0\rangle + \; \cdots \; \bigcirc\!\!\bigcirc \; |D_{\mathbf{i}}\rangle + \; \cdots \; \bigcirc \; |D_{\mathbf{j}}\rangle + \; \cdots$$

- Death occurs with probability $(\langle D_{\mathbf{i}}|H|D_{\mathbf{i}}\rangle - E_r)\delta\tau$
- Cloning occurs (population becomes more negative or positive) if $(\langle D_{\mathbf{i}}|H|D_{\mathbf{i}}\rangle - E_r) > 0$
- E_r is initially set to the energy of the Hartree-Fock until the psip population reaches the desired level.
- After which E_r is periodically updated to keep the population at the desired level.

FCIQMC Population Dynamics 3. Annihilation

- Positive and negative psips residing on the same determinant annihilate:

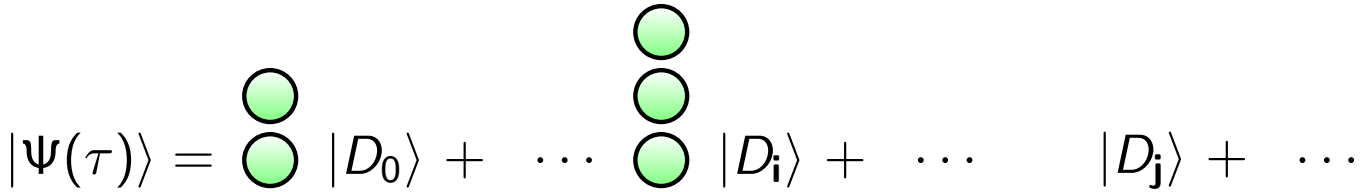
$$|\Psi(\tau)\rangle = \quad |D_0\rangle + \quad \cdots \quad |D_{\mathbf{i}}\rangle + \quad \cdots \quad |D_{\mathbf{j}}\rangle + \quad \cdots$$



J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes J. Chem. Phys. **136**, 054110 (2012)

FCIQMC Population Dynamics 3. Annihilation

- Positive and negative psips residing on the same determinant annihilate:



J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes J. Chem. Phys. **136**, 054110 (2012)

FCIQMC Population Dynamics 3. Annihilation

- Positive and negative psips residing on the same determinant annihilate:

$$|\Psi(\tau)\rangle = \;  \; |D_0\rangle + \cdots \;  \; |D_{\mathbf{i}}\rangle + \cdots \qquad |D_{\mathbf{j}}\rangle + \cdots$$

- Essential so that the step-wise average of psip vector is the eigenvector with smallest eigenvalue of the FCI matrix.
- If we use enough psips.

J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes J. Chem. Phys. **136**, 054110 (2012)

Accumulating Expectation Values

- Projected energy:

$$E = \frac{\langle D_0|\hat{H}e^{\hat{H}\tau}|D_0\rangle}{\langle D_0|e^{\hat{H}\tau}|D_0\rangle} = \frac{\langle D_0|\hat{H}|\Psi_0\rangle}{\langle D_0||\Psi_0\rangle} \tag{9}$$

- $\langle D_0|\hat{H}e^{\hat{H}\tau}|D_0\rangle$ equals a factor of:
  - $\langle D_\mathbf{0}|H|D_\mathbf{i}\rangle$
  - $-\langle D_\mathbf{0}|H|D_\mathbf{i}\rangle$

  sum over every psip, sum over every determinant.
- $\langle D_0||\Psi_0\rangle$ equals the number of psips on the Hartree–Fock.
- Average over every step after the simulation has equilibrated.
- Need to estimate errors carefully due to serial correlation. As $\delta\tau$ is small the stochastic representation of the eigenvector changes only a small amount between iterations, correlating estimates close in iteration space.

The Initiator Approximation

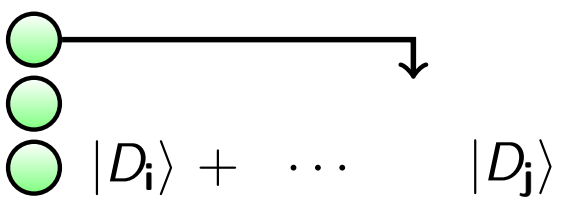
- Allow only psips on determinants with a population above a threshold (initiator determinants) to spawn onto unoccupied determinants.

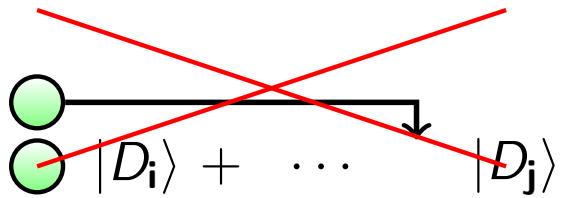
- Error bars converge faster and we need fewer psips.

D. Cleland, G. H. Booth, A. Alavi J. Chem. Phys. **132**, 041103 (2010)

The Initiator Approximation

- Allow only psips on determinants with a population above a threshold (initiator determinants) to spawn onto unoccupied determinants.





- Error bars converge faster and we need fewer psips.

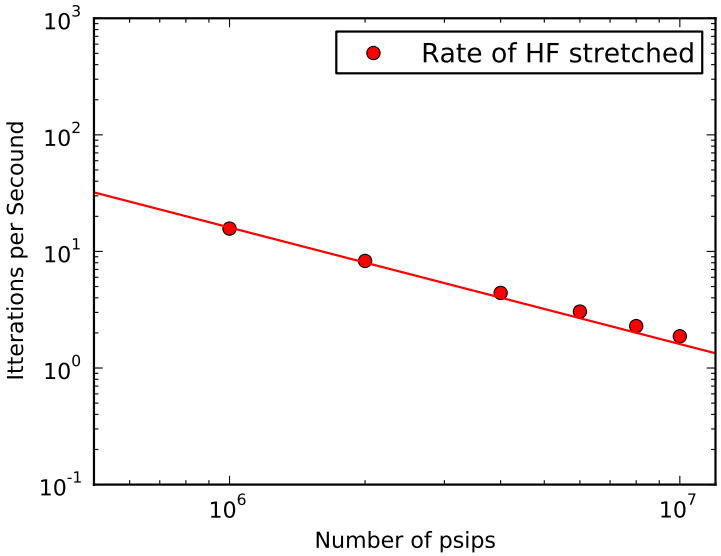
D. Cleland, G. H. Booth, A. Alavi J. Chem. Phys. **132**, 041103 (2010)

Part II

How can we use current algorithms most effectively

Computer time as a function of the number of psips

- In FCIQMC more psips equivalent to more steps in terms of computer time.
- $\mathcal{O}(l \log l)$ sorting of newly spawned psips before of annihilation is negligible (l length of list to sort).
- Fits to $\frac{c}{N_p}$ shown (N_p number of psips).



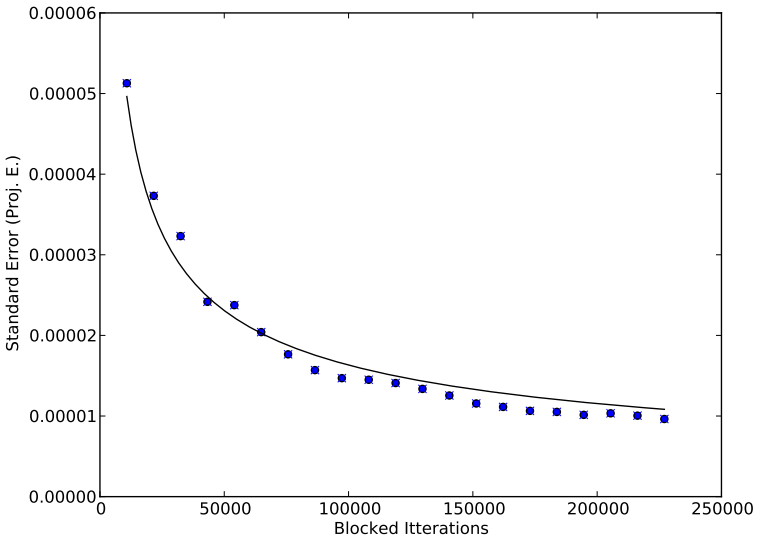
- What about the error bar?
- Is it better to run for more steps or use more psips (which is a better use of resources)?

Quantifying Efficiency Empirically 1

- For any Monte Carlo algorithm the stochastic error (in some expectation value of the simulation) as a function of the number of steps N :

$$\sigma = \frac{a}{\sqrt{N}} \qquad (10)$$

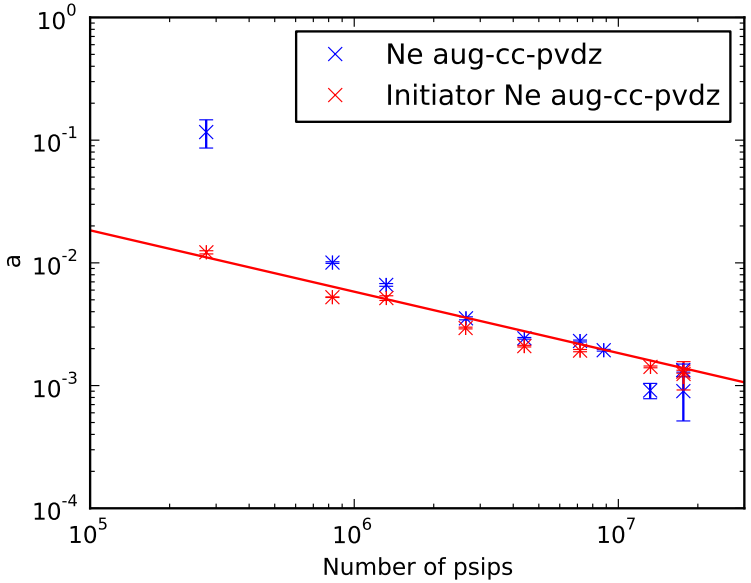
- So can use a to quantify the efficiency of FCIQMC, provided no systematic error is introduced.
- Small a is good, errors converge fast.



Empirical Results: Scaling with the number of psips 1

$$\sigma = \frac{a}{\sqrt{N_s}} \qquad (11)$$

- How does a depends on the number of psips N_p in the simulation (N_s number of iterations).
- One should fill the memory with psips if $a \to 0$ faster than $\mathcal{O}(\frac{1}{\sqrt{N_p}})$.

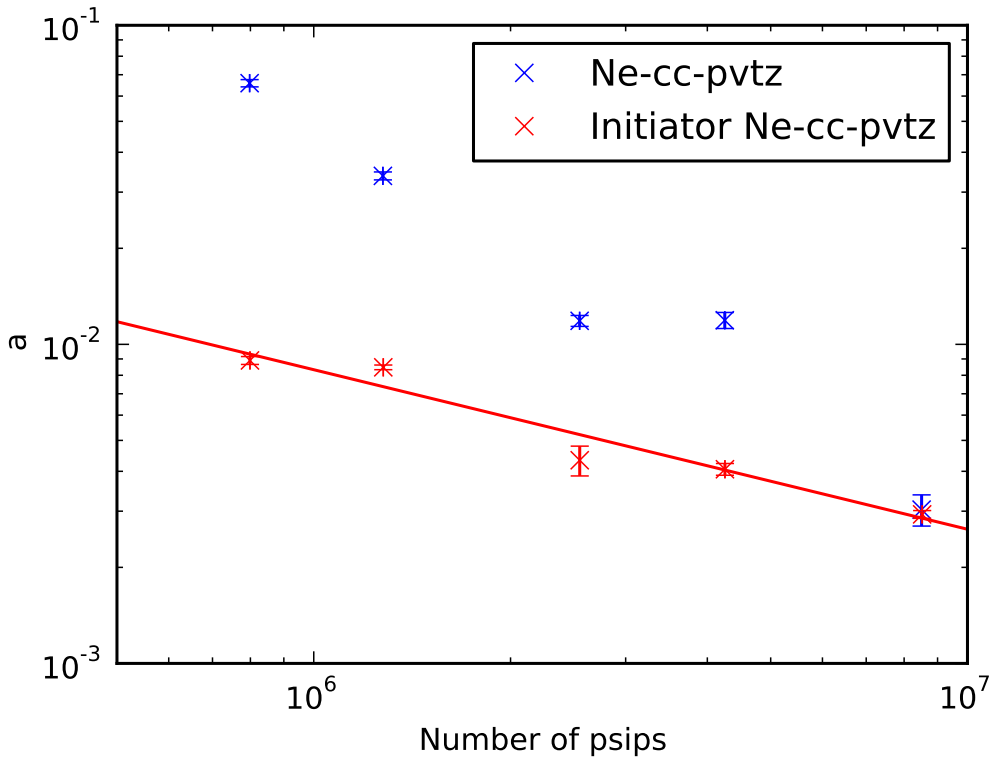


- For FCIQMC a decays faster than $\mathcal{O}(\frac{1}{\sqrt{N_p}})$, for i-FCIQMC seems to decay as $b\sqrt{N_p}$ meaning more psips equivalent to more iterations.

Empirical Results: Scaling with the number of psips 2

$$\sigma = \frac{a}{\sqrt{N_s}} \tag{12}$$

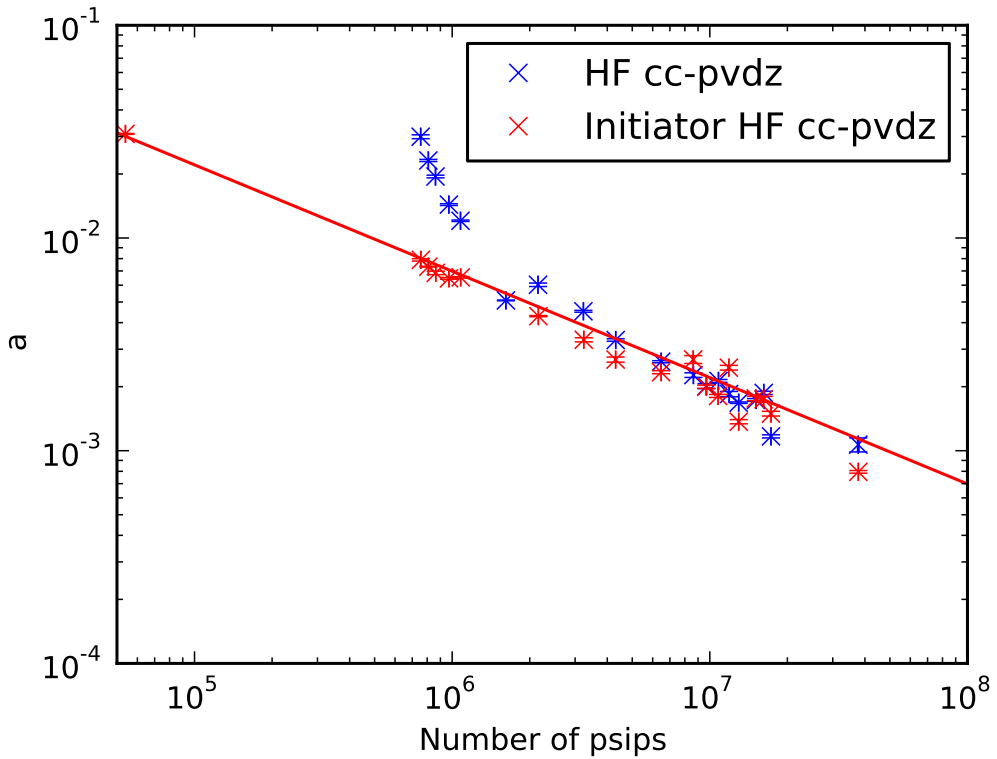
- Other systems show similar results.



Empirical Results: Scaling with the number of psips 3

$$\sigma = \frac{a}{\sqrt{N_s}} \tag{13}$$

- Other systems show similar results.



Conclusions

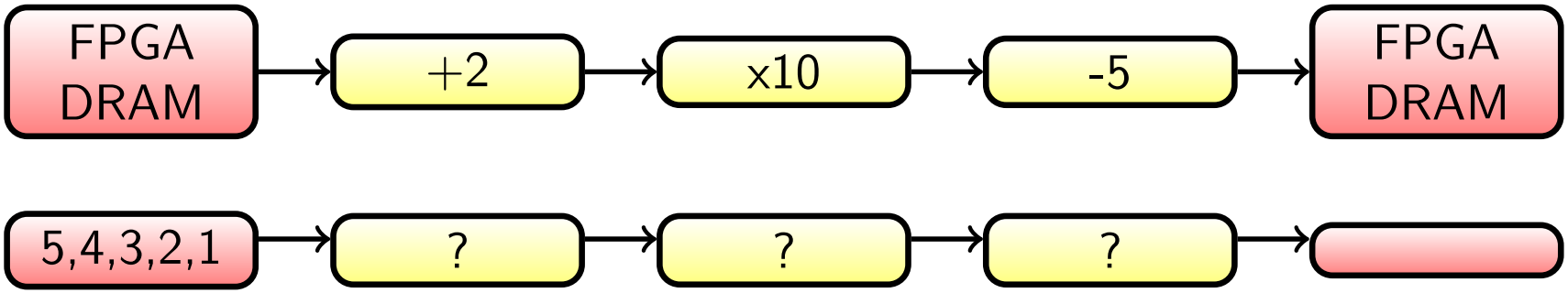
- For calculations with the initiator approximation more psips is equivalent to more steps in terms of $\sigma(E)$.
- For other systems the error decreases faster as the number of psips increase getting closer to the initiator limit.
- At the limit of a large number of psips, trivial parallelism mode (running multiple independent simulations).
- We now have a way of comparing different Monte Carlo algorithms.

Part III

Novel Hardware

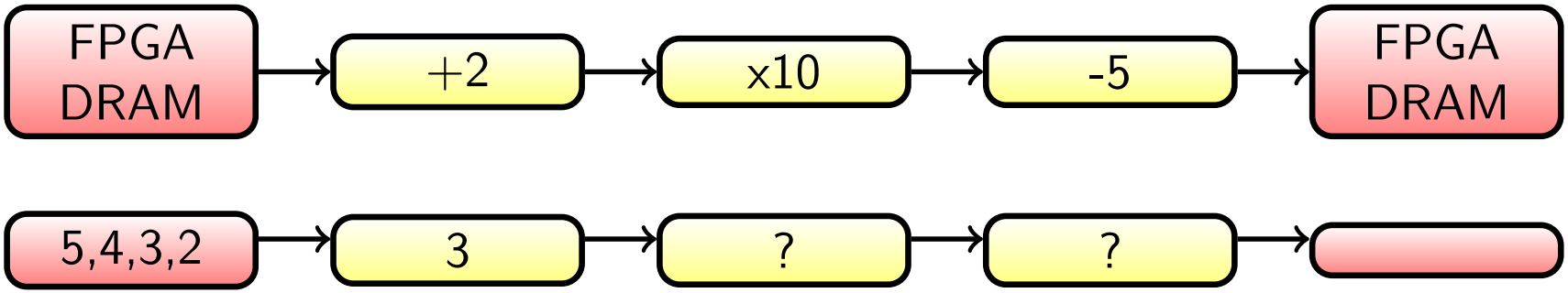
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



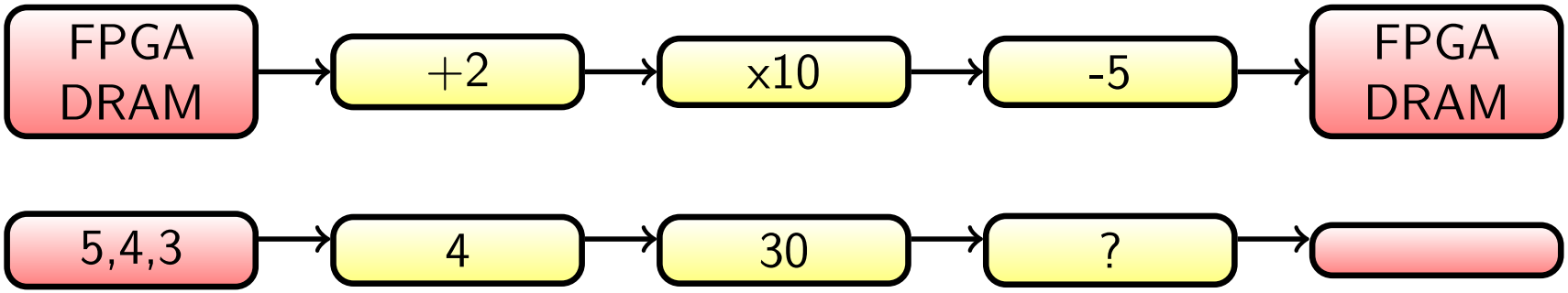
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



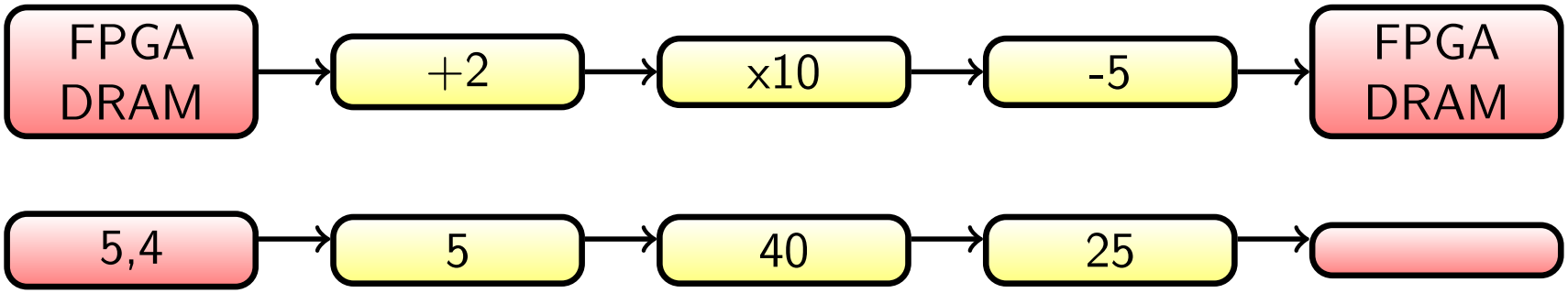
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



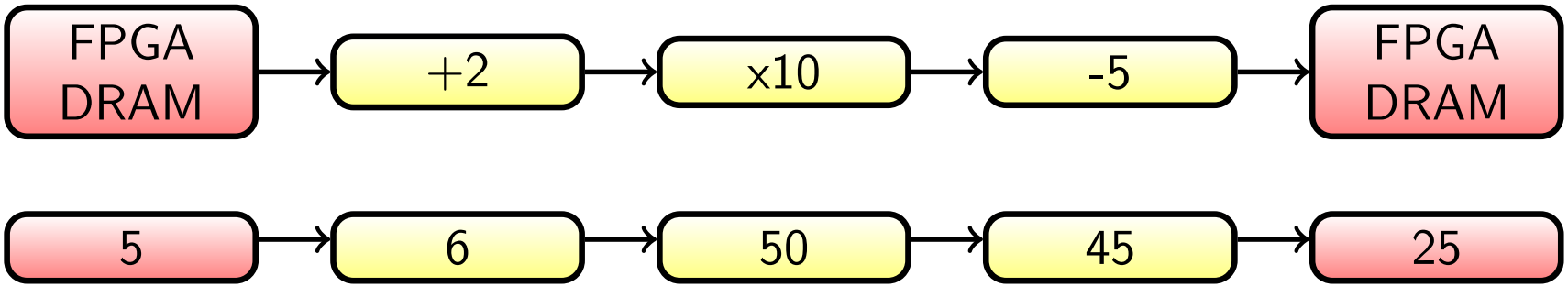
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



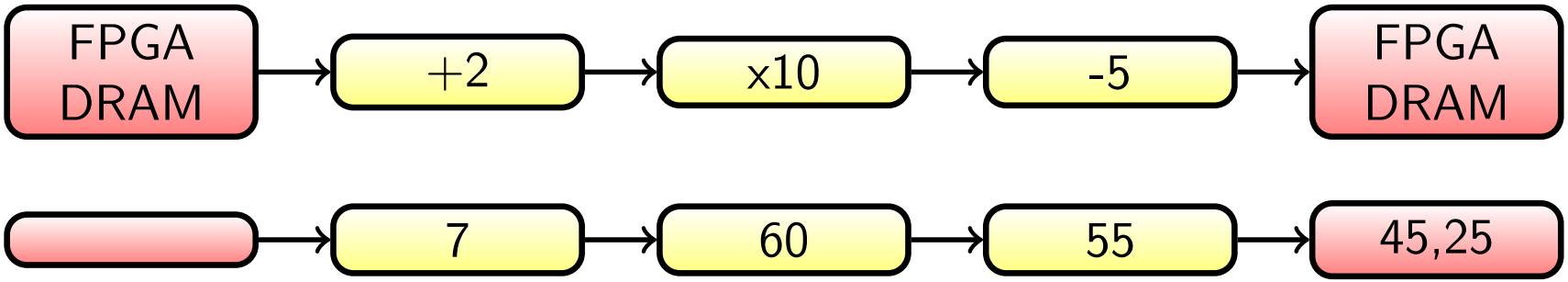
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



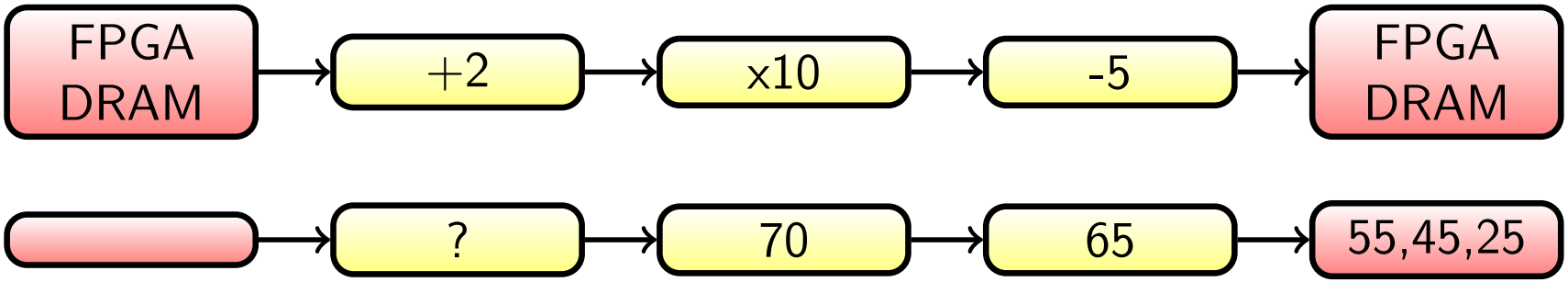
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



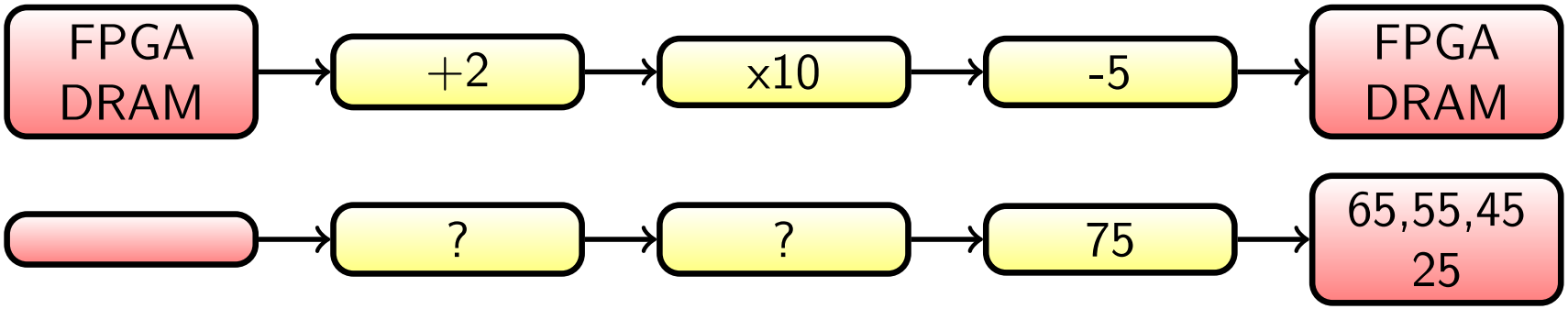
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



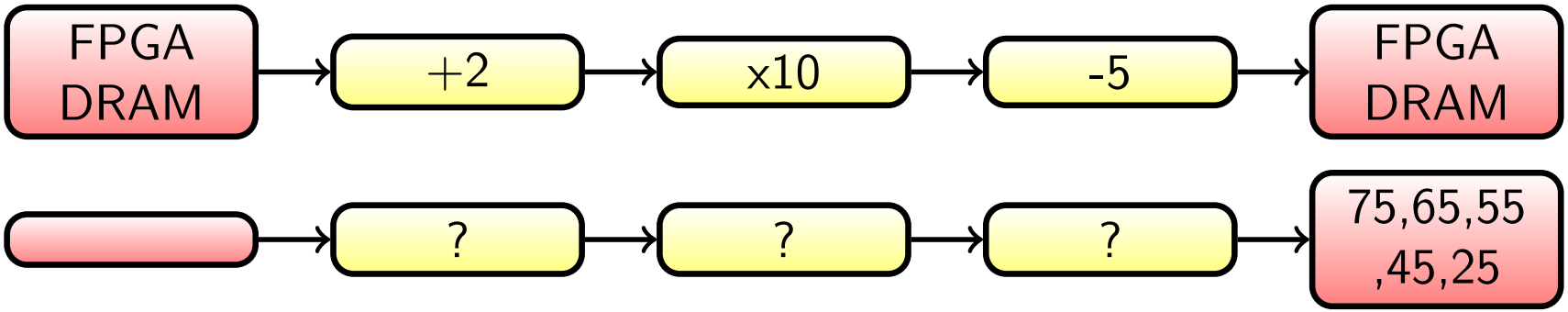
Field Programmable Gate Arrays (FPGA's)

- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



Field Programmable Gate Arrays (FPGA's)

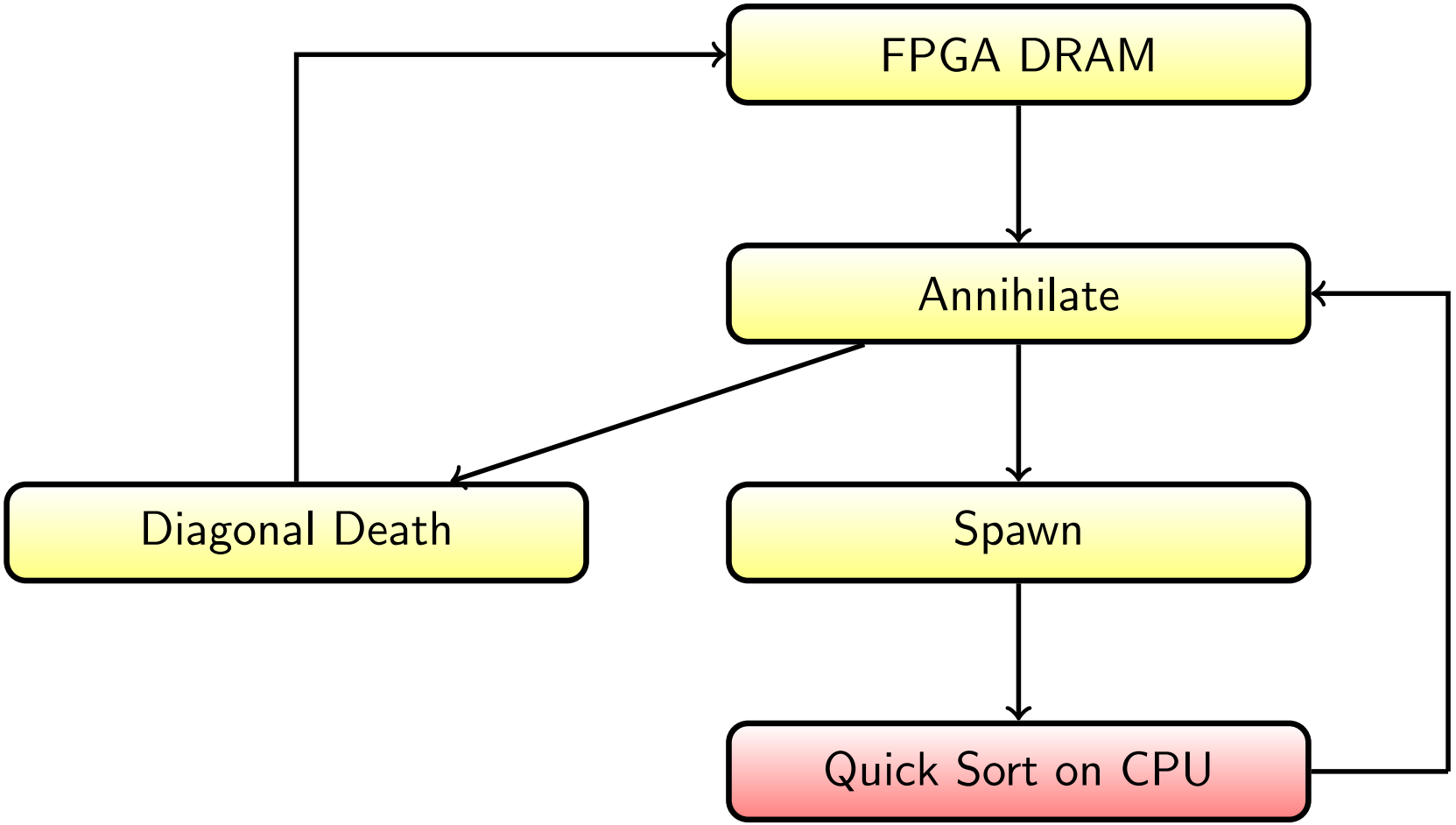
- A programmable grid of logic components.
- Components can be grouped and connected so that each group performs an operation and passes it too the next group each clock cycle.
- Sending data back up the chip can be tricky as we have to meet timings.
- Clock rate of $\sim$ 100 MHz c.f. CPU $\sim$ 1000 MHZ .



FCIQMC on FPGA's

- The psip vector makes a good candidate for streaming through the FPGA.
- Diagonal death $\mathcal{O}(L)$ and spawning $\mathcal{O}(N_p)$. Where L is the length of psip vector.
- Annihilation requires sorting the newly spawned psips $\mathcal{O}(l \log l)$
- Efficient sorting tricky.
- Send to CPU for sort.
- Currently investigating the Hubbard model (can store the integrals on chip as they have a simple structure).

FCIQMC on FPGA's



Conclusions and Directions

- Move to multiple FPGAs and CPUs.
- Investigate the Hubbard Model and quantify speed up.
- Real chemical systems require access of the integrals to generate the matrix elements, this will be tricky.
- It may be more efficient for all psips on a determinant to spawn at once Diagonal Death and spawning both $\mathcal{O}(L)$.
- Is this version more efficient for an FPGA ? (Work in progress)

Acknowledgements

- Alex Thom, Michael Bearpark, James Spencer.
- EPSRC for a studentship.
- Stephen Girdlestone, Craig Davies, and Robin Bruce at Maxeler Technologies
- All calculations ran using HANDE (developed at Imperial College) and the Imperial College High Performance Computing Service.

EPSRC
Pioneering research and skills

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Imperial College London

Serial Correlation

- As τ is small the vector of psips only changes a small amount between iterations.
- Thus expectation values close in iteration space are correlated.
- One has to remove this serial correlation by blocking: Average into blocks and compute standard error of blocks.
- We used the Iterative algorithms for optimal block size in:
  R. M. Lee, G. J. Conduit, N. Nemec, P. López Ríos, N. D. Drummond, Phys. Rev. E **83**, 066706 (2011)