The strange energy spectrum of MPS ground states

with Max Silvester and Giuseppe Carleo

1. Introduction to DMRG
2. Extrapolation with truncation error and with the energy variance
3. Perfect Sampling; fat tails for the variance
4. The unusual energy spectrum
5. Improved extrapolations

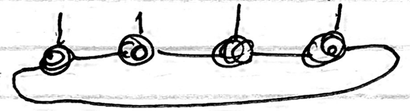
Max Silvester

Intro to DMRG

Diagrams: matrix multiply $\sum_j A_{ij} B_{jk}$

External lines: usually physical state
Internal: summed over

Periodic Matrix Product State



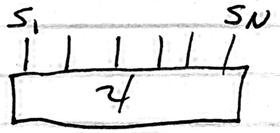
$= \mathrm{tr}\{A[s_1]\, B[s_2] \cdots\}$

Hamiltonians: simple, usually local

e.g. Heisenberg $H = \sum_{\langle ij \rangle} J \vec{S}_i \cdot \vec{S}_j$

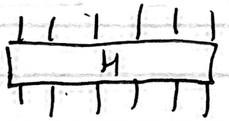
The Hilbert space is always ~~separable~~, but physical sites live on a lattice in 1D, 2D, or 3D

1D



$s_i = \uparrow$ or $\downarrow$, 0 or 1
2^N complex #'s

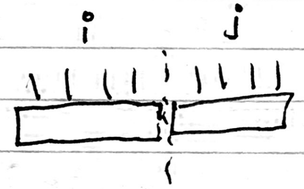
$H =$



Both are written as tensor networks

Matrix Product State (MPS) or MP operator MPO

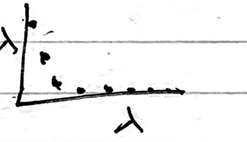
If we cut a wavefn in two, we can do an SVD



$$\psi_{ij} = U_{i\lambda} D_\lambda V^T_{\lambda j}$$

Schmidt Decomposition

The key to tensor networks is that the singular values are mostly near zero



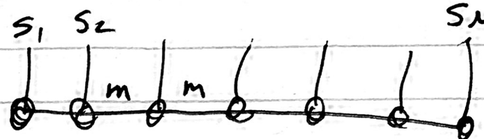
If we retain only m values, the exact truncation error is $\sum_{\lambda=m+1}^{N2^{N/2}} D_\lambda^2$

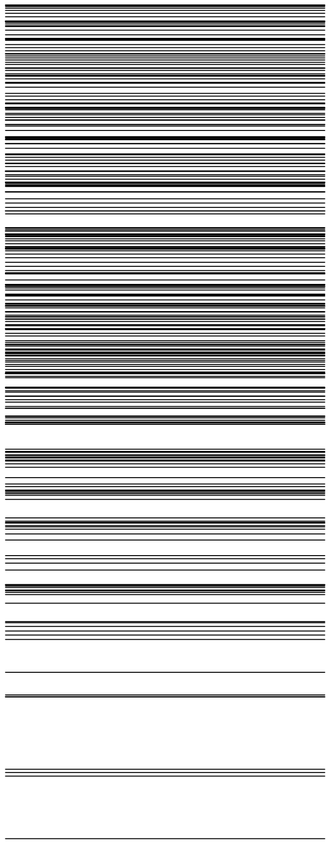
$= \varepsilon_{exact}$. This is a measure of the error in the Wfn.

The von Neumann entanglement entropy is

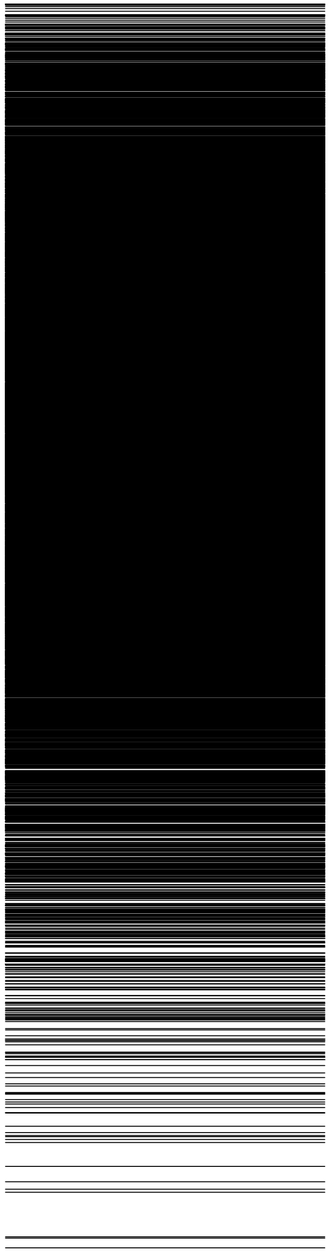
$$S = -\sum_\lambda D_\lambda^2 \ln D_\lambda^2$$

For ground states, S is small — area of the cut (Area Law). For Area law states in 1D, an MPS is ideal





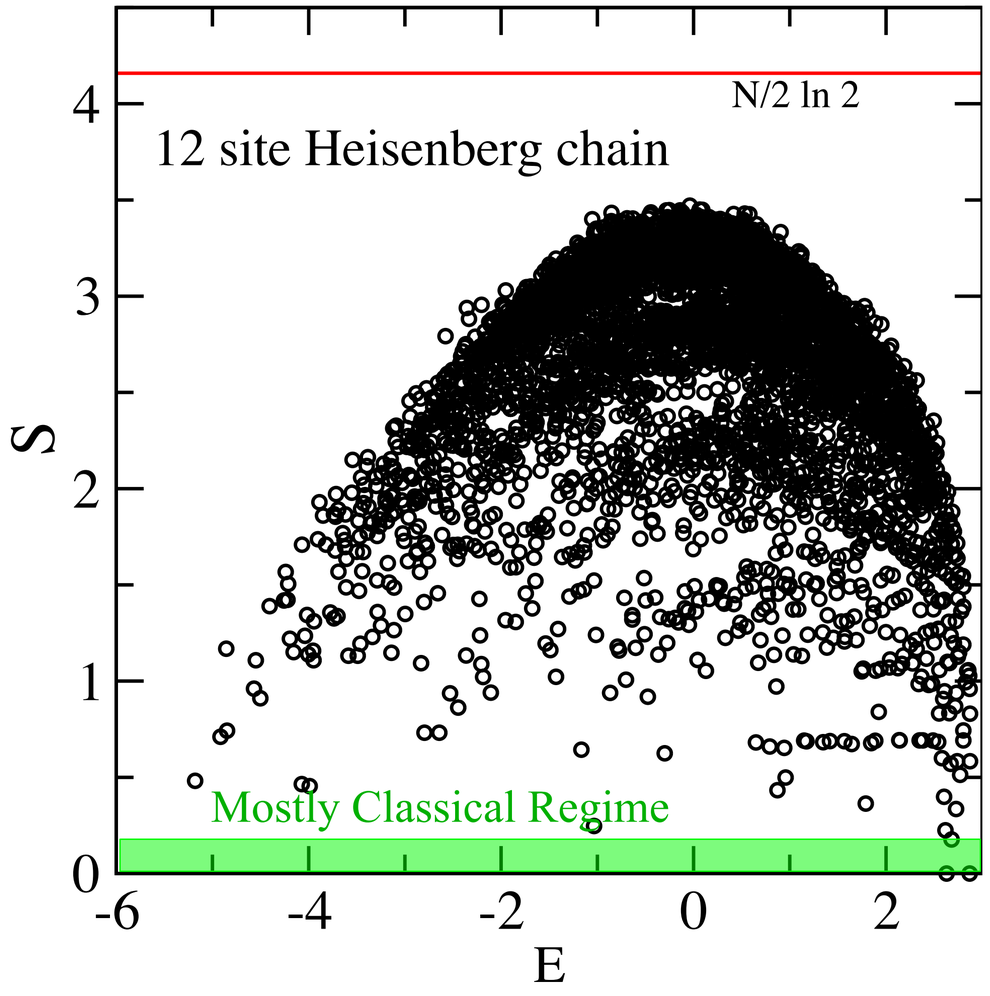
N=8



N=12

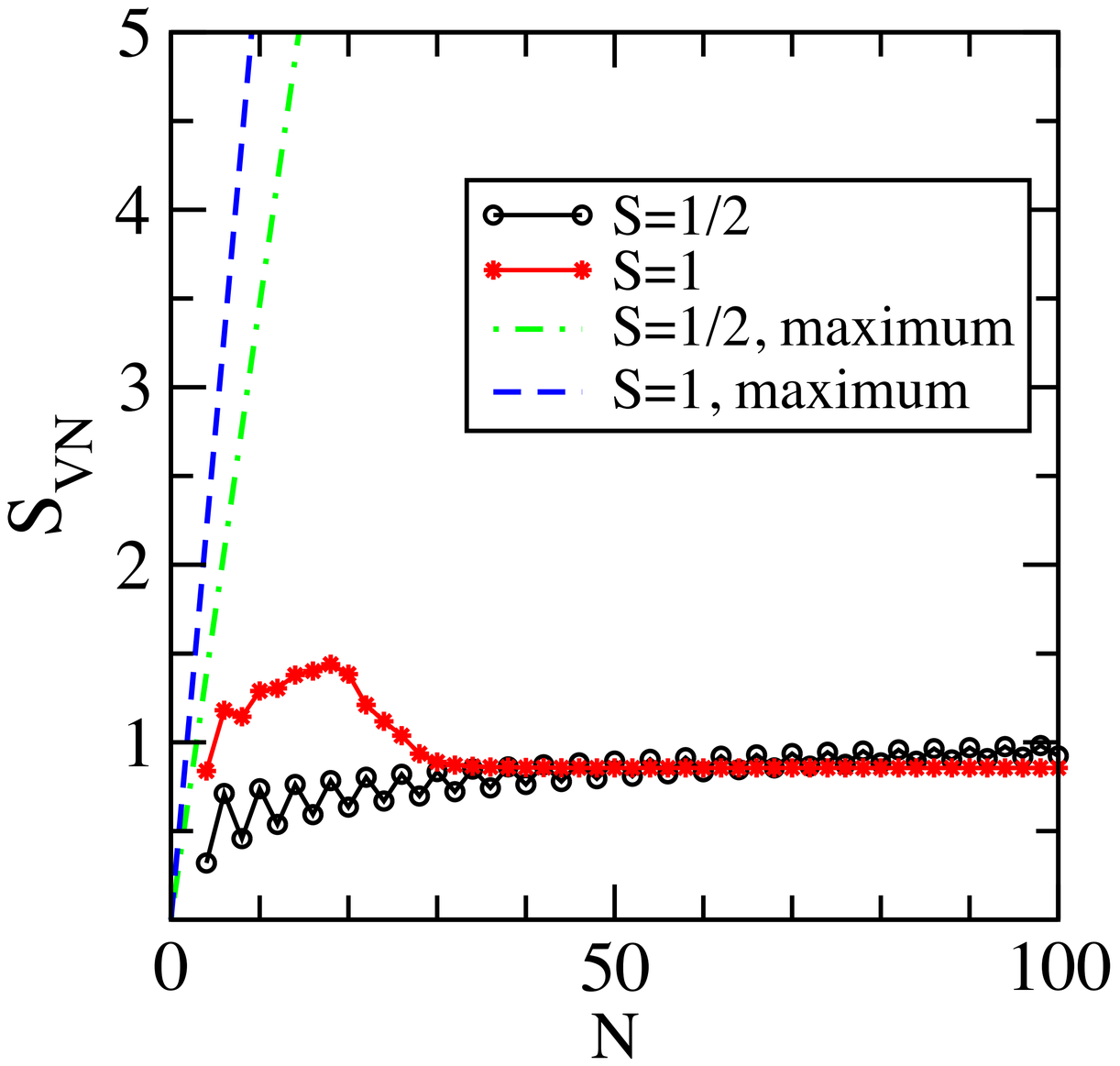
Energy levels of S=1/2 Heisenberg chains

Bulk eigenstates are "super-entangled"



Von Neumann Entanglement entropy S for every eigenstate (system divided in center)

Entanglement entropy of the ground state of N-site Heisenberg chains



S=1: gapped, bigger d, edge spinons

S=1/2: gapless, log correction, "RVB" oscillations

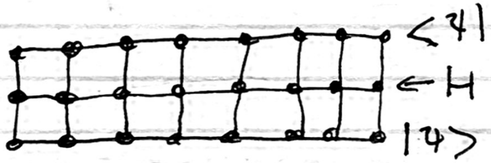
System are cut in two, symmetrically, with open boundaries

(Doing all these systems with DMRG took just a few minutes on my laptop)

For 2D problems, we can study a strip or cylinder with a snake path

Computational effort grows exponentially with the width.

(For wide widths, should use a 2D TN with an MPS for ψ & MPO for H,

$E =$ 

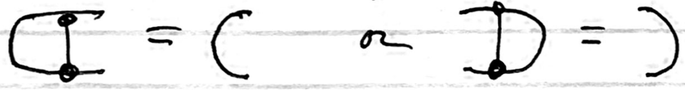
$= \langle\psi|H|\psi\rangle$

H-MPO:

W ~ 3-5 simple 1D
~ 20-40 wide 2D
~ 100-1000 quantum chemistry

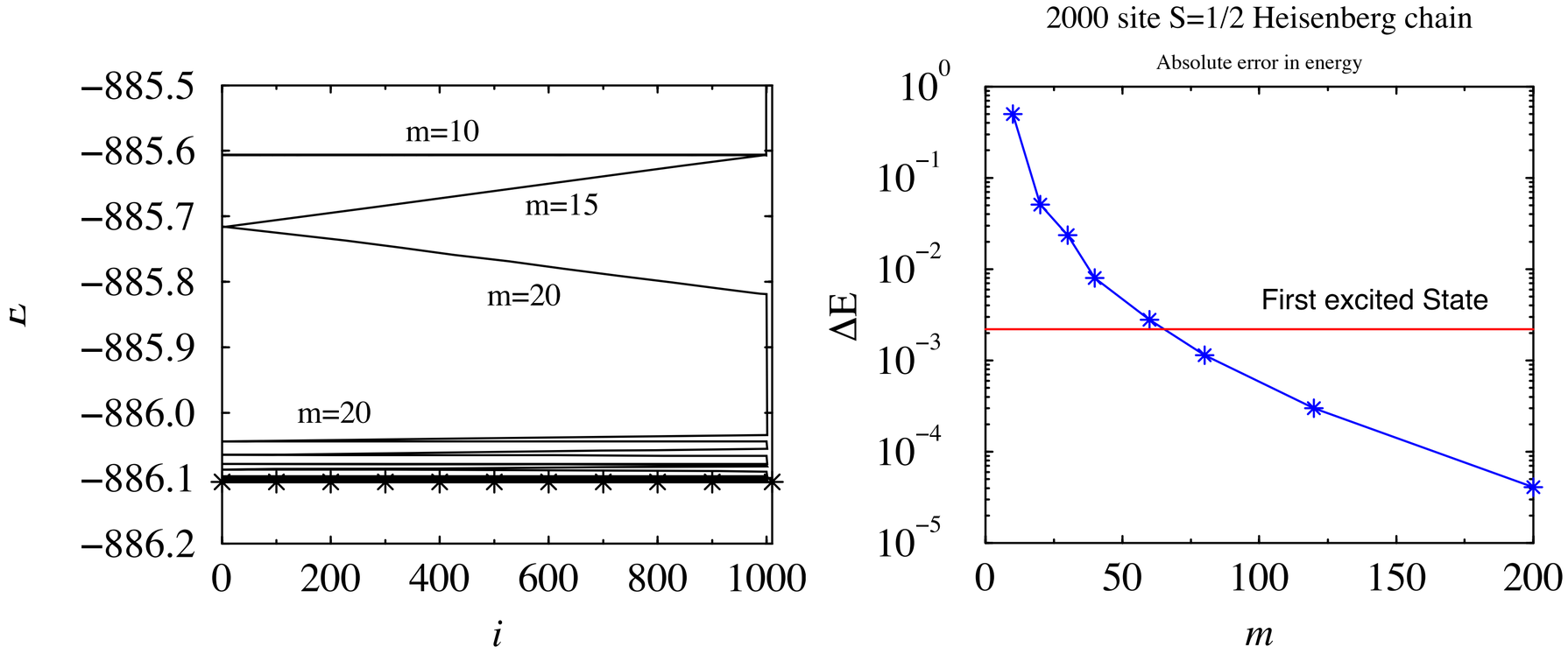
The density matrix renormalization group (DMRG) finds ψ + E with

1) Alternating least squares (sweeping)
2) Gauge choices/tensor orthogonality to create orthonormal subspaces

e.g. 

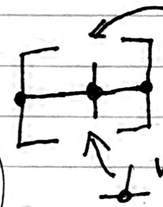
3) Efficient Lanczos or Davidson in the subspaces

Convergence in 1D

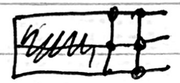


Comparison with Bethe Ansatz

Effective Hamiltonian during a sweep:

H_{eff} = 

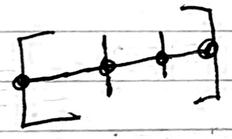
E = edge tensor

= 

$V = H_{eff} V$

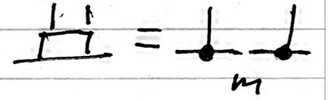
Orthogonality: $H\psi = E\psi$ Lanczos

Two site algorithm:



V =

(a) Find optimal two-site V

(b) SVD to factorize it 

with truncation

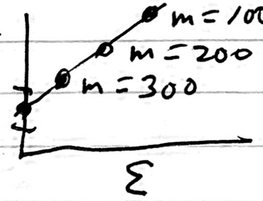
This:

(a) Speeds convergence significantly

(b) Allows m to increase efficiently

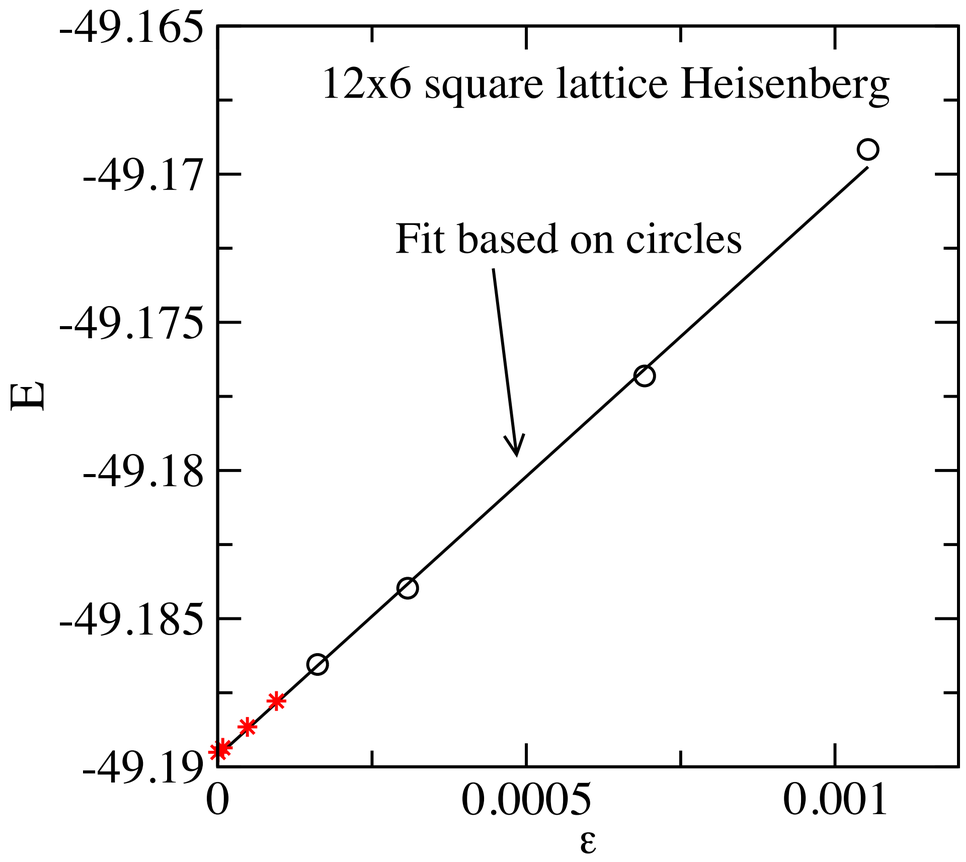
(c) Gives an estimate of the truncation error $\varepsilon = \sum_{m+1}^{\infty} D_\lambda^2$

which can be used to extrapolate E



very useful but delicate

Energy extrapolation



Probability of states thrown away
= truncation error (function of m)

Assign error bars to result:
if the fit is this good,
assign (extrapolation from
last point)/5

(no derivation, just
experience that this works
on lots of systems)

If the fit looks worse,
increase the error bar
(substantially) or don't use
that run/keep more states
or smaller size system.

Weaknesses of truncate error extrapolate

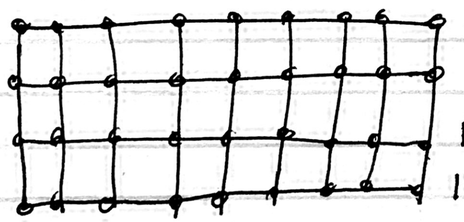
1) Tied to 2-site algorithm and sweeping history (increase of m)

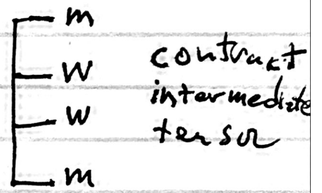
2) May fail for non-local Hams (quantum chemistry)

What about variance extrapolation?

$\sigma^2 = \langle\psi|(H-E)^2|\psi\rangle = 0$ for any eigenstate

$= \langle\psi|H^2|\psi\rangle - E^2$ $\quad E = \langle\psi|H|\psi\rangle$

$\langle H^2\rangle =$ 



Contraction time $\sim m^3 w^2 N$

vs $m^3 w N$ for DMRG — extra factor of w

Can this be sped up with sampling?

Prob of a spin config: $P(s_1, s_2 \ldots) = |\psi(s_1, s_2 \ldots)|^2$

There are Markov chain methods to sample/sweeps, but we don't need them

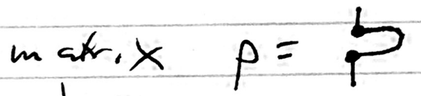
— perfect sampling —

Perfect Sampling – Mimics Physical Measurement

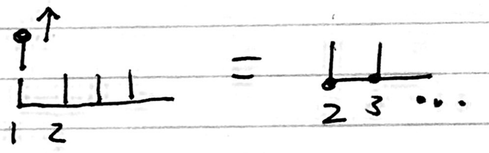
Measuring bit by bit gives the same $P(s\ldots)$ as all at once.

1. Adjust orthogonality center to site 1

   $\mathbb{D} \Rightarrow$ on site $2 \ldots N$

2. Single site density matrix $\rho =$ 
   gives $P(\uparrow) + P(\downarrow)$ choose with random # gen.

3. Apply ~~projector~~ state obtained to MPS site



Repeat for all sites.

Perfect indep sampling from $P(s_1, s_2 \ldots)$

Cost Nm^2 / sample

$$E = \langle\psi|H|\psi\rangle = \sum_s \langle\psi|s\rangle\langle s|H|\psi\rangle$$

$$= \sum_s P(s) \frac{\langle s|H|\psi\rangle}{\langle s|\psi\rangle}$$

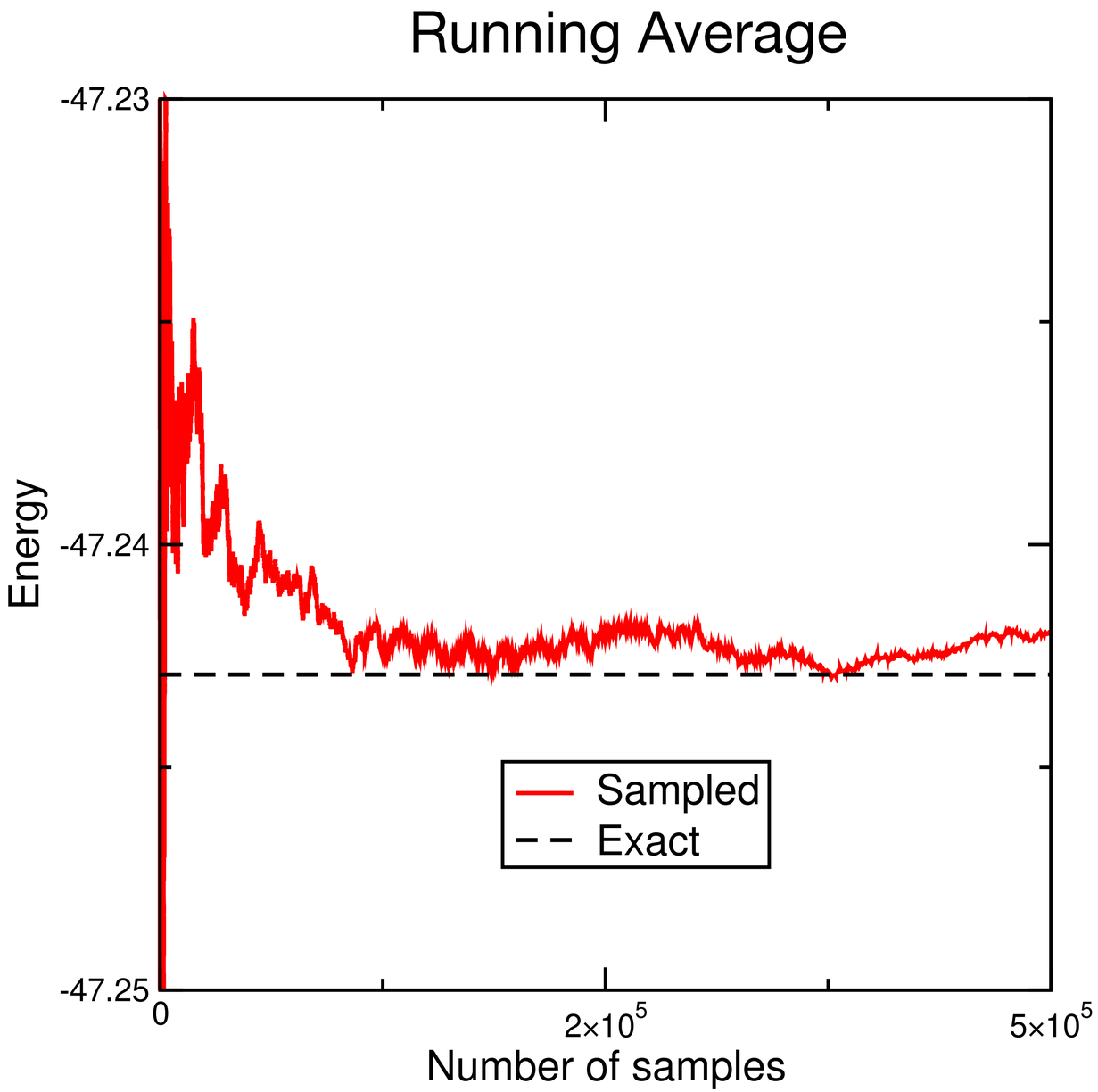
$\langle s|H|\psi\rangle$ Cost Nwm^2 but already have E

$$\sigma^2 = \sum_s \langle\psi|H-E|s\rangle\langle s|H-E|\psi\rangle$$

$$= \Delta_s + \sum_{s \text{ s.t. } P(s)\neq 0} P(s)\left(E_s^L - E\right)^2 \qquad E_s^L = \frac{\langle s|H|\psi\rangle}{\langle s|\psi\rangle}$$

Bias in sampling σ^2

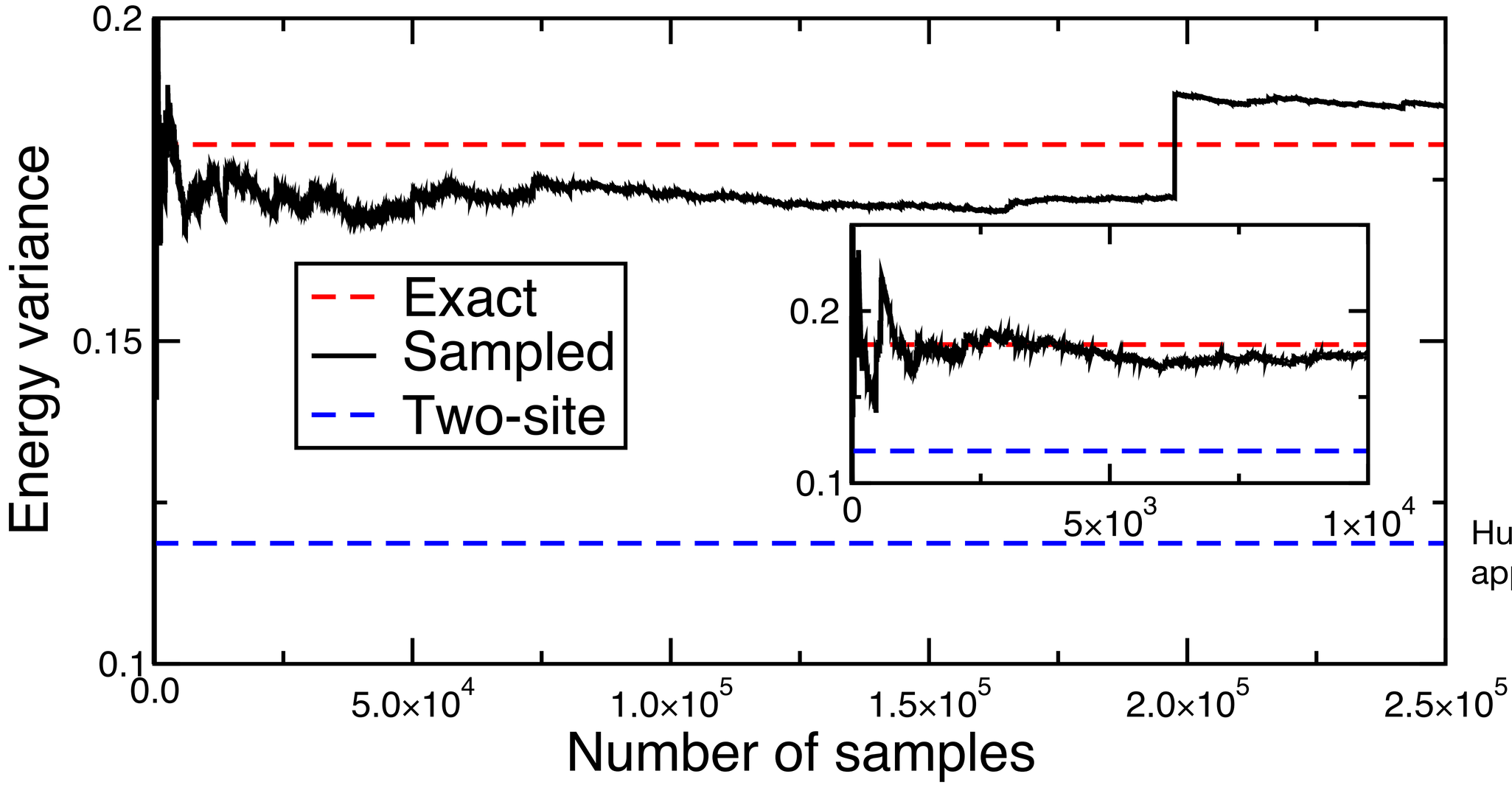
Energy sampling



12x6 Heisenberg square lattice

m = 120

Variance sampling

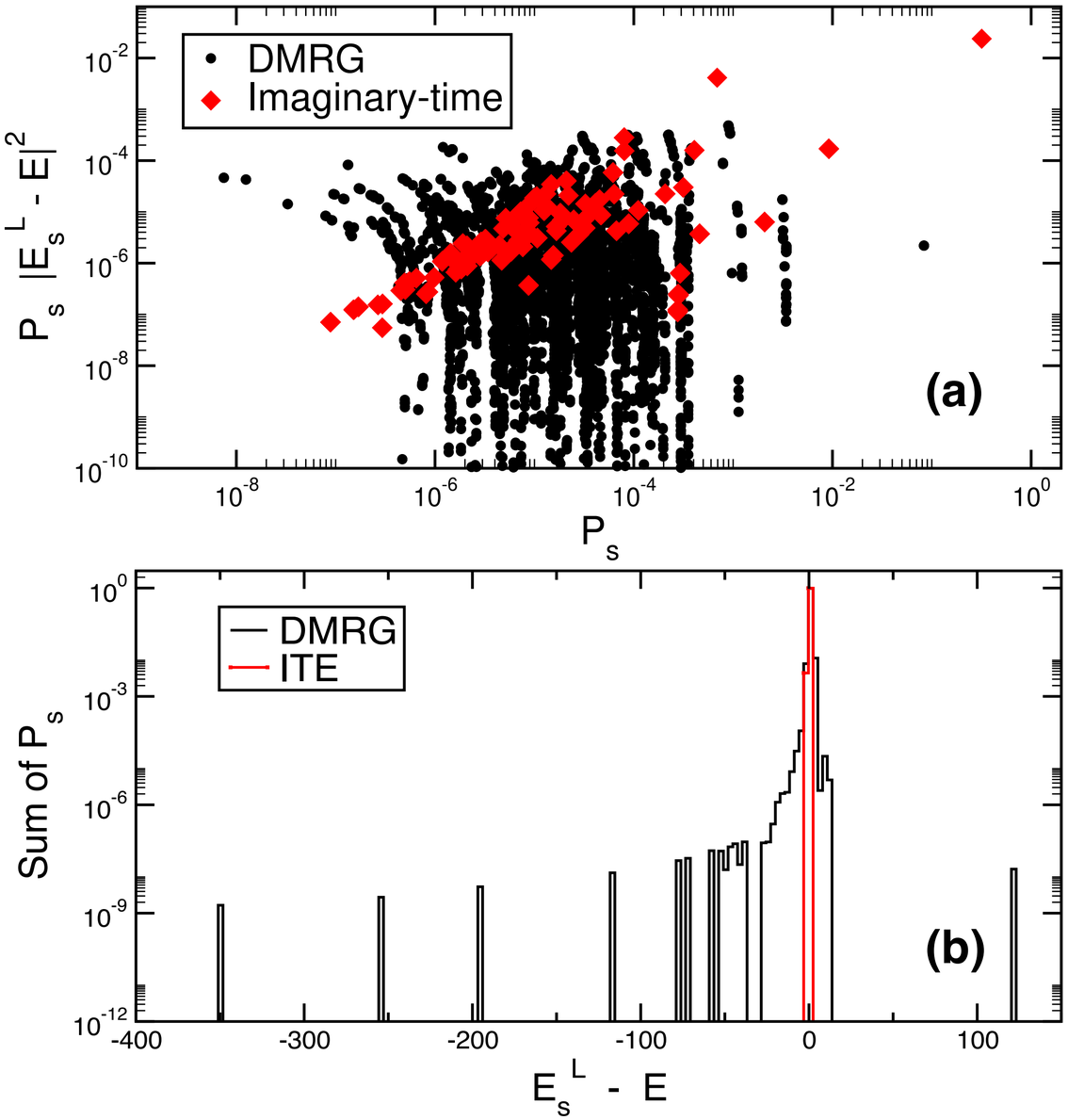


Variance has fat tails-- terrible sampling

Hubig, et al approximate variance

4x4 Heisenberg square lattice torus (pbc) m = 80

Small system exact calculation



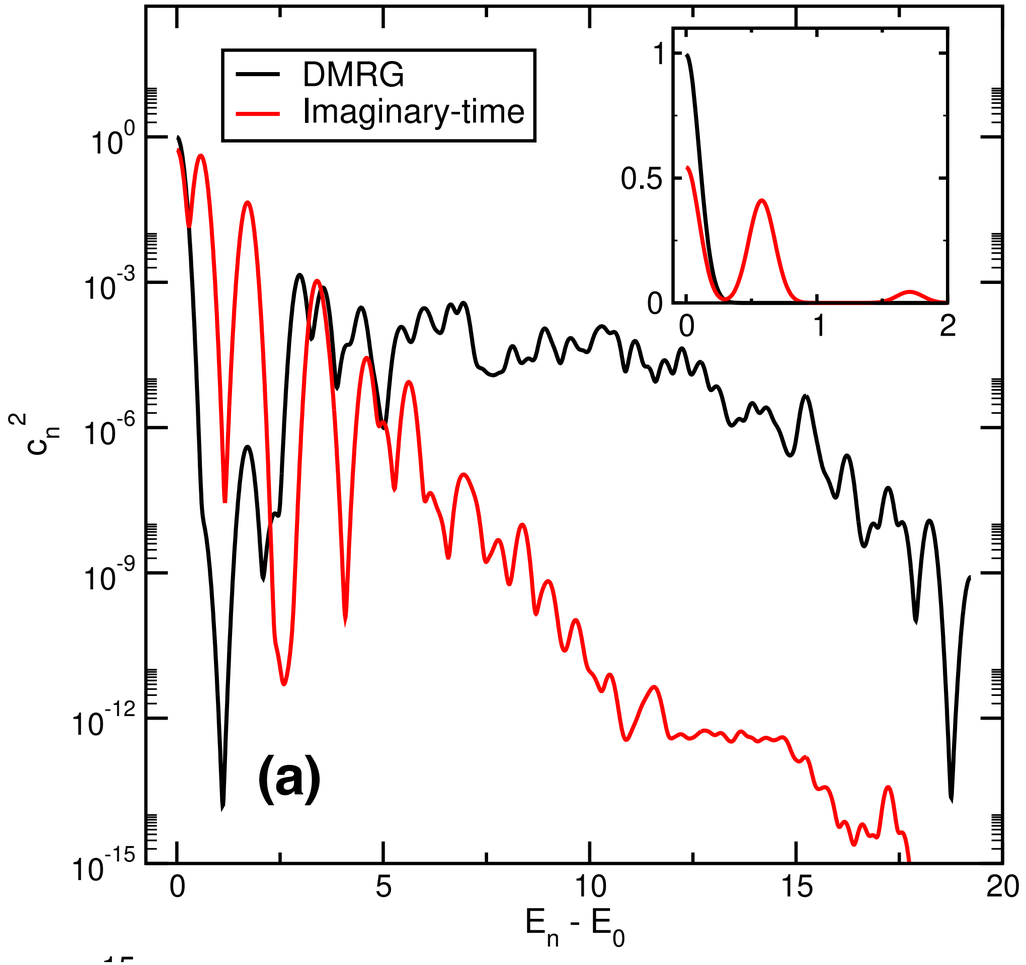
Here we compare the sampling distribution for a DMRG wavefunction and a wavefunction obtained from imaginary time evolution. We plot the contribution to the variance of each sample. (4x4 torus)

Histogram of distribution of E-local, DMRG versus imaginary time evolution

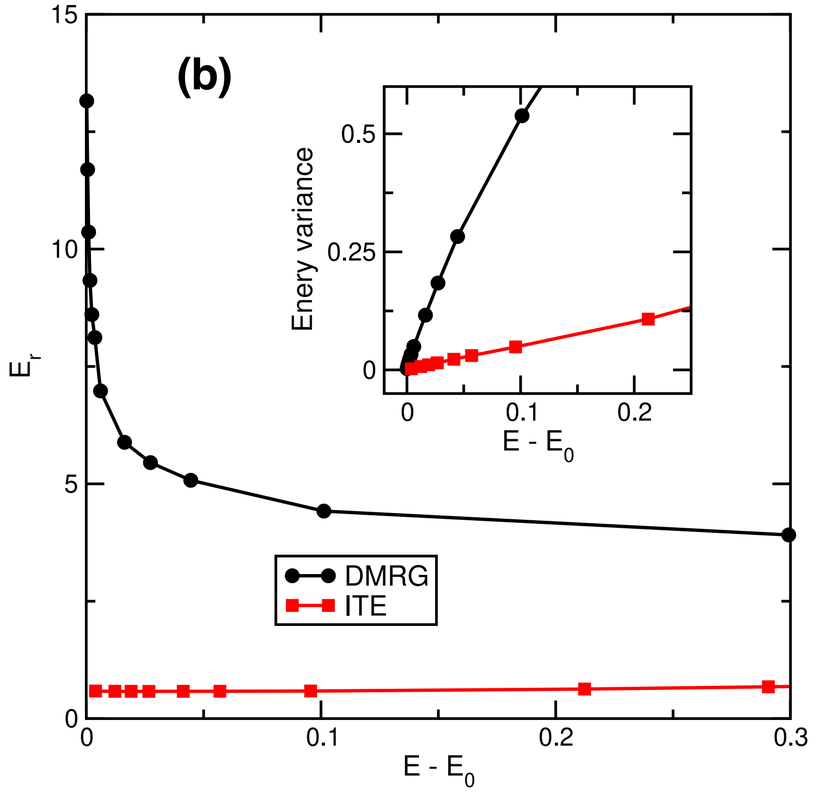
Energy Spectrum

To try to understand this, we decomposed the approximate ground states into superpositions of exact eigenstates

$$|\psi\rangle = \sum_n c_n |n\rangle \qquad \text{n an eigenstate}$$



We can look at the average energy of the excited state part of the wavefunction E_r as we increase the bond dimension (move to left)



Why do DMRG states have this strange Energy Spectrum?

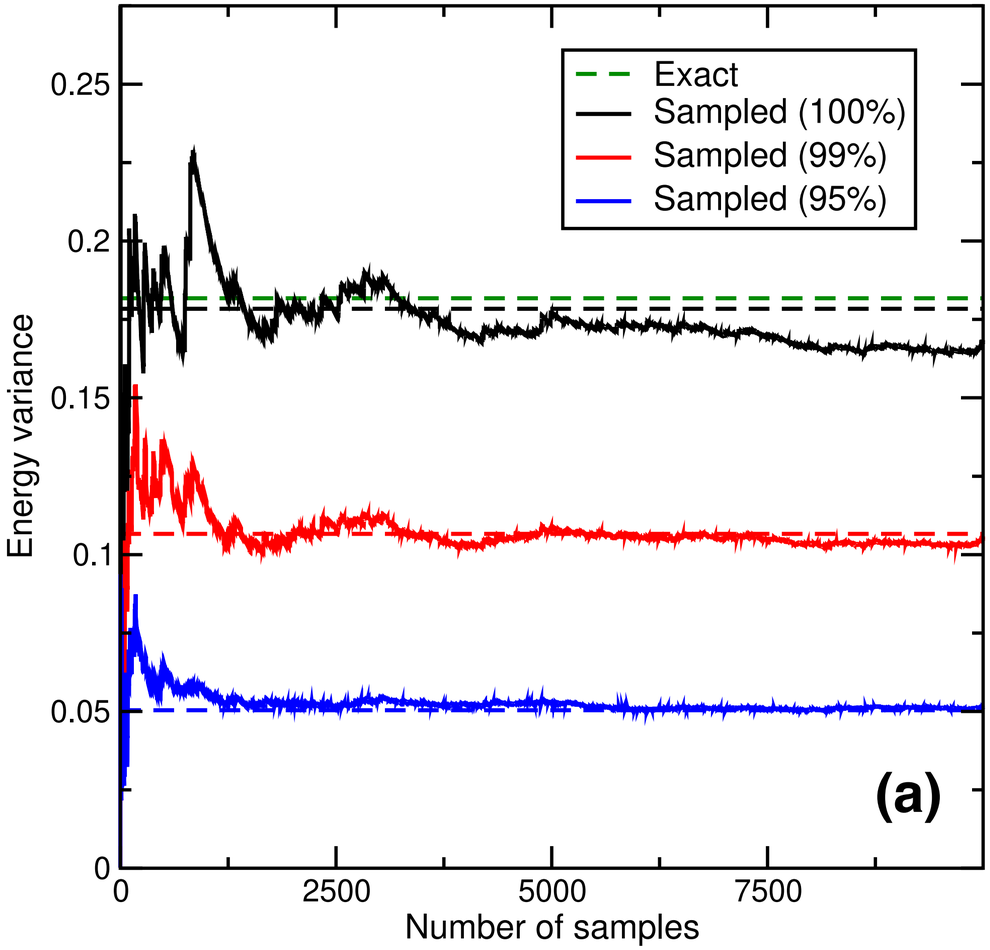
Optimized MPS ground states satisfy two criteria: low energy, and small bond dimension. This is incompatible with being a superposition of a few low energy states.

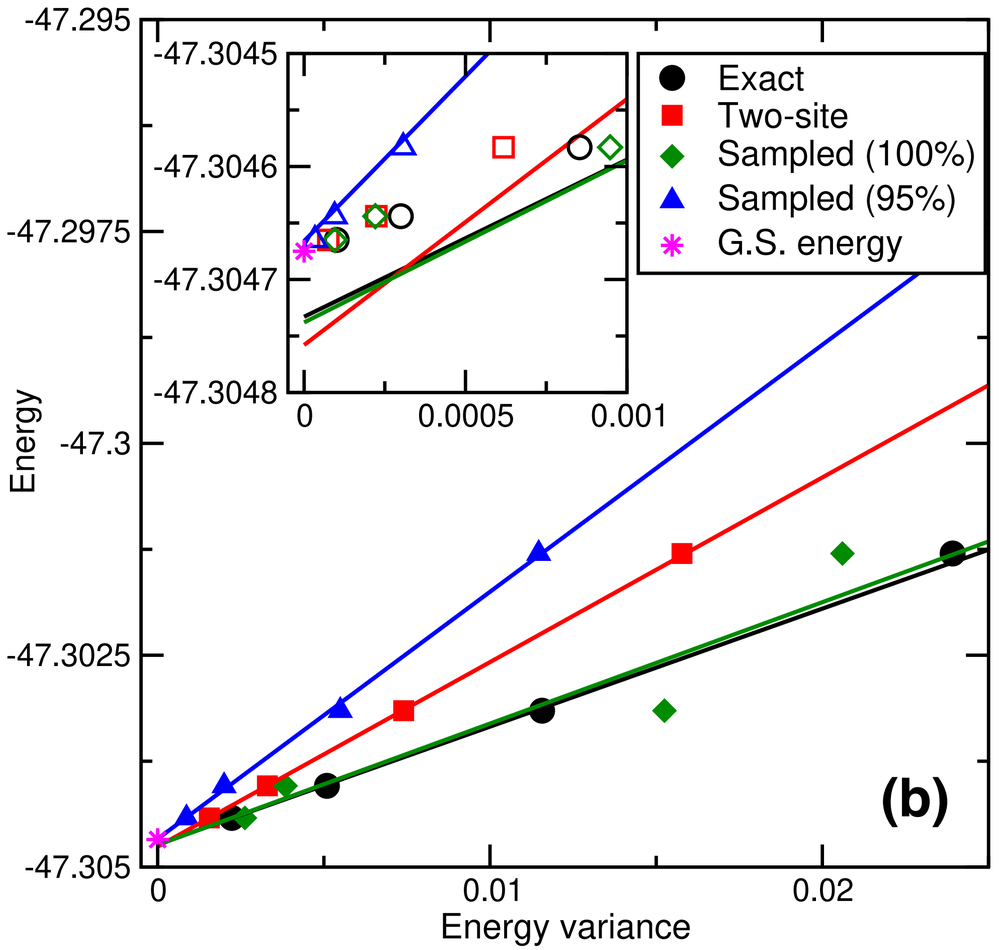
Suppose, for high accuracy, we needed m~1000 for the ground state. Low lying states are slightly more entangled, but say we still need m~1000 for them. Say our approximate MPS has m=100. If it was a superposition of the excited states, it would say a few MPS with high bond dimension could combine and cancel out almost all the high bond dimension parts.

Apparently, the only way to get low bond dimension is a superposition of many states—which must have a broad spectrum.

Can we revive our extrapolation based on sampled variance?

Suppose we impose a window on our sampled E-local, throwing away extreme samples. This is highly biased, but the bias goes away as the bond dimension increases. If the bias decreases smoothly, the extrapolation can deal with it.





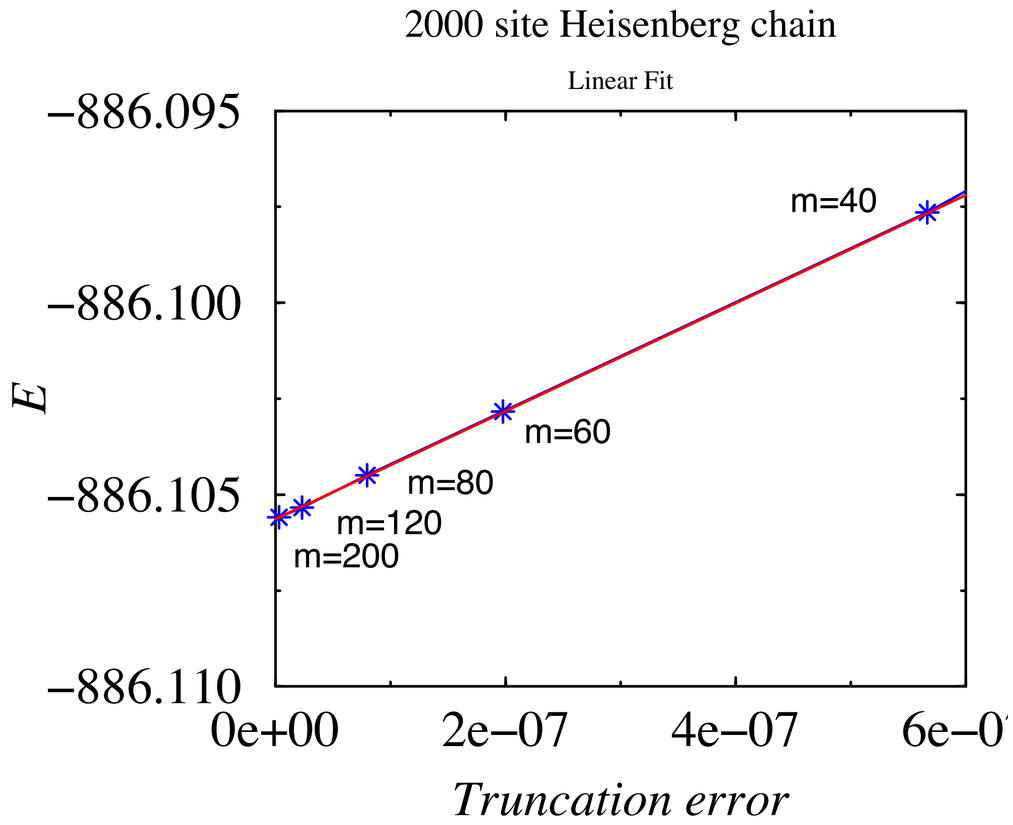
We actually find <u>better</u> extrapolations with the truncated sampled method (at 95%) than any of the alternatives—truncation error, exact variance, two site variance…

Conclusions

DMRG ground states inherently have a strange energy spectrum: lots of weight in the exact ground state, and very spread out tiny contributions going to very high energies! This is very different from imaginary time approximate wavefunctions.

Biased truncated energy extrapolations work quite well, providing an excellent alternative to truncation error extrapolations.

Extrapolation of the energy



Extrapolation improves the energy by a factor of 5-10 and provides an error estimate.