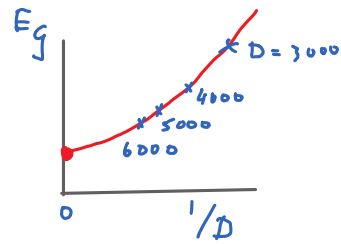


1. Energy variance

[Hubig2018]

When doing MPS computations involving SVD truncations of virtual bonds, the results should be computed for several values of the bond dimension, D , to check convergence as $D \rightarrow \infty$. Often it is also necessary to extrapolate the results to $D = \infty$, e.g. by plotting results versus $1/D$ or some power thereof.

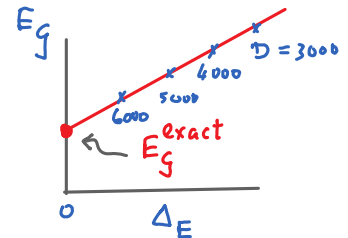


However, for some computational schemes, it is not *a priori* clear how the observable of interest scales with D , nor how it should be extrapolated to $D = \infty$. An example is ground state energy when computed using 1-site DMRG with subspace expansion [Hubig2015], because it does not rely on SVD truncation of bonds.

Thus, it is of interest to have a reliable error measure without requiring costly 2-site DMRG. A convenient scheme was proposed in [Hubig2018], based on a smart way to approximate the full energy variance,

$$\Delta_E := \|(\hat{H} - E)\psi\|^2 = \langle \psi | (\hat{H} - E)^2 | \psi \rangle \quad (= \text{zero for an exact eigenstate}) \quad (1)$$

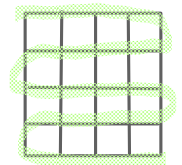
$$= \langle \psi | \hat{H}^2 | \psi \rangle - E^2, \quad \text{with } E = \langle \psi | \hat{H} | \psi \rangle \quad (2)$$



Then extrapolations can be done by computing quantity of interest for several D , but plotting the results via Δ_E , and extrapolating to $\Delta_E \rightarrow 0$

If quantity of interest is energy, then extrapolation is linear, $E_g(\Delta_E) = E_g^{\text{exact}} + a \cdot \Delta_E$ (3)

Computing $\langle \psi | \hat{H}^2 | \psi \rangle$ directly is costly for large systems with long-ranged interactions, such as 2D systems treated by DMRG snakes. Also, computing Δ_E as the difference between two potentially large numbers is prone to inaccuracies. [Hubig2018] found a computation scheme in which the subtraction of such large numbers is avoided *a priori*.



Key idea: use projectors $P^{n\perp}$ onto mutually orthogonal, irreducible spaces $V^{n\perp}$

$$\text{Recall (2.11): } 1_V = 1_d^{\otimes L} = \sum_{n=0}^L P^{n\perp}, \quad P^{n\perp} P^{n'\perp} = \delta^{nn'} P^{n\perp} \quad (4) \quad \text{completeness} \quad \text{orthogonality} \quad (5)$$

$$\text{with } P^{0\perp} = |\psi\rangle\langle\psi| \quad (6)$$

$$P^{1\perp} = \sum_{l=1}^L \text{diagram}, \quad P^{2\perp} = \sum_{l=1}^L \text{diagram} \quad (7) \quad (2.16) \quad (2.17)$$

$$\text{Insert completeness into definition of variance: } \Delta_E^{(4)} = \langle \psi | (\hat{H} - E) \sum_{n=0}^L P^{n\perp} (\hat{H} - E) | \psi \rangle =: \sum_{n=0}^L \Delta_E^{n\perp} \quad (8)$$

Now two crucial simplifications occur:

Now two crucial simplifications occur:

$$\Delta_E^{0\perp} \stackrel{(5)}{=} \langle \psi | (\hat{H} - E) \underbrace{|\psi\rangle}_{(b) \ P^{0\perp}} \langle \psi | (\hat{H} - E) | \psi \rangle = (E - E)(E - E) = 0 \quad (9)$$

largest contribution to variance cancels by construction!

$n > 0$

$$\Delta_E^{n\perp} = \langle \psi | (\hat{H} - E) P^{n\perp} (\hat{H} - E) | \psi \rangle = \langle \psi | \hat{H} P^{n\perp} \hat{H} | \psi \rangle, \quad \text{since } P^{(n>0)\perp} | \psi \rangle \stackrel{(5,6)}{=} 0 \quad (10)$$

$$= P^{n\perp} P^{n\perp} \quad (\text{TS-II.2.11})$$

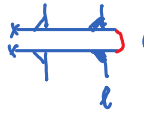
$$= \| P^{n\perp} \hat{H} \psi \|^2 \quad (11)$$

In practice, approximate Δ_E by the first two nonzero terms:

$$\Delta_E \simeq \Delta_E^{2s} = \Delta_E^{1\perp} + \Delta_E^{2\perp} = \langle \psi | \hat{H} P^{1\perp} \hat{H} | \psi \rangle + \langle \psi | \hat{H} P^{2\perp} \hat{H} | \psi \rangle \quad (12)$$

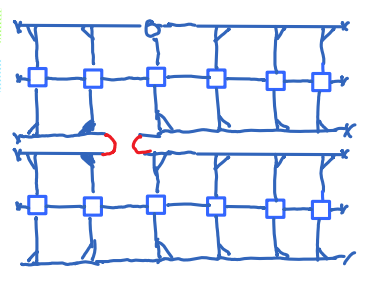
(11) is exact if longest-range terms in $\hat{H}$ are nearest-neighbor, because then $P^{(n>3)\perp} \hat{H} | \psi \rangle = 0$ [Gleis2022a] (13)

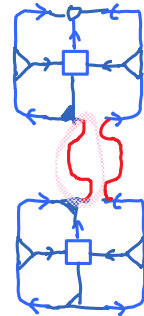
Explicit computations:

$n=1$: Recall $P^{1\perp} \stackrel{(\text{TS-II.2.16})}{=} \sum_{\ell=1}^L$  $= \sum_{\ell=1}^L P_{\ell, \ell+1}^{DK}$ (14)

mutually orthogonal! (TS-I.4.15)

$$\Delta_E^{1\perp} \stackrel{(10)}{=} \langle \psi | \hat{H} P^{1\perp} \hat{H} | \psi \rangle \stackrel{(14)}{=} \sum_{\ell} \langle \psi | \hat{H} P_{\ell, \ell+1}^{DK} \hat{H} | \psi \rangle \quad (15)$$

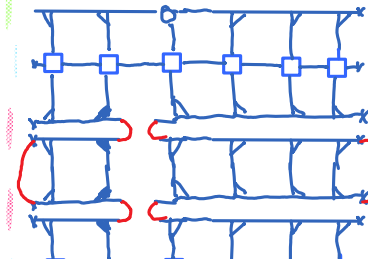
$$= \sum_{\ell=1}^L$$


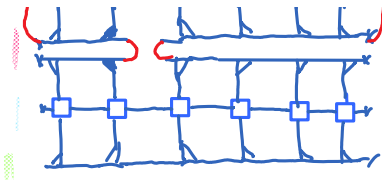
$$= \sum_{\ell=1}^L$$


$$= \sum_{\ell=1}^L \left\| P_{\ell, \ell+1}^{DK} \hat{H} \psi \right\|^2 \quad (16)$$

We would like to avoid computing D_d explicitly, because of its large image dimension. So, exploit

completeness of kept together with discarded isometries: $\begin{array}{c} \rightarrow \\ \leftarrow \end{array} = \begin{array}{c} \rightarrow \\ \rightarrow \end{array} - \begin{array}{c} \rightarrow \\ \leftarrow \end{array} \quad (17)$

$$\Delta_E^{1\perp} \stackrel{(15)}{=} \sum_{\ell} \langle \psi | \hat{H} P_{\ell, \ell+1}^{DK} P_{\ell, \ell+1}^{DK} \hat{H} | \psi \rangle = \sum_{\ell=1}^L \left\| P_{\ell, \ell+1}^{DK} \hat{H} \psi \right\|^2 = \sum_{\ell=1}^L$$




 (18)

$$\begin{aligned}
 &= \sum_{l=1}^L \left[\text{Diagram with a red loop on the left edge} \right] \\
 &= \sum_{l=1}^L \left[\text{Diagram with a red loop on the bottom edge} \right] - \sum_{l=1}^L \left[\text{Diagram with a red loop on the top edge} \right] \quad (19)
 \end{aligned}$$

also follows from (16) using $\square = \bigcirc$

$N = 2$: Recall $P^{z\downarrow} = \sum_{l=1}^{L-1} \left[\text{Diagram with arrows and a red loop} \right] = \sum_{l=1}^L P_{l,l+1}^{DD} \quad (20)$

$$\Delta_E^{z\downarrow} = \langle \psi | \hat{H} \tilde{P}^{z\downarrow} \hat{H} | \psi \rangle = \| P^{z\downarrow} H \psi \|^2 = \sum_{l=1}^{L-1} \| P_{l,l+1}^{DD} H \psi \|^2 \quad (21)$$

$$= \sum_{l=1}^{L-1} \left[\text{Diagram with two squares and arrows} \right] = \sum_{l=1}^{L-1} \left[\text{Diagram with two squares and arrows, boxed} \right] \quad (22)$$

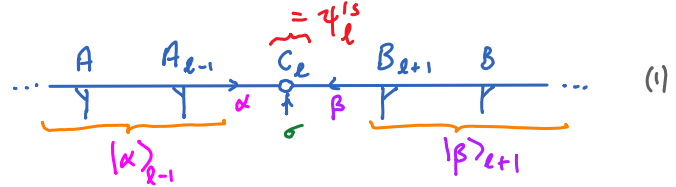
again use $\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} = \begin{array}{c} | \\ - \\ | \end{array} - \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array}$
 and $\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} = \begin{array}{c} | \\ - \\ | \end{array} - \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \quad (23)$

Problem: when exploiting symmetries, 1-site DMRG performs poorly, because it does not explore subspaces with different quantum numbers. An early remedy for this is 2-site DMRG, but that is computationally much more expensive than 1-site DMRG. Subsequent suggestions for 1-site DMRG with symmetries are 'density matrix perturbation' [White2005], the 'center matrix wave function formalism' [McCulloch2007], 'subspace expansion' [Hubig2015], and 'controlled bond expansion' (CBE) [Gleis2022], which performs best.

Reminder of 1-site DMRG,

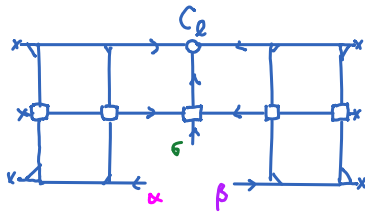
in site-canonical representation:

Local basis: $|\alpha, \sigma, \beta\rangle := |\alpha\rangle_{l-1} |\sigma\rangle_l |\beta\rangle_{l+1}$

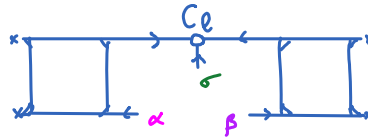


Minimize energy with constraint of fixed normalization, 1 site at a time:

$$\frac{\partial}{\partial C_l} \left[\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle \right] = 0 \quad (2)$$



=

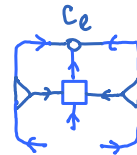


close zipper

(3)

$$\left[H_l^{1s} \right]^{a'} \left[\psi_l^{1s} \right]^a = E \left[\psi_l^{1s} \right]^a$$

$$a = (\alpha, \sigma, \beta)$$



cost $\sim \mathcal{O}(D^3 d \omega)$

(4)

Solve for 'eigenvector' with lowest eigenvalue, say $\tilde{\psi}_l^{1s}$, then do SVD on it to move to next site:

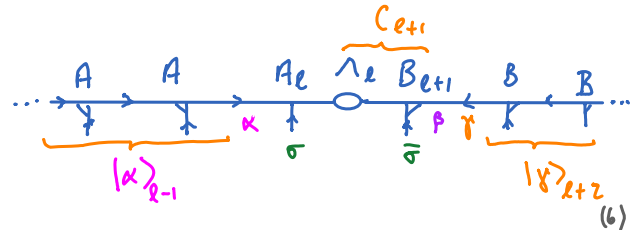
$$D \rightarrow \psi_l^{1s} \rightarrow B_{l+1} \rightarrow D \xrightarrow{\text{SVD}} D \rightarrow (U) (S) (V^\dagger) \rightarrow B_{l+1} \rightarrow D \xrightarrow{\text{reshape}} D \rightarrow A_l \rightarrow C_{l+1} \rightarrow D \quad (5)$$

Important: dimensions of $\tilde{C}$ are fixed, hence truncation is neither needed nor possible!

Reminder of 2-site DMRG,

in 2-site-canonical representation:

Local basis: $|\alpha, \sigma, \bar{\sigma}, \gamma\rangle := |\gamma\rangle_{l+2} |\bar{\sigma}\rangle_{l+1} |\sigma\rangle_l |\alpha\rangle_{l-1}$

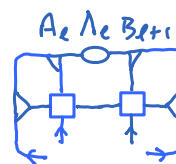


(6)

Minimize energy two sites at a time:

$$\left[H_l^{2s} \right]^{a'} \left[\psi_l^{2s} \right]^a = E \left[\psi_l^{2s} \right]^a$$

$$a = (\alpha, \sigma, \bar{\sigma}, \gamma)$$



cost $\sim \mathcal{O}(D^3 d^3 + D^3 d^2 \omega)$

(7)

Solve for 'eigenvector' with lowest eigenvalue, $(\tilde{M}B)$, then do SVD and truncate (!) to move to next site:

$$D \rightarrow \tilde{\psi}_l^{2s} \rightarrow D \xrightarrow{\text{SVD}} D \rightarrow (U) (S) (V^\dagger) \rightarrow D \rightarrow \tilde{A}_l \rightarrow \tilde{\Lambda}_l \rightarrow \tilde{B}_{l+1} \rightarrow D \xrightarrow{\text{truncate}} \tilde{A}_l \rightarrow \tilde{\Lambda}_l \rightarrow \tilde{B}_{l+1} \rightarrow D \xrightarrow{\text{QR}} \tilde{A}_l \rightarrow \tilde{A}_{l+1} \rightarrow D \quad (8)$$

Problem of single-site optimization: it is constrained to a variational space defined by outgoing state spaces $|\alpha\rangle_{l-1}, |\beta\rangle_{l+1}$. If the ranges of quantum numbers Q_α and Q_β for these spaces are too small to accurately represent the ground state, single-site DMRG has no way to enlarge them.

Two-site optimization does not have this problem: the action of H on two sites enlarges bond dimension in between, adding the full range of quantum numbers needed on that bond. If a certain quantum number was missing on that bond before the action of H, but appears afterwards with non-negligible weight, it will survive after SVD and truncation. Hence: two-site optimization can add missing quantum numbers, if needed.

But this comes at a cost: effective two-site Hamiltonian has dimension $D^2 d^2 \times D^2 d^2$. (9)
By contrast, effective one-site Hamiltonian has dimension $D^2 d \times D^2 d$.

Example: charge conservation:

$$Q_{l-1} + \sigma_l = Q_l \quad (10)$$

states carry two labels:

$$|\alpha_l\rangle = |Q_l, i_l\rangle \in \text{span}\{|\sigma_1, \sigma_2, \dots, \sigma_l\rangle\}$$

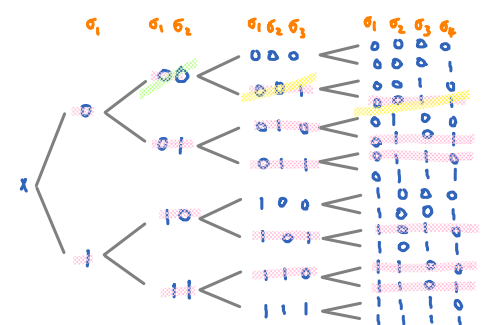
charge enumerates distinct states with same charge

$$Q_{l-1}, i_{l-1} \quad Q_l, i_l$$

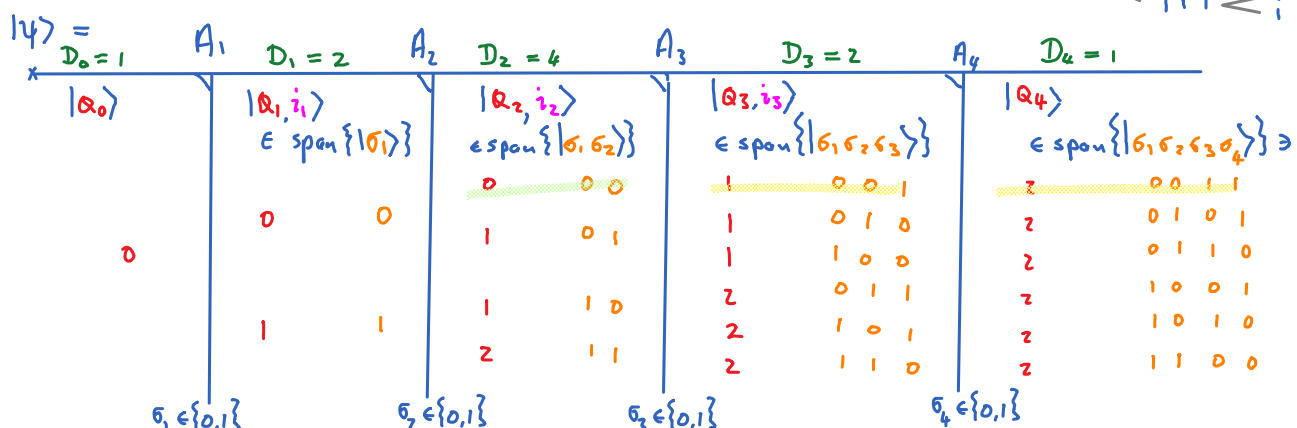
$$\sigma_l \in \{0, 1\}$$

{empty site, occupied site}

all possible charge configurations of 4-site chain of spinless fermions:

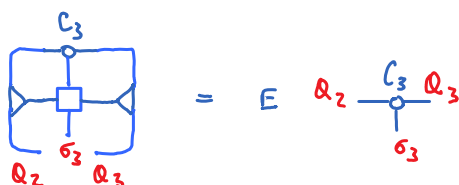


Consider 4-site chain of spinless fermions, with total charge $Q_4 = 2$



If bond $l=2$ is truncated (e.g. dropping $Q_2=0$), that causes missing states on all later bonds!

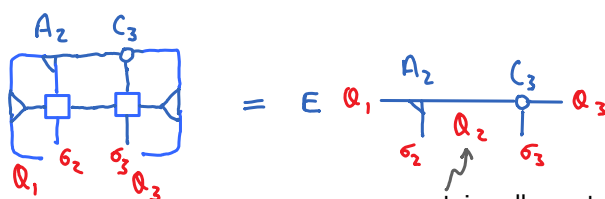
1s DMRG:



updated C_3 has same bond dimensions as initial C_3 (13)

1-site optimization of truncated C_3 will never find a good ground state if latter has non-negligible contributions from missing blocks.

2s DMRG:



updated A_2, C_3 can have larger bond dimensions than initial A_2, C_3 (14)

contains all quantum numbers consistent with $Q_1 + \sigma_2 = Q_2 = Q_3 - \sigma_2$

2-site optimization can reinstate missing blocks!

Which part of 2-site space is missed by H_ℓ^{1s} and $H_{\ell+1}^{1s}$?

Recall kept+discarded decomposition:

$$\mathcal{D} + \bar{\mathcal{D}} = d\mathcal{D}, \quad \bar{\mathcal{D}} = \mathcal{D}(d-1)$$

Orthonormality of kept and discarded isometries:

$$\frac{A_\ell}{\mathcal{D}} \frac{\bar{A}_\ell}{\bar{\mathcal{D}}} \oplus \frac{\bar{A}_\ell}{\bar{\mathcal{D}}} \frac{A_\ell}{\mathcal{D}} = \frac{A_\ell^\dagger A_\ell}{\mathcal{D} \mathcal{D}}, \quad \frac{B_\ell}{\mathcal{D}} \frac{\bar{B}_\ell}{\bar{\mathcal{D}}} \oplus \frac{\bar{B}_\ell}{\bar{\mathcal{D}}} \frac{B_\ell}{\mathcal{D}} = \frac{B_\ell^\dagger B_\ell}{\mathcal{D} \mathcal{D}} \quad (15)$$

$$A_\ell^\dagger A_\ell = \begin{matrix} A_\ell \\ \bigcap \\ A_\ell^* \end{matrix} = \left(= \mathbb{1}_\ell^K, \quad B_\ell B_\ell^\dagger = \begin{matrix} B_\ell \\ \bigcap \\ B_\ell^* \end{matrix} = \right) = \mathbb{1}_{\ell-1}^K \quad (16a)$$

$$\begin{matrix} \bigcap \\ \ell \end{matrix} = \left(= \mathbb{1}_\ell^D, \quad \begin{matrix} \bigcap \\ \ell \end{matrix} = 0, \quad \begin{matrix} \bigcap \\ \ell \end{matrix} = \right) = \mathbb{1}_{\ell-1}^D, \quad \begin{matrix} \bigcap \\ \ell \end{matrix} = 0 \quad (16b)$$

Completeness:

$$\begin{matrix} \bigcap \\ \ell \end{matrix} + \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \mathbb{1}_\ell^P, \quad \begin{matrix} \bigcap \\ \ell \end{matrix} + \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \mathbb{1}_{\ell-1}^P \quad (16c)$$

Compare action of 1-site and 2-site Hamiltonians:

$$\begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} \oplus \begin{matrix} \bigcap \\ \ell \end{matrix} \quad (17a)$$

$$\begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} \oplus \begin{matrix} \bigcap \\ \ell \end{matrix} \quad (17b)$$

$$\begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} = \begin{matrix} \bigcap \\ \ell \end{matrix} \oplus \begin{matrix} \bigcap \\ \ell \end{matrix} \oplus \begin{matrix} \bigcap \\ \ell \end{matrix} \oplus \begin{matrix} \bigcap \\ \ell \end{matrix} \quad (17c)$$

This part is not captured by 1s DMRG

This can also seen by considering energy variance:

$$\Delta_E^{1\perp} = \sum_{\ell=1}^{\mathcal{L}} \left\| \begin{matrix} \bigcap \\ \ell \end{matrix} \right\|^2 = \|P^\perp H \psi\|^2, \quad (18a)$$

$$\Delta_E^{2\perp} = \sum_{\ell=1}^{\mathcal{L}-1} \left\| \begin{matrix} \bigcap \\ \ell \end{matrix} \right\|^2 = \|P^\perp H \psi\|^2, \quad (18b)$$

Minimized by 1s DMRG, vanishes for converged 1s-GS.

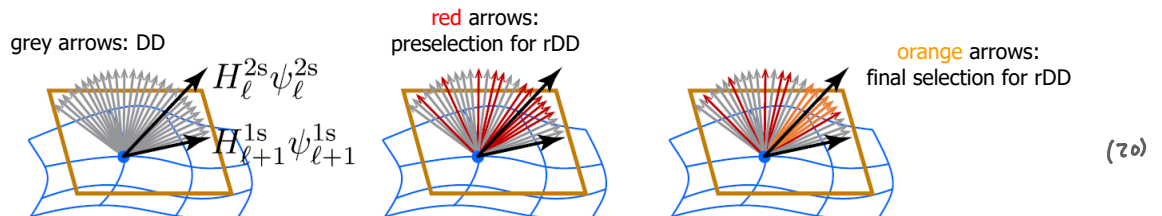
Minimized by 2s DMRG, vanishes for converged 2s-GS.

Subspace missed by 1s DMRG but explored by 2s DMRG is the DD subspace: $\text{image}(\bar{A}_\ell \otimes \bar{B}_{\ell+1}(\nabla \otimes \nabla))$

It contains 'missing' symmetry sectors (good!), but is huge (bad!)

dimension: $\bar{\mathcal{D}} \times \bar{\mathcal{D}} = \text{huge!}$ (19)

Key insight [Gleis2022]: $H_\ell^{2s} \psi_\ell^{2s}$ has significant weight only on small subspace of DD, the 'relevant DD' (rDD) !



Hence, it suffices to expand bond to include only the rDD ! View rDD as $\text{image}(\tilde{A}_\ell^{\text{tr}} \otimes \bar{B}_{\ell+1}(\nabla \otimes \nabla))$ or $\text{image}(\bar{A}_\ell \otimes \tilde{B}_{\ell+1}^{\text{tr}}(\nabla \otimes \nabla))$

Truncated isometries ∇ or ∇ can be found via 'shrewd selection' = (i) preselection, then (ii) final selection (see next section)

(i) Compute truncated isometry $\tilde{A}_\ell^{\text{tr}}$ (🔴)

(ii) expand bond ℓ :

$$\text{replace } \frac{A_\ell}{D_d} \rightarrow \frac{A_\ell}{D_d} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} = \frac{A_\ell^{\text{ex}}}{D_d \tilde{D} + \tilde{D}} , \quad C_{\ell+1} \rightarrow \frac{C_{\ell+1}^{\text{ex},i}}{D_d \tilde{D} + \tilde{D}} = \text{initial } C_{\ell+1} \quad (21)$$

so that initialized version of expanded bond = old bond:

$$A_\ell^{\text{exp}} C_\ell^{\text{exp}} = A_\ell C_\ell$$

$$\frac{A_\ell}{D_d} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} = \frac{A_\ell}{D_d} C_{\ell+1} \quad (22)$$

since $\frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} = 0$

and construct expanded 1s Hamiltonian:

$$H_{\ell+1}^{1s,\text{ex}} = \left[\frac{A_\ell}{D_d} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} \right]_{\ell+1} = \frac{D + \tilde{D}}{D} \left[\frac{d}{\ell+1} \right] \quad (23)$$

(iii) Find GS of expanded 1s Hamiltonian:

(e.g. Lanczos eigensolver), as in 1s DMRG:

$$\left[\frac{A_\ell}{D_d} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} \right]_{\ell+1} = E \frac{\tilde{C}_{\ell+1}^{\text{exp}}}{\tilde{C}_{\ell+1}} \quad (24)$$

(iv) Shift isometry center from $\ell+1$ to ℓ :

$$\frac{A_\ell}{D_d} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D_d \tilde{D}} \xrightarrow{\text{SVD, truncate}} \frac{\tilde{C}_\ell}{D_d} \frac{\tilde{B}_{\ell+1}}{D_d} \quad (25)$$

The truncated weight at step (iv), say ξ , serves as error measure.

In practice: suppose we want to gradually grow the bond dimension by a factor α per sweep. Then, for each update, we need to increase bond dimension from an initial D_i to a final $D_f = \alpha D_i$, with $\alpha > 1$. (26)

Thus, expand from D_i to $D_i + \tilde{D} = D_f(1 + \delta)$, with $\delta > 0$ (27)

and in (iv), truncate from $D_f(1 + \delta)$ to D_f . (28)

Typical choices: $\alpha = 1.1$, $\delta = 0.1$. (29)

Note: (i) to (iii) constitute a strictly variational algorithm, guaranteed to lower the energy.

However, (iv) involves truncation, which can lead to slight energy increase.

Hence the CBE algorithm, just as 2s-DMRG, is not strictly variational. [McCulloch2024]

3. Shrewd selection

CBE.3

Goal: truncate $\bar{A}_\ell(\nabla) \rightarrow \tilde{A}_\ell^{\text{tr}}(\nabla)$ to minimize $C_1 = \left\| \begin{array}{c} \text{orthogonal complement} \\ \text{truncated complement} \end{array} \right\|$ (1)

Optimal truncation can be achieved via SVD; but that has 2s costs, $\mathcal{O}(D^3 d^3)$

Instead, use 'shrewd selection' (cheap, efficient, practical, though not strictly optimal), involving two steps:

(i) Preselection: truncate $\bar{A}_\ell(\nabla) \rightarrow \hat{A}_\ell^{\text{pr}}(\nabla)$ to minimize $C_2 = \left\| \begin{array}{c} \text{orthogonal complement} \\ \text{preselected complement} \end{array} \right\|$ (2)

Truncate central bond in presence of its environment, with MPO bond open (to reduce numerical costs)

(ii) Final selection: truncate $\hat{A}_\ell^{\text{pr}}(\nabla) \rightarrow \tilde{A}_\ell^{\text{tr}}(\nabla)$ to minimize $C_3 = \left\| \begin{array}{c} \text{preselected complement} \\ \text{truncated complement} \end{array} \right\|$ (3)

Truncate again, now in with MPO bond closed, as appropriate for $H_\ell^{\text{zs}} \psi_\ell^{\text{zs}}$

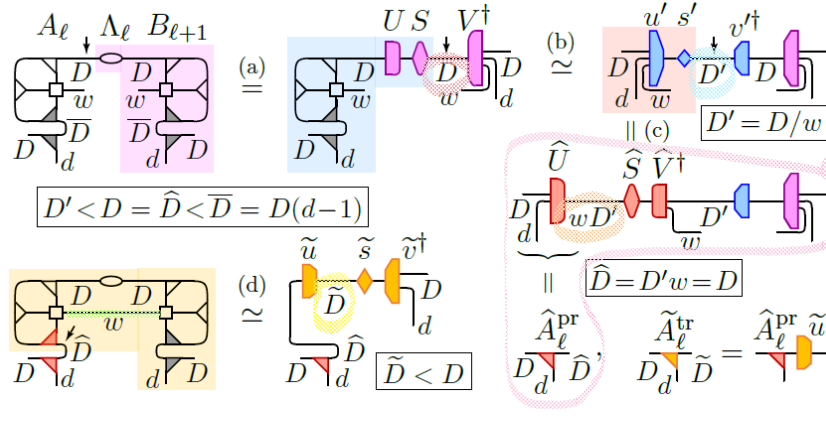
Details of preselection [steps (a-c)] and final selection [step (d)]:

- arrows indicate bond being opened before doing SVD

- shading and symbols in matching colors indicate SVD input and output

- output is written as $USV^\dagger$ or $usv^\dagger$ when involving no or some truncation, respectively

- use $\begin{array}{c} \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array} - \begin{array}{c} \text{---} \\ \text{---} \end{array}$
 $\begin{array}{c} \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array} - \begin{array}{c} \text{---} \\ \text{---} \end{array}$



(a) Canonicalize right side (shaded pink) of diagram, assigning its weights to central MPS bond.

(b) Truncate central MPS bond, $D \rightarrow D' = D/w$ (reason for this choice: see (d))

(c) Regroup, to combine truncated MPS bond and MPO bond into composite bond of dimension $\hat{D} = D'w = D$

If using exact arithmetic, this would involve no truncation. In practice (numerically) zero singular values

$\mathcal{O}(10^{-16})$, may arise. They must be truncated to ensure $\begin{array}{c} \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \end{array}$, so that $\text{image}(\nabla) \subset \text{image}(\nabla)$.

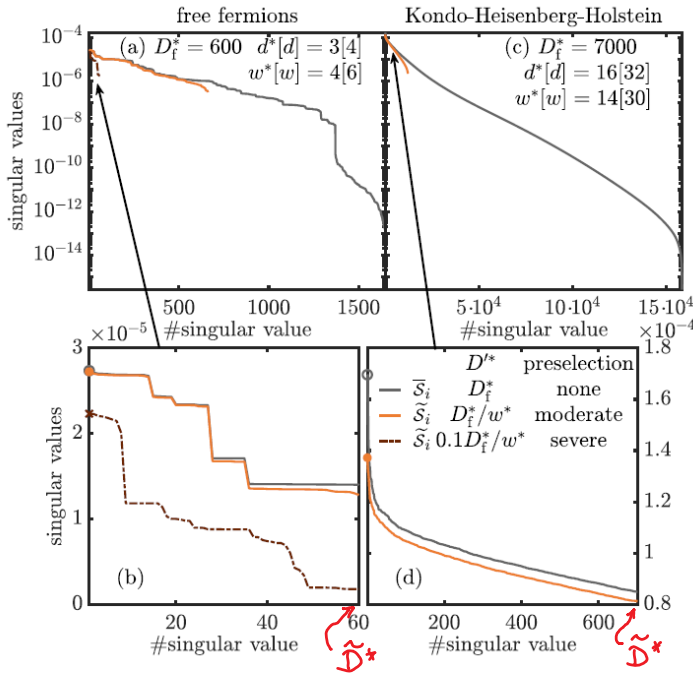
(d) Final selection: close MPO bond, then truncate central MPS bond: $\hat{D} \rightarrow \tilde{D} < D$ (e.g. $\tilde{D} = 0.1 D$).

To ensure 1s costs for this step, we need $D'w = \hat{D} = D$, hence choose $D' = D/w$ in (b).

Important: By design, every step has at most 1s costs, $\mathcal{O}(D^3 dw)$

Moreover, CBE captures the most most relevant contributions from H_ℓ^{zs}

2s accuracy and convergence per sweep, at 1s cost !!



Comparison of three truncation settings:

grey: optimal truncation via SVD (grey)

serves as a reference

orange: moderate preselection, $D_f^* = D_f^*/w^*$

then final selection $\tilde{D}^* = 0.1 D_f^*$

agrees rather well with reference!

brown: severe preselection, $D_f^* = 0.1 D_f^*/w^*$

then final selection $\tilde{D}^* = 0.1 D_f^*$

misses some information from reference

Take-home message: optimal truncation requires computation of a huge amount of singular values, most of which are discarded anyway. Those that are kept can be very well captured using shrewd selection!



It's not necessary to line up all kids from tallest to shortest if you just need a few big ones to help you with something!

Results for CBE-DMRG:

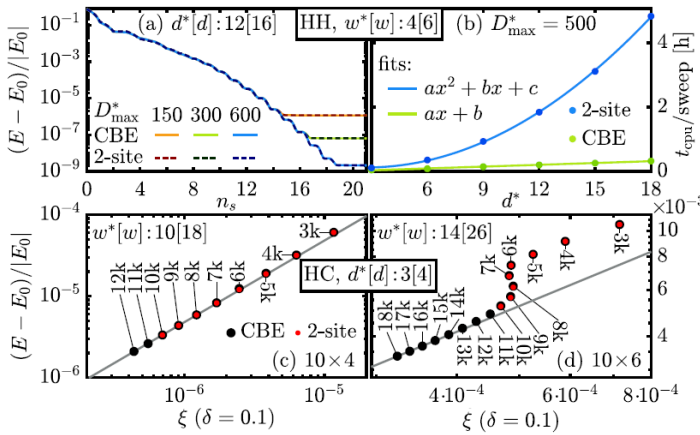


FIG. 3. Hubbard-Holstein (HH) model: (a) Convergence of the GS energy versus number of half-sweeps n_s at fixed $d^* = 3(N_{\text{ph}}^{\text{max}} + 1)$. E_0 was obtained by linear ξ extrapolation of data from $D_{\text{max}}^* \in [1000, 1200]$. (b) CPU time per sweep for various D_{max}^* at fixed D_{max}^* , showing d^* (CBE) vs d^{*2} (2s) scaling. Hubbard cylinders (HC): Error in GS energy vs ξ for (c) 10×4 and (d) 10×6 HCs, obtained with CBE (black) and 2s (red) DMRG, for various D_{max}^* (legends). Since 2s CPU times far exceed those of CBE, 2s data is only shown for $D_{\text{max}}^* \leq 10k$. Reference energies $E_0 = -27.8816942$ (10×4) and -41.7474961 (10×6) are obtained by linear ξ extrapolation of the four most accurate CBE results to $\xi = 0$ (gray line).

$$H_{\text{HH}} = -\sum_{\ell\sigma} (c_{\ell\sigma}^\dagger c_{\ell+1\sigma} + \text{H.c.}) + 0.8 \sum_{\ell} n_{\ell\uparrow} n_{\ell\downarrow} + 0.5 \sum_{\ell} b_{\ell}^\dagger b_{\ell} + \sqrt{0.2} \sum_{\ell} (n_{\ell\uparrow} + n_{\ell\downarrow} - 1) \times (b_{\ell}^\dagger + b_{\ell})$$

$$H_{\text{HC}} = -\sum_{\langle \ell, \ell' \rangle, \sigma} (c_{\ell\sigma}^\dagger c_{\ell'\sigma} + \text{H.c.}) + 8 \sum_{\ell} n_{\ell\uparrow} n_{\ell\downarrow}$$

(a) CBE and 2s DMRG have same convergence rate per sweep.

(b) CBE has 1s costs $\sim \mathcal{O}(d)$, much faster than 2s DMRG $\sim \mathcal{O}(d^2)$

(c,d) Reliable convergence with increasing D^* , decreasing ξ

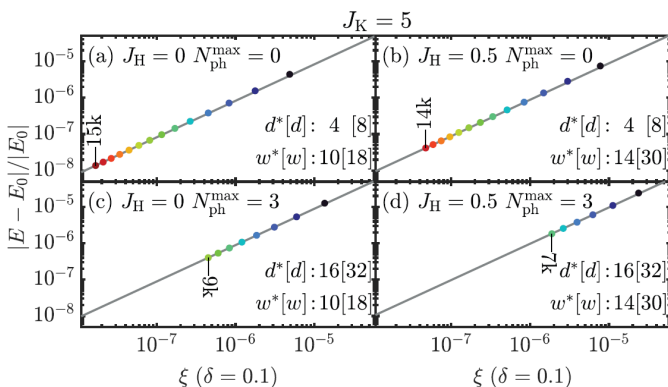


FIG. S-10. Error in GS energy versus discarded weight for the Kondo-Heisenberg-Holstein (KHH) model on a 10×4 cylinder, with (a) only Kondo coupling, (b) Kondo and Heisenberg coupling, (c) Kondo and Holstein coupling and (d) Kondo, Heisenberg and Holstein coupling. Legends state our choices for J_H and $N_{\text{ph}}^{\text{max}}$, and corresponding values of $d^*[d]$ and $w^*[w]$.

$$H_{\text{KH}} = -\sum_{\langle \ell, \ell' \rangle, \sigma} (c_{\ell\sigma}^\dagger c_{\ell'\sigma} + \text{H.c.}) + J_K \sum_{\ell} \mathbf{S}_{\ell} \cdot \mathbf{S}_{\ell} + \frac{1}{2} \sum_{\langle \ell, \ell' \rangle} \mathbf{S}_{\ell} \cdot \mathbf{S}_{\ell'}$$

J_K tunes quantum phase transition between two phases with different Fermi surface volumes.

Recall 1s TDVP:

Schrödinger equation for MPS:

$$i \frac{d}{dt} |\tilde{\psi}[M(t)]\rangle = \hat{P}^{1s} \hat{H} |\tilde{\psi}[M(t)]\rangle \quad (1)$$

$$\Psi = \begin{array}{c} A_1 \quad A_2 \quad \dots \quad A_\ell \quad C_{\ell+1} \quad B_{\ell+2} \quad \dots \quad B_{\mathcal{L}-1} \quad B_{\mathcal{L}} \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array} \quad (2)$$

$$\mathcal{P}^{1s} = \sum_{\ell'=1}^{\mathcal{L}} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array} \quad (3)$$

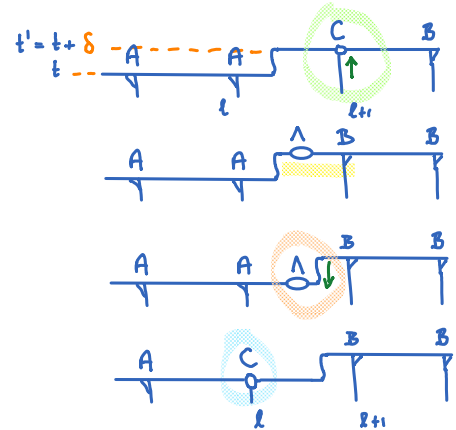
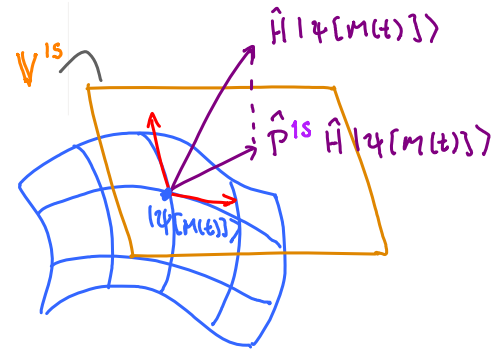
1s TDVP algorithm (sweeping right-to-left):

(1) Integrate $i \dot{C}_{\ell+1} = H_{\ell+1}^{1s} C_{\ell+1}$ from $t \rightarrow t' = t + \delta$

(2) QR factorize $C_{\ell+1}(t') = \Lambda_{\ell+1}(t') B_{\ell+1}(t')$

(3) Integrate $i \dot{\Lambda}_{\ell} = -i H_{\ell}^b \Lambda_{\ell}$ from $t' \rightarrow t$

(4) Update $A_{\ell}(t) C_{\ell+1}(t) \rightarrow A_{\ell}(t) \Lambda_{\ell}(t) B_{\ell+1}(t') \equiv C_{\ell}(t)$



Advantages of 1s TDVP: applicable to long-ranged Hamiltonians, numerical stability, unitary time-evolution, energy conservation (because truncation happens before, not after, time step!)

1s-TDVP has two leading errors:

(i) Lie-Trotter error, can be reduced by higher-order integration schemes, e.g. third-order, with error $\mathcal{O}(\delta^3)$

(ii) Projection error, quantified by $\Delta_P = \|(1 - \hat{P}^{1s}) \hat{H} \psi(t)\|$ (4)

Projection error can be reduced by using 2s TDVP,

$$i \frac{d}{dt} |\tilde{\psi}[M(t)]\rangle = \hat{P}^{2s} \hat{H} |\tilde{\psi}[M(t)]\rangle \quad (5)$$

Then projection error becomes

$$\Delta_P = \|(1 - \hat{P}^{2s}) \hat{H} \psi(t)\|$$

$$\mathcal{P}^{2s} = \sum_{\ell=1}^{\mathcal{L}-1} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array} \quad (6)$$

However, after time step, another truncation is needed to bring down bond dimension from Dd to D .

This truncation-after-time-step leads to non-unitary time-evolution, non-conservation of energy.

CBE-TDVP

Key idea: use CBE to reduce 2s contribution to Δ_P , given by $\Delta_P^{2\perp} = \|\hat{P}^{2s} (1 - \hat{P}^{1s}) \hat{H} \psi(t)\|$ (7)

$$\mathcal{P}^{2\perp} = \sum_{\ell=1}^{\mathcal{L}-1} \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array}, \quad \Delta_P^{2\perp} = \sum_{\ell=1}^{\mathcal{L}-1} \left\| \begin{array}{c} \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \quad \text{---} \end{array} \right\|^2 \quad (8)$$

finds that part of $\hat{H} \psi$ that is not in 1s space but in 2s space

$\Delta_P^{2\perp}$ is the same object is that minimized for CBE-DMRG! Hence, CBE is also useful here!

We add just one step (0) to 1s-TDVP algorithm (when sweeping right-to-left), using:

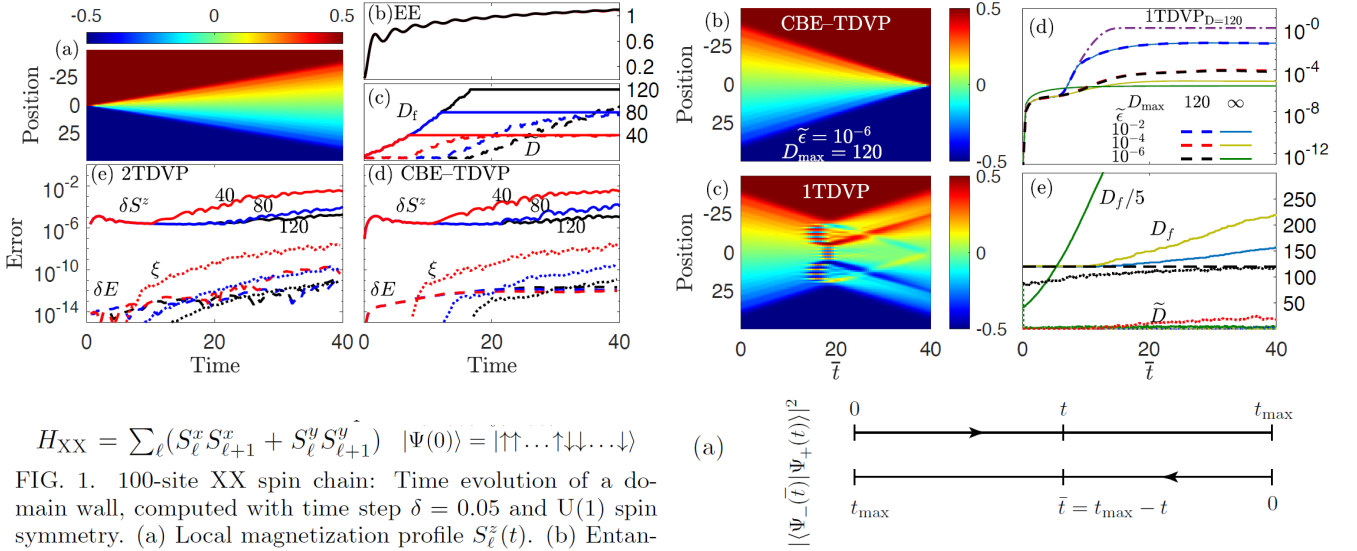
Step (0): Expand $D \rightarrow D + \tilde{D}$ for bond ℓ , using $A_{\ell} \rightarrow A_{\ell}^{ex}$, $C_{\ell+1} \rightarrow C_{\ell+1}^{exp}$, $H_{\ell+1}^{1s} \rightarrow H_{\ell+1}^{1s, exp}$

$$\frac{A_\ell}{D} \underset{d}{\nwarrow} \underset{d}{\nearrow} D \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D} \underset{d}{\nwarrow} \underset{d}{\nearrow} \tilde{D} = \frac{A_\ell^{\text{ex}}}{D} \underset{d}{\nwarrow} \underset{d}{\nearrow} (D + \tilde{D}) \underset{d}{\nwarrow} \underset{d}{\nearrow} C_{\ell+1}^{\text{ex}} = \left[\text{circuit} \right]_{\ell+1}, \quad (9)$$

$$H_{\ell+1}^{(1,\text{ex})} = \left[\text{circuit} \right]_{\ell+1} = \left[\text{circuit} \right]_{\ell+1}^{D+\tilde{D}}. \quad (10)$$

Other steps remain as before, except that in step (2), QR factorization is replaced by SVD, to 'trim bond dimension from $D + \tilde{D}$ to final value D_f , chosen such that truncation error is $< 10^{-12}$ (for early times), or such that $D_f = D_{\text{max}}$ (for later times, to limit computational costs). Trimming error is characterized by discarded weight, $\xi(t)$, which can be controlled or monitored. TDVP properties of unitary time evolution and energy conservation hold within $\mathcal{O}(\xi(t))$.

Benchmarking CBE-TDVP for exactly solvable XX model:



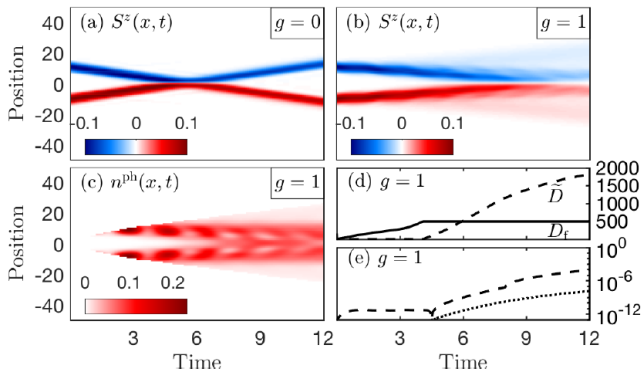
$$H_{\text{XX}} = \sum_\ell (S_\ell^x S_{\ell+1}^x + S_\ell^y S_{\ell+1}^y) \quad |\Psi(0)\rangle = |\uparrow\uparrow \dots \uparrow\downarrow \dots \downarrow\rangle$$

FIG. 1. 100-site XX spin chain: Time evolution of a domain wall, computed with time step $\delta = 0.05$ and U(1) spin symmetry. (a) Local magnetization profile $S_\ell^z(t)$. (b) Entanglement entropy $EE(t)$ between the left and the right half of the chain. (c) Bond dimension $D_f(t)$ and its pre-trimming expansion $\tilde{D}(t)$ per time step, for $D_{\text{max}} = 120$. (d,e) Error analysis: magnetization $\delta S_\ell^z(t)$ (solid line), i.e., the maximum deviation (over ℓ) of $S_\ell^z(t)$ from the exact result, energy $\delta E(t)$ (dashed line), and discarded weight $\xi(t)$ (dotted line) for $D_{\text{max}} = 40$ (red), 80 (blue) and 120 (black), computed with (d) CBE-TDVP or (e) 2TDVP. Remarkably, the errors are comparable in size, although CBE-TDVP has much smaller computational costs.

$$F(\bar{t}) = |\langle \Psi_-(\bar{t}) | \Psi_+(t) \rangle|^2, \quad \bar{t} = t_{\text{max}} - t \in [0, t_{\text{max}}]$$

FIG. S-1. (a) Forward-backward time evolution for the computation of $F(t)$. (b,c) Back-evolution of the domain wall, described by $|\Psi_-(\bar{t})\rangle$, computed using (b) CBE-TDVP and (c) 1TDVP. (d) Time evolution of $\delta F(\bar{t}) = 1 - F(\bar{t})$, computed via 1TDVP with $D = 120$ (dash-dotted line), and via CBE-TDVP using three values of $\tilde{\epsilon}$, and either with $D_{\text{max}} = 120$ (dashed lines) or $D_{\text{max}} = \infty$ (solid lines). (e) Time evolution of the corresponding bond dimensions $D_f(\bar{t})$ (solid lines) and $\tilde{D}(\bar{t})$ (dots). (The solid green curve shows $D_f/5$.)

Phonon-induced pair attraction during electron-electron scattering



$$H_{\text{PH}} = \sum_\ell U n_{\ell\uparrow} n_{\ell\downarrow} + \sum_\ell \omega_{\text{ph}} b_\ell^\dagger b_\ell$$

$$+ \sum_{\ell\sigma} (c_{\ell\sigma}^\dagger c_{\ell+1\sigma} + \text{h.c.}) (-t + b_\ell^\dagger + b_\ell - b_{\ell+1}^\dagger - b_{\ell+1})$$

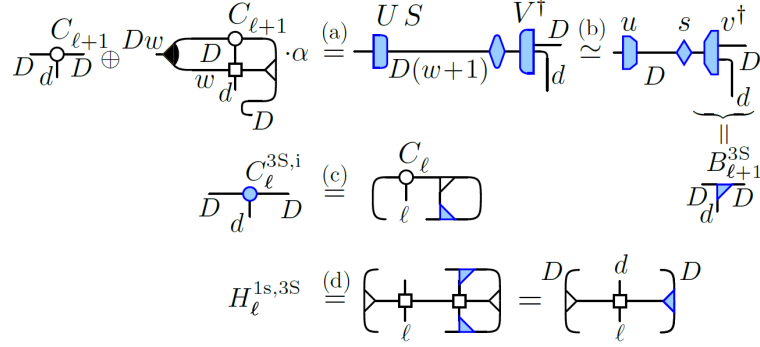
$$n_{\text{ph}}^{\text{max}} = 8, \quad d = 4(n_{\text{ph}}^{\text{max}} + 1) = 36$$

FIG. 4. Peierls-Hubbard model: Real-space scattering of two electron wave packets, for $U=10$ and $\omega_{\text{ph}}=3$, computed with $\delta = 0.05$, $n_{\text{ph}}^{\text{max}} = 8$ and U(1) spin symmetry. (a,b) Spin magnetic moment $S^z(x,t)$ for $g=0$ and $g=1$. (c) Phonon density $n^{\text{ph}}(x,t)$, (d) bond dimensions, and (e) error analysis: energy $\delta E(t)$ (dashed line) and discarded weight $\xi(t)$ (dotted line), all computed for $g=1$, with $D_{\text{max}} = 500$.

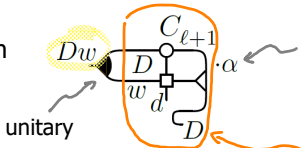
- (a) Without electron-phonon coupling, two wave packets bounce off each other due to strong U repulsion.
- (b) With electron-phonon coupling, the wave packets tend to stick together, while (c) phonons get activated.

Another scheme for bond expansion with 'strictly single-site (3S) costs was proposed by [Hubig2015]. The states to be added are generated by acting with 'part of a single-site' Hamiltonian onto $C_{\ell+1}$:

During right-to-left sweep:



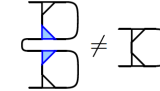
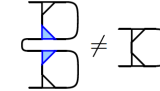
The bond between sites ℓ and $\ell+1$ is expanded from D to $D + Dw$

through a direct sum with  'mixing' parameter, typically < 1 , which has to be adapted during sweeping (see Sec. VI of [Hubig2015])

This 'enriches' the left bond with states generated by

(a) SVD to identify the dominant bond states.

(b) Truncate back to original bond size, $D + Dw \rightarrow D$ and identify $D_d^T D$

(c,d) Use $D_d^T D$ to expand $C_{\ell}^{3S,i}$ and $H_{\ell}^{1S,3S}$, as in step (ii) of CBE.  has to be normalized explicitly, since 

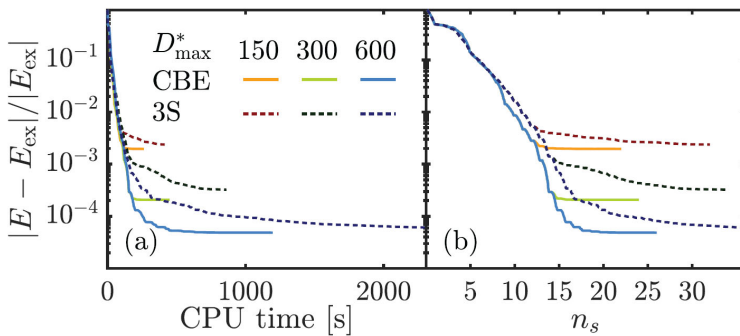
Finally, update C_{ℓ}^{3S} with ground state of $H_{\ell}^{1S,3S}$

Mixing with Dw new states can bring in good quantum numbers (good!),

and do so efficiently if α is large (e.g. $\alpha = 1$).

Step (b) can eliminate bad quantum numbers (good!).

But since no D projection is involved, some K states that are actually useful for ground state can also be truncated, hence (b) can actually cause the ground state energy to increase (not so good!), by an amount depending on α . Further sweeping will push energy back down.



CBE and 3S initially converge at the same rate w.r.t. CPU time, but 3S eventually slows down and takes longer to reach final convergence compared to CBE.

FIG. S-8. Error in energy for spinful free fermions on a 10×4 cylinder versus (a) CPU time and (b) number of half-sweeps n_s . E_{ex} is the exact ground-state energy.

But: 3S with strong mixing (e.g. $\alpha = 1$) can be more efficient than CBE for getting out of a 'wrong' minimum.

So, we recommend [Gleis2022, Section S-3(B), final paragraph] to use CBE + mixing (CBE + α), to combine strengths of both methods:

- 3S with strong mixing facilitates efficiently escaping from metastable minima;
- CBE yields rapid descent to bottom of close-by accessible minima.

CBE + mixing [for right-to-left sweep]

(i) - (iii): Just as for pure CBE: (i) Compute $\tilde{A}_\ell^{\text{tr}}$ (). (ii) Expand bond ℓ :

$$\text{replace } \frac{A_\ell}{D \begin{array}{|c|} \hline d \\ \hline \end{array} D} \rightarrow \frac{A_\ell}{D \begin{array}{|c|} \hline d \\ \hline \end{array} D} \oplus \frac{\tilde{A}_\ell^{\text{tr}}}{D \begin{array}{|c|} \hline d \\ \hline \end{array} \tilde{D}} = \frac{A_\ell^{\text{ex}}}{D \begin{array}{|c|} \hline d \\ \hline \end{array} \tilde{D} + \tilde{D}}, \quad \begin{array}{c} \tilde{C}_{\ell+1} \\ \downarrow \\ \tilde{D} + \tilde{D} \end{array} \rightarrow \begin{array}{c} C_{\ell+1}^{\text{ex},i} \\ \downarrow \\ \tilde{D} \end{array} \overset{\text{initial}}{=} \begin{array}{c} C_{\ell+1} \\ \downarrow \\ \tilde{D} \end{array}$$

and construct expanded 1s Hamiltonian:

$$H_{\ell+1}^{1s,\text{ex}} = \begin{array}{c} \begin{array}{|c|} \hline \tilde{D} + \tilde{D} \\ \hline \end{array} \\ \downarrow \\ \begin{array}{|c|} \hline d \\ \hline \end{array} \end{array} = \begin{array}{c} D + \tilde{D} \\ \downarrow \\ \begin{array}{|c|} \hline d \\ \hline \end{array} \end{array}$$

(iii) Find GS of expanded 1s Hamiltonian:

(e.g. Lanczos eigensolver), as in 1s DMRG:

$$\begin{array}{c} \tilde{C}_{\ell+1}^{\text{exp}} \\ \downarrow \\ \begin{array}{|c|} \hline \tilde{D} + \tilde{D} \\ \hline \end{array} \\ \downarrow \\ \begin{array}{|c|} \hline d \\ \hline \end{array} \end{array} = E \begin{array}{c} \tilde{C}_{\ell+1}^{\text{exp}} \\ \downarrow \\ \tilde{D} \end{array}$$

(iv) If $\alpha = 0$: Shift isometry center from $\ell+1$ to ℓ :
(just as for pure CBE), and skip (v)

$$\begin{array}{c} A_\ell \\ \downarrow \\ \tilde{D} + \tilde{D} \end{array} \begin{array}{c} \tilde{C}_{\ell+1}^{\text{exp}} \\ \downarrow \\ \tilde{D} \end{array} \xrightarrow{\text{SVD, truncate}} \begin{array}{c} \tilde{C}_\ell \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} \tilde{B}_{\ell+1} \\ \downarrow \\ \tilde{D} \end{array} \quad (\tau_S)$$

If $\alpha \neq 0$: SVD and truncate on bond ℓ
(but do not shift isometry center):

$$\begin{array}{c} A_\ell \\ \downarrow \\ \tilde{D} + \tilde{D} \end{array} \begin{array}{c} \tilde{C}_{\ell+1}^{\text{exp}} \\ \downarrow \\ \tilde{D} \end{array} \xrightarrow{\text{SVD, truncate}} \begin{array}{c} \tilde{A}_\ell \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} \tilde{C}_{\ell+1}^S \\ \downarrow \\ \tilde{D} \end{array}$$

(v) Only if $\alpha \neq 0$:

Perform 3S mixing with α on bond ℓ
(but do not find another GS):

$$\begin{array}{c} \tilde{C}_{\ell+1}^S \\ \downarrow \\ \tilde{D} \end{array} \oplus \begin{array}{c} \tilde{C}_{\ell+1}^S \\ \downarrow \\ \tilde{D} \end{array} \cdot \alpha \xrightarrow{(a)} \begin{array}{c} U \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} S \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} V^\dagger \\ \downarrow \\ \tilde{D} \end{array} \xrightarrow{(b)} \begin{array}{c} u \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} s \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} v^\dagger \\ \downarrow \\ \tilde{D} \end{array} \xrightarrow{(c)} \begin{array}{c} \tilde{A}_\ell \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} \tilde{C}_{\ell+1}^S \\ \downarrow \\ \tilde{D} \end{array}$$

Then replace

$$\begin{array}{c} \tilde{A}_\ell \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} \tilde{C}_{\ell+1}^S \\ \downarrow \\ \tilde{D} \end{array} \rightarrow \begin{array}{c} C_\ell^{3S} \\ \downarrow \\ \tilde{D} \end{array} \begin{array}{c} B_{\ell+1}^{3S} \\ \downarrow \\ \tilde{D} \end{array}$$

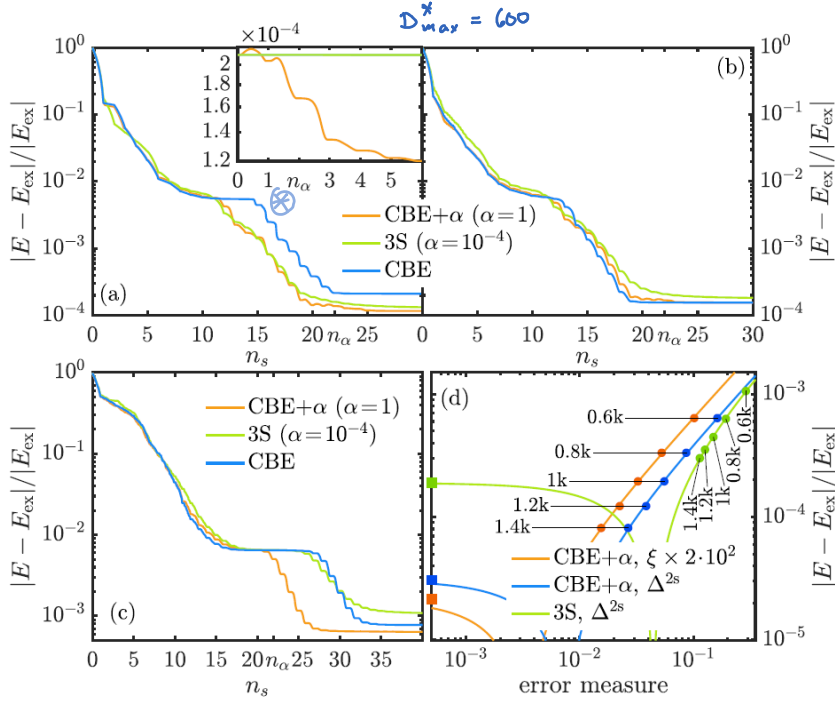
Now isometry center is located at site ℓ .

Use $\alpha = 1$ in bond-growing phase, then set $\alpha = 0$ for a few sweeps before measuring errors and observables.

To perform a series of computations at different bond dimensions $D_1 < D_2 < \dots$ to monitor convergence:

- Converge an D_1 MPS; then measure errors and observables.
- Converge a D_2 MPS: initialize it with the previous D_1 MPS,
then perform one or two sweeps with $\alpha = 1$ to escape metastable minima,
followed by a few sweeps with $\alpha = 0$ to reach bottom of new minimum,
then measure errors and observables.
- and so on for larger D_n .

This strategy leads to quick convergence while avoiding local minima.



(periodic boundary conditions)
 (a) 1D ring at half-filling: CBE gets stuck for a while in a metastable minimum $\otimes$: the ground state for open chain. CBE + α and 3S do not get stuck there.

Inset of (a):
 Initial state = CBE ground state for $D_{\text{max}}^* = 600$
 CBE + $\alpha = 1$ gets unstuck, while CBD and 3S with $\alpha = 10^{-4}$ remain stuck.

(b) 1D ring away from half-filling: CBE converges fastest.

(c,d) 10 x 40 cylinder:

(c) CBE + α converges fastest.

(d) For given bond dimension, CBE + α reaches lower energy than 3S

FIG. S-1. (a-d) Convergence of the GS energy versus number of half-sweeps n_s using different bond expansion methods at $D_{\text{max}}^* = 600$. Spinful free fermions on a $\mathcal{L} = 100$ ring with nearest-neighbor hopping at (a) half-filling, (b) filling $N = 0.9\mathcal{L}$. Inset in (a): attempt to converge the CBE MPS using pure 3S and CBE+ α (see text for details). (c) Spinful free fermions on a $\mathcal{L}_x \times \mathcal{L}_y = 10 \times 4$ cylinder with nearest-neighbor hopping at filling $N = 0.9\mathcal{L}_x\mathcal{L}_y$. For CBE+ α , we used $\alpha = 1$ for $n_s \leq n_\alpha = 22$ and $\alpha = 0$ for $n_s > n_\alpha$. (d) Energy versus error measure (Δ^{2s} : 2s variance, ξ : discarded weight) for CBE+ α and 3S at different D_{max}^* . Dots are data points, lines linear fits and squares indicate the extrapolated energy.

Conclusion: CBE + α yields very favorable convergence properties !