Mode optimized tensor network state methods and applications to strongly correlated systems

A journey from mathematical aspects towards industrial perspectives

Synergies among physics, chemistry, math and computer science

Örs Legeza

Strongly Correlated Systems "Lendület" Research Group
Wigner Research Centre for Physics, Budapest, Hungary

Institute for Advanced Study, Technical University of Munich, Germany

Parmenides Foundation, Pöcking, Germany

Entanglement and topology in interacting quantum matter

Benasque, 20.02.2026

in collaboration with with colleagues from academia

- Andor Menczer, Gergely Barcza, Szilárd Szalay, Imre Hagymási, Ádám Ganyecz, Mihály Máté, András Olasz, Tamás Mosoni
- Florian Gebhard, Reinhard M. Noack, Georg Ehlers (Marburg)
- Frank Verstraete, Klaas Gunst, S. Wooters, D. van Neck
- Jens Eisert, Christian Krumnow, Edoardo Fertitta, Beate Paulus
- Reinhold Schneider, Max Pfeffer, Gero Friesecke
- Libor Veis, Jiri Pittner, Jiri Brabec, Andrej Antalik, Jan Brandejs
- Frank Neese, Ali Alavi, Peter Saalfrank
- Gero Friesecke, Mi-Song Dupuy, Benedikt Graswald
- Alexander Tichai, Achim Schwenk, Takai Miyagi
- Simen Kvaal, Fabian Faulstich, Andre Laestadius
- Moca Pascu, Miklós Werner, Kornél Kapás, Gergely Zaránd
- Uli Schollwöck, Martin Grundner, Sam Mardazad, Claudius Hubig
- Karol Kowalski, Sotiris Xantheas, ... (incomplete list)

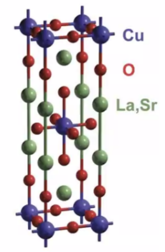
in collaboration with

- Our computer program package is used by more than 30 research groups worldwide for more than two decades in condensed matter physics, quantum chemistry, nuclear physics, quantum information theory, applied mathematics and computer science, etc...
- High-Performance Computing Center Stuttgart, Germany
- National Energy Research Scientific Computing Center (NERSC), USA

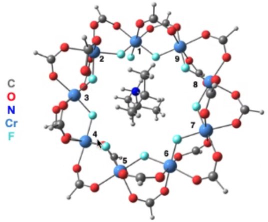
Recently there is also an increasing interest by industrial partners:

- NVIDIA, USA
- AMD, USA
- SandboxAQ, USA (Google startup)
- Riverlane LTD, UK
- Furukawa Electric Institute of Technology, Japan
- IBM, USA
- FACCTS, Germany
- Dynaflex LTD, Hungary

Strong correlations between electrons used by nature and in new technologies



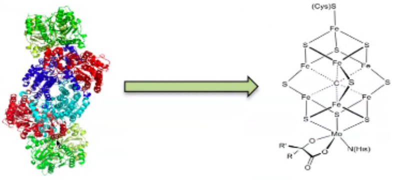
High T_{c} superconductors



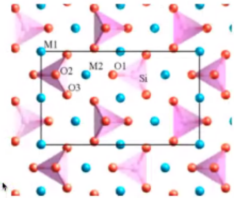
Lee, Small & Head-Gordon, *JCP*, **2018**, *149*, 244121

Single molecular magnets (SMM)

$N_2 + 3\,H_2 \rightarrow 2\,NH_3$ "FeMoco"

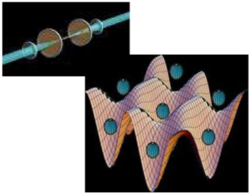


Nitrogen fixation



Battery technology

Experimental realizations: optical lattices
Numerical simulations: model systems



Atoms (represented as blue spheres) pictured in a 2D-optical lattice potential

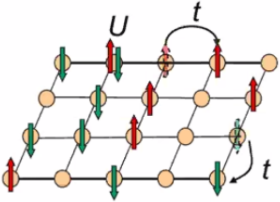
Potential depth of the optical lattice can be tuned.

Periodicity of the optical lattice can be tuned.

Hubbard model: lattice model of interacting electron system

$$H = t \sum_{\langle i,j \rangle,\sigma} c^{\dagger}_{i,\sigma} c_{j,\sigma} + \frac{U}{2} \sum_{\sigma \neq \sigma'} \sum_{i} n_{i,\sigma} n_{i,\sigma'}$$

t hopping amplitude
U on-site Coulomb interaction
$\sigma \in \uparrow, \downarrow$ spin index



Classical or quantum computers?

Topics to be covered

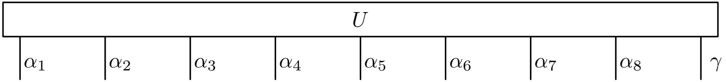
1. Tensor product factorization (mathematically exact, loop free):
2. Lattice models and ab initio framework
3. Adaptiv basis optimization: Fermionic mode transformation
   - 2D Hubbard-like systems with adaptive mode transformation
   - Strongly correlated molecular clusters
   - Long time evolution with adaptive mode transformation
4. Basis update but Galerkin (BUG): novel time evolution method
   - (2+1)D Z_2 Higgs model on the honeycomb lattice
   Quantum simulation on Q-chip and via TNS
5. Embedding methods: Capturing static and dynamic correlations
   - Self Consistent Field DMRG (DMRG-SCF)
6. Parallelization strategies, mixed precision arithmetic, exascale

Tensor product approximation

State vector of a quantum system in the discrete tensor product spaces

$$|\Psi_\gamma\rangle = \sum_{\alpha_1=1}^{q_1} \cdots \sum_{\alpha_d=1}^{q_d} U(\alpha_1, \ldots, \alpha_d, \gamma)\, |\alpha_1\rangle \otimes \cdots \otimes |\alpha_d\rangle \in \bigotimes_{i=1}^{d} \Lambda_i := \bigotimes_{i=1}^{d} \mathbf{C}^{q_i} \ ,$$

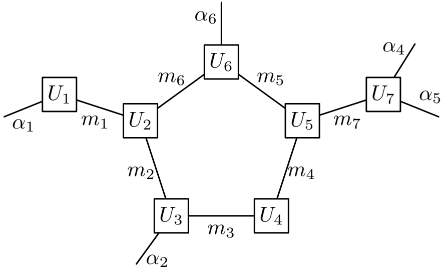
where $span\{|\alpha_i\rangle : \alpha_i = 1, \ldots, q_i\} = \Lambda_i = \mathbf{C}^{q_i}$ and $\gamma = 1, \ldots, m$.



$\dim \mathcal{H}_d = \mathcal{O}(q^d)$ Curse of dimensionality! (exponential scaling)

We seek to reduce computational costs by parametrizing the tensors in some data-sparse representation.

A general tensor network representation of a tensor of order 5.



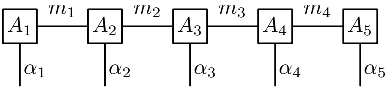
Matrix product state (MPS) representation / DMRG / TT
Exponential scaling $\rightarrow$ polynomial scaling

Affleck, Kennedy, Lieb Tagasaki (87); Fannes, Nachtergale, Werner (91), White (92)

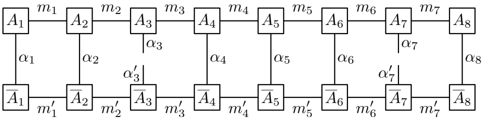
The tensor U is given elementwise as

$$U(\alpha_1, \ldots, \alpha_d) = \sum_{m_1=1}^{r_1} \cdots \sum_{m_{d-1}=1}^{r_{d-1}} A_1(\alpha_1, m_1) A_2(m_1, \alpha_2, m_2) \cdots A_d(m_{d-1}, \alpha_d).$$

We get d component tensors of order 2 or 3. Scaling: m^3.



Calculation of ρ_{ij} corresponds to the contraction of the network except at modes i and j.



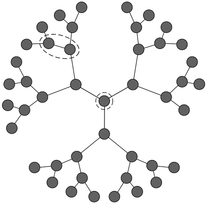
von Neumann quantum information entropy, $s = -\sum_\alpha \lambda_\alpha^2 \ln \lambda_\alpha^2$.

Mutual information, $I = s_i + s_j - s_{ij}$.

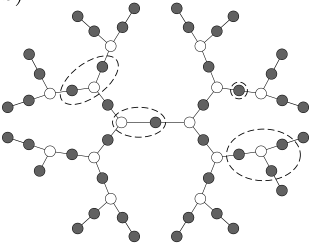
Ö.L & Sólyom, (03), Rissler, Noack, White (06)

T3NS a new tensor format Gunst, Verstraete, Wooters, Ö.L., van Neck (2018)

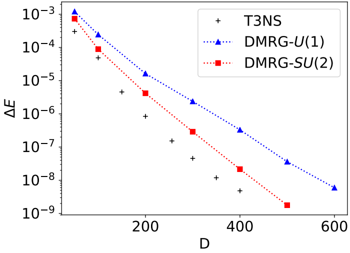
(a)



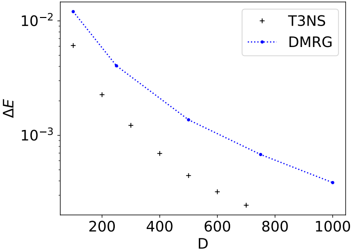
(b)



LiF



N_2



TNS/DMRG provide state-of-the-art results in many fields

$$\mathcal{H} = \sum_{ij\alpha\beta} T_{ij}^{\alpha\beta} c_{i\alpha}^{\dagger} c_{j\beta} + \frac{1}{2} \sum_{ijkl\alpha\beta\gamma\delta} V_{ijkl}^{\alpha\beta\gamma\delta} c_{i\alpha}^{\dagger} c_{j\beta}^{\dagger} c_{k\gamma} c_{l\delta} + \dots\,,$$

- T_{ij} kinetic and on-mode terms, V_{ijkl} two-particle scatterings
- We consider usually lattice models in real space (DMRG)
- In quantum chemistry modes are electron orbitals (QC-DMRG)
- In UHF QC spin-dependent inetractions (UHF-QCDMRG)
- In relativistic quantum chemistry modes are spinors (4c-DMRG)
- In nuclear problems modes are proton/neutron orbitals (JDMRG)
- In k-space modes are momentum eigenstates (k-DMRG)
- For particles in confined potential modes → Hermite polynomials
- Major aim: to obtain the desired eigenstates of $\mathcal{H}$.

• Symmetries: Abelian and non-Abelian quantum numbers, double groups, complex integrals, quaternion sym. etc

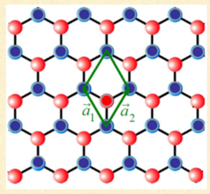
• # of block states: 1 000 – 60 000. Size of Hilbert space up to 10^8.

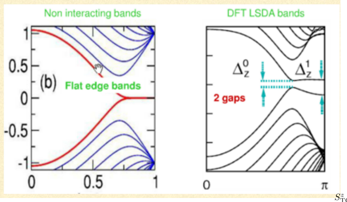
• In ab inito DMRG the CAS size is: 100 electrons on 100 orbitals.

• 1-BRDM and 2-BRDM, finite temperature, dynamics

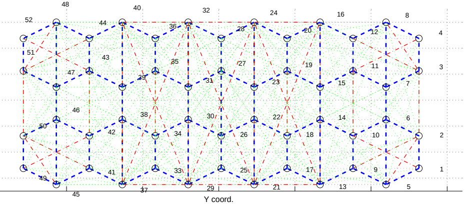
• Massively parallel implementations CPU/GPU→ exascale on HPC

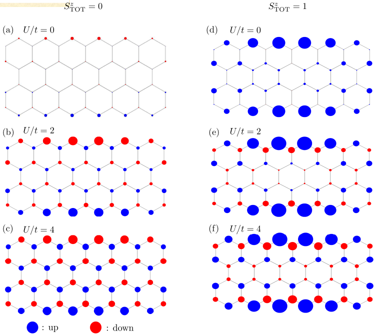
Example on graphene nanoribbons I. Hagymási, Ö.L (2016)





- Flat bands disappear when interaction is switched on
- Need TNS due to strong quantum fluctuations





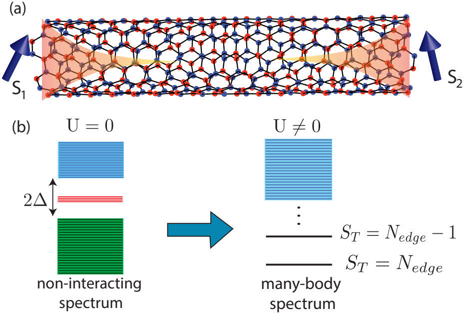
- DMRG $D = 20000$, $U = 2$

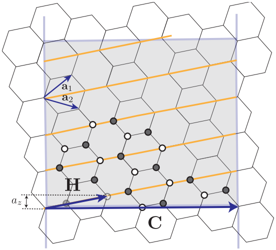
Problem revisited with modetransformation, monitoring emerging modes for zigzag, armchair, periodic BC etc.. Fraction of D is needed

Mate, Vizkeleti, Szalay, Hagymasi, Ö.L.

(left) S^z_i for the ground state in the presence of a pinning magnetic field at the bottom zigzag

Topologically protected, correlated end spin formation in carbon nanotubes **Moca, Izumida, Dóra, Ö.L., Zaránd (PRL, 2019)**



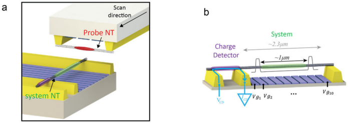


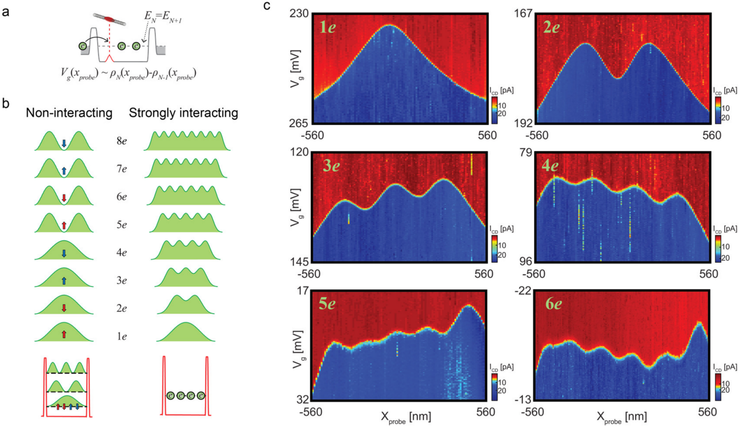
- $H_0 = -\sum_{\mathbf{x},\mathbf{x}',s} t(\mathbf{x}-\mathbf{x}') c_s^\dagger(\mathbf{x}) c_s(\mathbf{x}')$ with $\mathbf{r} = \mathbf{r}(\mathbf{x}) = \mathbf{r}(\nu, \mathrm{l}, \tau)$
- Construct and diagonalize H_0 and obtain the corresponding eigenfunctions; express the Coulomb interaction in this basis.
- $S_1 = S_2 = \frac{N_{\text{edge}}}{2}$
- Topological nanotubes spontaneously form double dot devices.
- Sign of the interaction can be changed by changing the dielectric constant of the environment.
- Coupling between ferromagnetic edge states is length and chirality

Particles in a harmonic trap: modes are Hermite polynomials

Shapir, Hamo, Pecker, Moca, ÖL, Zarand, Ilani (Science, 2019)

Scanning charge technique:

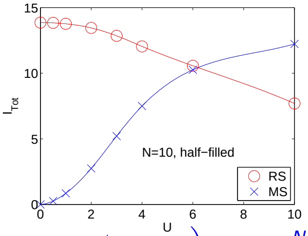




k-space: modes are momentum eigenstates (k-DMRG)

Xiang(1996), Nishimoto,Jeckelmann,Gebhard,Noack(2002),Ö.L.,Sólyom(2003), ...

Entropy behavior:	$U = 0$	$U = \infty$
momentum space:	$I_{\mathrm{Tot}} = \sum_p S_p = 0$	$I_{\mathrm{Tot}} = N \ln 4$
real space:	$I_{\mathrm{Tot}} = N \ln 4$	$I_{\mathrm{Tot}} = N \ln 2$



$$\mathcal{H} = -t \sum_{j=1,\sigma}^{N} \left(c_{j,\sigma}^{+} c_{j+1,\sigma} + c_{j+1,\sigma}^{+} c_{j,\sigma} \right) + U \sum_{j=1}^{N} n_{j,\uparrow} n_{j,\downarrow}$$

$\epsilon(k) = \sum_r e^{-ikr} t(r)$, where $k_i = (2\pi n)/N$, $-N/2 < n \leq N/2$

$T_{ij} = -2t \cos(k_i) \delta(i-j)$ and $V_{ijkl} = (U/N) \delta(i+j-k-l)$

$$\mathcal{H} = \sum_{k\sigma} \epsilon_k c_{k\sigma}^{\dagger} c_{k\sigma} + \frac{U}{N} \sum_{pkq} c_{p-q\uparrow}^{\dagger} c_{k+q\downarrow}^{\dagger} c_{k\downarrow} c_{p\uparrow},$$

momentum space exact at $U = 0$, crossover with U

Global fermionic mode optimization via swap gates

Krumnow,Veis,ÖL,Eisert(2015-2016), Friesecke,Werner,Kapas,Menczer,ÖL(2024)

- Finding an optimal representation of a quantum many body wave function, i.e., a parametrization with the minimum number of parameters for a given error margin is a task of utmost importance in modern quantum physics and chemistry

$$\Psi = \sum_{\mu_1,\ldots,\mu_N=0}^{1} C(\mu_1,\ldots,\mu_N)\Phi_{\mu_1,\ldots,\mu_N}$$

- Computatinal complexity $\sim$ block entropy area (BEA)

$$\mathcal{B}\alpha(C) = \sum_{\ell=1}^{N-1} S_\alpha(\rho_{1,2,\ldots,\ell})$$

$$\rho_{1,2,\ldots,\ell}(\mu_1,..,\mu_\ell;\mu'_1,\ldots,\mu'_\ell) = \sum_{\mu_{\ell+1},\ldots,\mu_N} C(\mu_1,\ldots,\mu_N)C^*(\mu'_1,\ldots,\mu'_\ell,\mu_{\ell+1},\ldots,\mu_N)$$

$S_\alpha(\rho) = \frac{1}{1-\alpha}\ln\left(\mathrm{Tr}\,\rho^\alpha\right)$ is the Rényi entropy for some $0 < \alpha < 1$.

- The important feature needed is concavity, so that density operators are favoured whose eigenvalues are either very large or very small.

Under a single particle unitary mode transformation $U \in U(N)$

- New modes $\varphi_i' = \sum_j U_{ij}\varphi_j$, and $C' = G(U)^\dagger C$ where $G(U)$ is a unitary transformation on the space of many-body coefficient tensors.

- For time reversal symmetric case, C and the φ_i are real-valued and $U \in O(N)$ or, discarding an immaterial overall sign factor, $U \in SO(N)$.

- U can be parametrized as $U = e^A U_*$ with U_* an arbitrary fixed matrix in $SO(N)$ and A real and skew-symmetric, the parametrization being unique for U close to U_*.

- Thus stationarity of a scalar function f on $SO(N)$ at U_* is equivalent to

$$0 = \frac{d}{dt}\Big|_{t=0} f(e^{tA}U_*) = \mathrm{Tr}\,\frac{\partial f}{\partial U}(U_*)U_*^T A^T\,,\ \forall A^T = -A$$

that is to say $\frac{\partial f}{\partial U}(U_*)U_*^T$ symmetric.

- Reduction to pairwise rotations: To achieve stationarity minimize f over all pairwise rotations $U_{ij}(\theta)$ given by $e^{\theta E_{ij}}$, $E_{ij} = e_i e_j^T - e_j e_i^T$, where e_i is the unit vector of $\mathbb{R}^N$ whose i-th component is 1 and whose other components are zero. This corresponds to the mode transformation $\varphi_i' = \cos\theta\varphi_i + \sin\theta\varphi_j$, $\varphi_j' = -\sin\theta\varphi_i + \cos\theta\varphi_j$ which leaves all other modes the same.

Reduction to permutations and nearest neighbor rotations

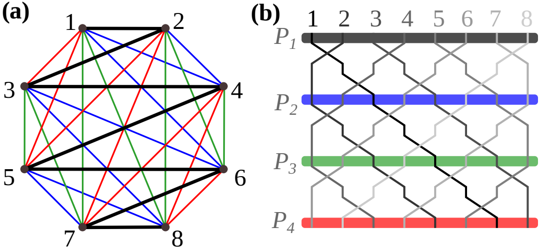
- Set of all pairwise rotations $U_{ij}(\theta)$ can be realized by $N/2$ global re-orderings of the orbitals and the $N-1$ nearest-neighbor rotations for each ordering.

- if $\tau_1, ..., \tau_{N/2}$ are the specific permutations such that any pair of orbitals become nearest neighbours under one of these permutations (that is, for all $i < j$ there exist ν and ℓ such that $\{\tau_\nu(i), \tau_\nu(j)\} = \{\ell, \ell+1\}$), then

$$U_{ij}(\theta) = \tau_\nu^{-1}\, U_{\ell,\ell+1}(\pm\theta)\, \tau_\nu$$

with '+' if $\tau_\nu(i) < \tau_\nu(j)$ and '−' otherwise.

Swap gates controlled permutations: The optimal set of permutations $\tau_1, ..., \tau_{N/2}$ can be generated by Walecki's method (1882):



- The modes are placed in a zig-zag line to the vertices of a regular polygon with N vertices

Local mode optimization and block entropy area

Krumnow,Veis,ÖL,Eisert(2015-2016)

- Consecutive permutations are easily generated by two layers of nearest neighbor swap operations placed in a checkerboard pattern ,i.e., full forward mode optimization sweep with fixed, but alternating angles of $\pi/2$ and 0 and a backward sweep in a reversed order.

- To avoid truncation of the wavefunction, the bond dimension has to be increased by a factor of q.

Local mode optimization and block entropy area: $H(U) = G(U)^\dagger H G(U)$ is constructed iteratively from two-mode unitary operators by optimizing $\theta_{l,l+1}$ while sweeping through the network for $l = 1 \ldots N-1$

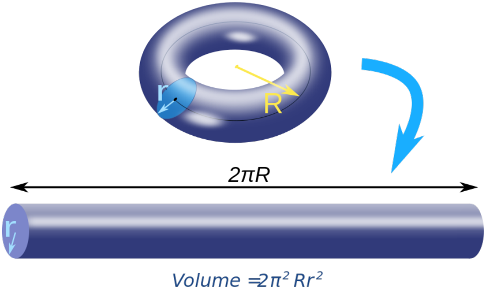
$$U^{\text{loc}}_{\text{opt}} = \text{argmin}_{U \in V} f_l\big(|\psi(\mathbb{I}_l \oplus U \oplus \mathbb{I}_{N-l-2})\rangle\big),$$

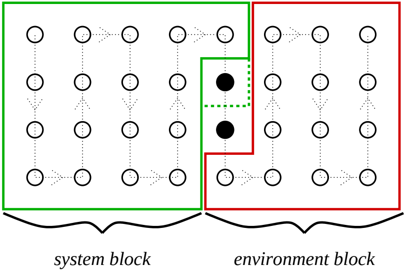
- At each micro-iteration step, f_l, i.e., the half-Rényi block entropy $S_{1/2}(\rho_{\{1,2,\ldots,l\}})$ is minimized by a two-mode rotation (disentanglers)

Locality of the BEA: Nearest neighbor rotations by $\theta_{l,l+1}$ change only the block entropy measured when the cut is at mode l, while all other block entropies remain invariant.

2d spinless fermions with PBCs Krumnow, Veis, Eisert, Ö.L (2019-2021)

$$H = \sum_{\langle i,j \rangle} c_i^\dagger c_j + \sum_{\langle i,j \rangle} V n_i n_j, \qquad U = \begin{pmatrix} exp(i\theta_1)cos(\theta_2) & exp(i\theta_1)sin(\theta_2) \\ -exp(-i\theta_1)sin(\theta_2) & exp(-i\theta_1)cos(\theta_2) \end{pmatrix}$$

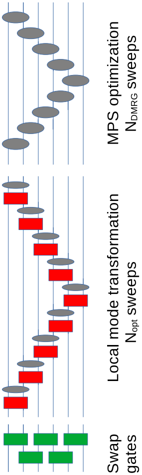


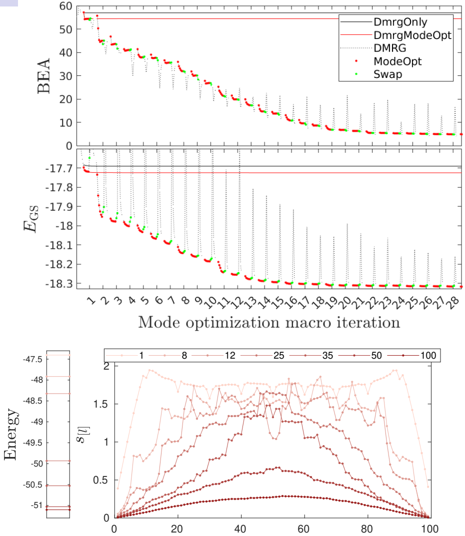


- Optimization on the mps manifold and on the Grassman manifold
- Hamiltionian becomes long ranged

$$H = \sum_{i,j=1}^{n} t_{i,j} c_i^\dagger c_j + \sum_{i,j,k,l=1}^{n} v_{i,j,k,l} c_i^\dagger c_j^\dagger c_k c_l,$$

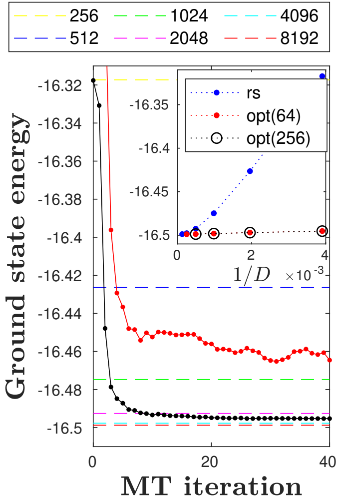
2d-spinless Hubbard, $N = 6 \times 6$ and 10×10 with $D = 80$

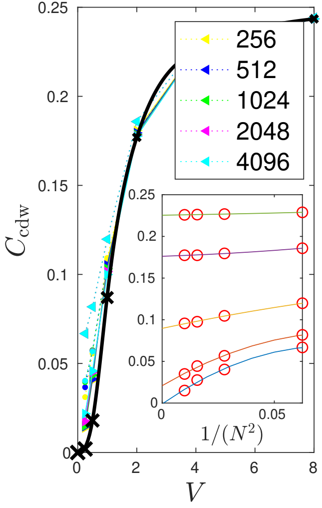




Consistency: Perturbation induced by the swap layers can be eliminated by subsequent application of several DMRG sweeps via nearest neighbor mode optimization, once the algorithm has found the stationary solution for both energy and BEA.

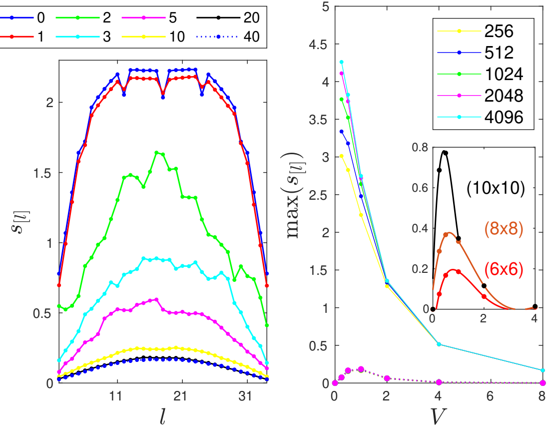
2d spinless fermions on a torus Krumnow, Veis, Eisert, Ö.L (2019-2021)





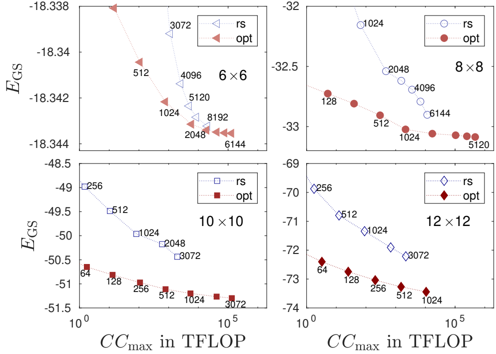
- $C_{\rm cdw} = (1/N^4) \sum_{i,j} \eta_{i,j} (n_i - 1/2)(n_j - 1/2)$
- One- and two-particle reduced density matrices. $\rho^{(2)}_{i,j,k,l} = \langle c_i^\dagger c_j^\dagger c_k c_l \rangle$
- Torus geometry: very fast finize size scaling

Residual entropy and KT-QPT



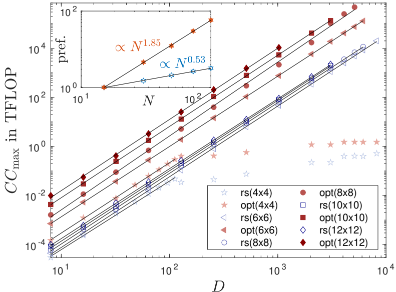
- Maximum of block entropy is reduced by a factor 10 (entropy is a logarithmic function)
- Highly simmetric and smooth entropy profiles is obtained
- Scaling of residual entropy, area law etc
- physics determined real correlations, quasi-particle picture etc
- Different mode picture for higher dimensional TNS methods

Direct connection to computer science: complexity in FLOPS



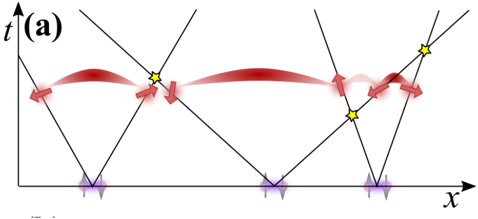
Menczer,Kapas,Werner,ÖL,2023

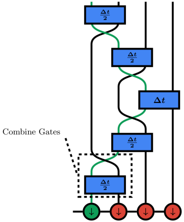
- Half-filled $N \times N$ Spinless model on a torus geometry
- $t = 1,\ t' = 0.4,\ V = 0.8$
- opt with $D = 80$
- Computational complexity in teraFLOPS



- The solid lines are first-order polynomial fits leading to exponents $\nu \simeq 3 \pm 0.2$
- inset: scaling of the prefactor as a function of system size N with fitted exponents 0.53 and 1.85 for the real space and for the optimized basis, respectively.

Long time evolution Krumnow, Eisert, Ö.L. (2019)

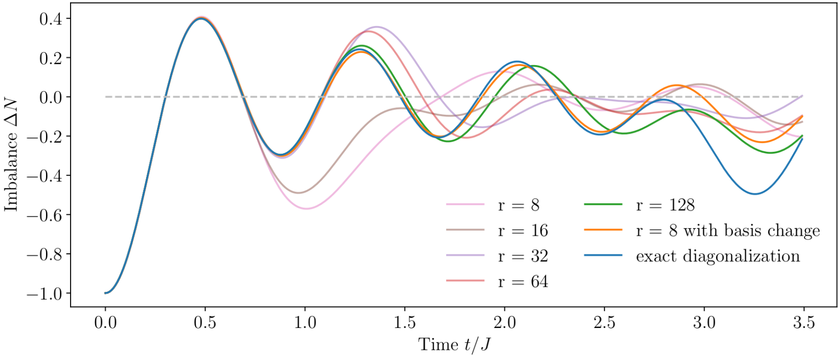




- At time $t = 0$ we perturb the system.
- After the quench the quasiparticles collide with each other. , the scattering events are denoted by stars.
- There are different time-evolution methods for MPS which are currently in use to solve the time-dependent Schrödinger equation (TDSE).
- application of $\hat{U}(\delta_t) = e^{-i\delta_t \hat{H}}$, i.e. , $|\psi(t)\rangle \rightarrow |\psi(t+\delta_t)\rangle$
- time-evolving block decimation (TEBD), $MPOW^{\mathrm{I,II}}$, Krylov, time-dependent variational principle (TDVP)
- each has advatages and disadvateges.
- TDVP $\rightarrow$ general non-local Hamiltonians (quantum chemsitry)

Long time evolution Krumnow, Eisert, Ö.L. (2019)

- correlations in the system spread and bond dimension increases
- Combination of time evolution and adaptive mode transformation:
- optimization of modes after a full dmrg sweep, i.e. dt time step, or after several dt time steps
- Ex: evolution of the imbalance $\Delta N(t) = (N_{\mathrm{even}}(t) - N_{\mathrm{odd}}(t))/N$. in 1D spinless Hubbard model, $|\psi(0)\rangle = |101010\ldots\rangle$



Introduction to BUG

General non-linear PDE

$$\dot{Y} = \Phi(Y) \qquad Y = Y_{ij} \in \mathbb{C}_{D_1 \times D_2}$$

Low-rank approximation

$$Y = USV^\dagger$$

U: $D_1 \times M$, S: $M \times M$, $V^\dagger$: $M \times D_2$

- U and V are (partial) isometries, with orthonormalized columns
- Similar to SVD, but S is not necessarily diagonal

Naive low-rank equations of motion

$$\dot{U} = \Phi(USV^\dagger)VS^{-1}$$
$$\dot{S} = U^\dagger \Phi(USV^\dagger)V$$
$$\dot{V}^\dagger = S^{-1}U^\dagger \Phi(USV^\dagger)$$

⟺ Naive discretization is hard due to possibly large curvature (S^{-1})

New variables: **No size increase!!**

$$X = US, \; Z = SV^\dagger$$

$$\dot{X} = \Phi(XV^\dagger)V$$
$$\dot{Z} = U^\dagger \Phi(UZ)$$

No S^{-1}

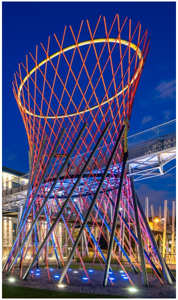
We can go through high curvature points without even noticing them!

Discretized scheme

$$\tilde{X} = X + \delta t \Phi(XV^\dagger)V$$
$$\tilde{Z} = Z + \delta t U^\dagger \Phi(UZ)$$
$$\hat{S} = S + \delta t U^\dagger \Phi(USV^\dagger)V$$

$\{\tilde{U}, \tilde{V}\}$ From QR decompositions of $\{\tilde{X}, \tilde{Z}\}$

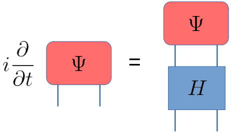
$\tilde{S} = \tilde{U}^\dagger U \hat{S} V^\dagger \tilde{V}$ Basis transf.



[2] Gianluca Ceruti, Christian Lubich, and Dominik Sulz, SIAM Journal on Numerical Analysis 61, 194-222 (2023).

BUG for the Schrödinger equation – rank adaptive BUG

$$\mathcal{H} = \mathcal{H}_{\text{left}} \otimes \mathcal{H}_{\text{right}} \qquad |\Psi\rangle = \sum_{i,j} \Psi_{ij} |i\rangle |j\rangle \qquad i\dot{\Psi}_{ij} = \sum_{k,l} \underbrace{\langle i|\langle j|H|k\rangle|l\rangle}_{H_{ij,kl}} \Psi_{kl}$$



Low rank approximation

$$\Psi_{ij} = \sum_{\alpha,\beta}^{M} U_{i\alpha} S_{\alpha\beta} V^*_{j\beta} \qquad S_{\alpha\beta} \in \mathcal{H}^{\text{trunc}}_{\text{left}} \otimes \mathcal{H}^{\text{trunc}}_{\text{right}}$$

U and V* define truncated subspaces in $\mathcal{H}_{\text{left}}$ and $\mathcal{H}_{\text{right}}$

$$X_{i\beta} = \sum_{\alpha} U_{i\alpha} S_{\alpha\beta} = \sum_{j} \Psi_{ij} V_{j\beta} \qquad X_{i\beta} \in \mathcal{H}_{\text{left}} \otimes \mathcal{H}^{\text{trunc}}_{\text{right}}$$

$$Z_{\alpha j} = \sum_{\beta} S_{\alpha\beta} V^*_{j\beta} = \sum_{i} U^*_{i\alpha} \Psi_{ij} \qquad Z_{\alpha j} \in \mathcal{H}^{\text{trunc}}_{\text{left}} \otimes \mathcal{H}_{\text{right}}$$

$$\tilde{X} = X + \delta t \Phi(XV^\dagger)V = X - i\delta t H^{\text{eff}}_{\text{left}} X \iff \mathcal{H}_{\text{left}} \otimes \mathcal{H}^{\text{trunc}}_{\text{right}}$$

$$\tilde{Z} = Z + \delta t U^\dagger \Phi(UZ) = Z - i\delta t H^{\text{eff}}_{\text{right}} Z \iff \mathcal{H}^{\text{trunc}}_{\text{left}} \otimes \mathcal{H}_{\text{right}}$$

$$\hat{S} = S + \delta t U^\dagger \Phi(USV^\dagger)V = S - i\delta t H^{\text{eff}}_{\text{center}} S \iff \mathcal{H}^{\text{trunc}}_{\text{left}} \otimes \mathcal{H}^{\text{trunc}}_{\text{right}}$$

Evolution with projected Hamiltonians

Rank adaptive BUG

- "New" truncated basis from $\tilde{X} \& \tilde{Z}$ via QR decomposition ⇒ $\{\tilde{H}^{\text{trunc}}_{\text{left}}, \tilde{H}^{\text{trunc}}_{\text{right}}\}$
- S-equation is not performed the "old" basis, but in $\hat{\mathcal{H}}^{\text{trunc}}_{\text{left}} \otimes \hat{\mathcal{H}}^{\text{trunc}}_{\text{right}}$

$$\hat{\mathcal{H}}^{\text{trunc}}_{\text{left}} = \mathcal{H}^{\text{trunc}}_{\text{left}} \oplus \tilde{\mathcal{H}}^{\text{trunc}}_{\text{left}}$$

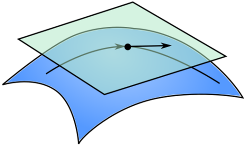
$$\hat{\mathcal{H}}^{\text{trunc}}_{\text{right}} = \mathcal{H}^{\text{trunc}}_{\text{right}} \oplus \tilde{\mathcal{H}}^{\text{trunc}}_{\text{right}}$$

MPS Simulation of the dynamics

$$i\partial_t|\Psi(t)\rangle = H|\Psi(t)\rangle$$ where $|\Psi(t)\rangle$ is approximated by an MPS.

Tangent space methods

MPS manifold



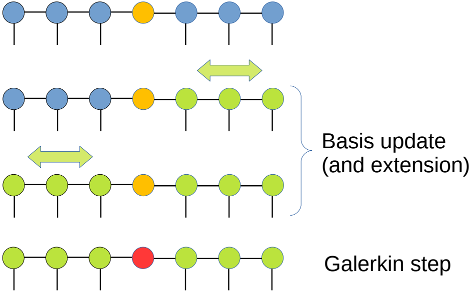
Low-rank approximation (based on Schmidt-decomposition)

$$|\Psi\rangle \approx \sum_{\alpha=1}^{M} \lambda_\alpha |\alpha\rangle_L \otimes |\alpha\rangle_R$$

Time evolution of Schmidt states and weights?

Basis update and Galerkin (BUG)

- Simple method
- 1-site scheme (cheap for heavy sites)
- Only forward steps
- General scheme
  (MPS, TTNS, nonlinear PDE’s)
- Straightforward parallelization



[2] Gianluca Ceruti, Christian Lubich, and Dominik Sulz,
SIAM Journal on Numerical Analysis 61, 194-222 (2023).

Quantum simulation on Q-chip

J. Cobos, J. Fraxanet, C. Benito, F. di Marcantonio, P. Rivero, K. Kapás, M. A. Werner, Ö. L., A. Bermudez, E. Rico

General Q-algorithms usually require arbitrary 2-qubit gates ⇔ Q-chip lattice grid (swap gates are necessary)

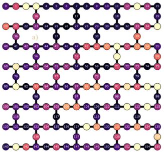
⇕

Noise limits circuit depth

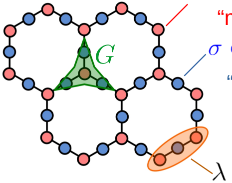
Idea: simulate a 2D lattice model compatible with the chip at hand!

(2+1)D Z_2 Higgs model on the honeycomb lattice

$$H = -m \sum_{\boldsymbol{n}} \tau^z_{\boldsymbol{n}} - g \sum_{(\boldsymbol{n},\boldsymbol{v})} \sigma^z_{(\boldsymbol{n},\boldsymbol{v})} - \lambda \sum_{(\boldsymbol{n},\boldsymbol{v})} \tau^x_{\boldsymbol{n}+\boldsymbol{v}} \sigma^x_{(\boldsymbol{n},\boldsymbol{v})} \tau^x_{\boldsymbol{n}}$$



⇔



$\tau \in \{\uparrow,\downarrow\} \Leftrightarrow N \in \{0,1\}$
"matter"

$\sigma \in \{\uparrow,\downarrow\} \Leftrightarrow E \in \{0,1\}$
"field"

Gauge invariance

$$G_n = \tau^z_n \sigma^z_{(n,1)} \sigma^z_{(n,2)} \sigma^z_{(n,3)}$$

$$[G_n, H] \equiv 0 \ \forall n$$

"physical" subspace:

$$G_n |\Psi\rangle = +1 |\Psi\rangle \ \forall n$$

Z2 Gauss law

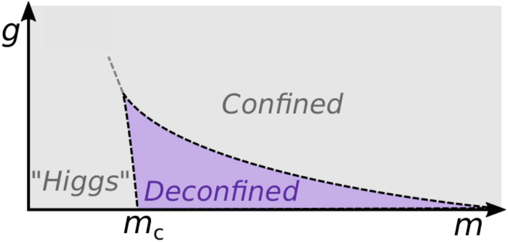
$$N_n = \sum_v E_{(n,v)} \ \mathrm{mod} 2$$

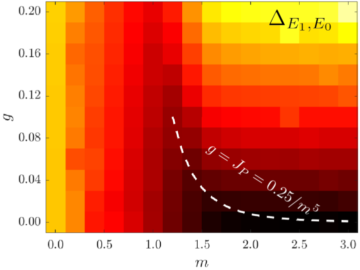
Absent plaquette term:

$$H_\mathrm{P} = J_\mathrm{P} \sum_\mathrm{P} \prod_{(n,v)\in \mathrm{P}} \sigma^x_{(n,v)}$$

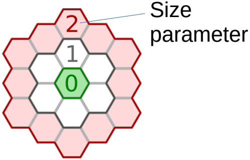
- Would generate direct field line fluctuations
- Is still generated perturbatively $J^\mathrm{eff}_\mathrm{P} \approx 0.25 \frac{\lambda^6}{m^5}$

Phase diagram of the Z2HM – DMRG results



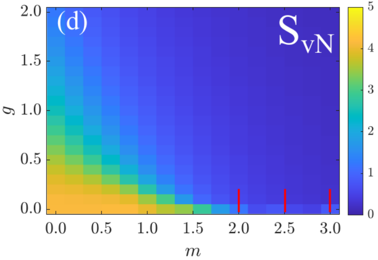


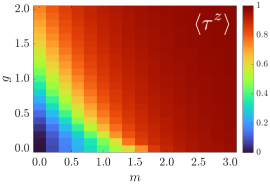
Finite-sized flakes

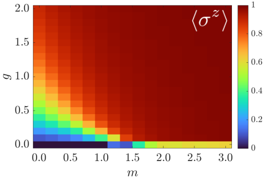


$$H = -m \sum_{\boldsymbol{n}} \tau^z_{\boldsymbol{n}} - g \sum_{(\boldsymbol{n},\boldsymbol{v})} \sigma^z_{(\boldsymbol{n},\boldsymbol{v})} - \lambda \sum_{(\boldsymbol{n},\boldsymbol{v})} \tau^x_{\boldsymbol{n}+\boldsymbol{v}} \sigma^x_{(\boldsymbol{n},\boldsymbol{v})} \tau^x_{\boldsymbol{n}} \quad \text{with } \lambda = 1$$

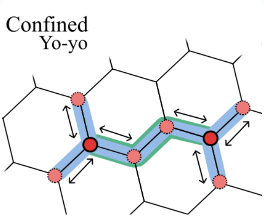
Half-system entanglement

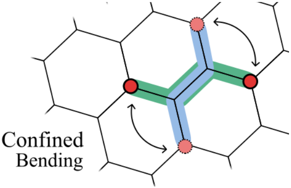


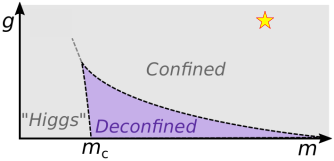


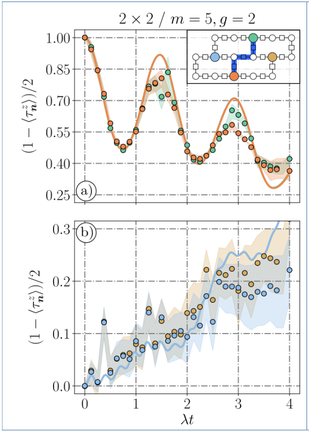


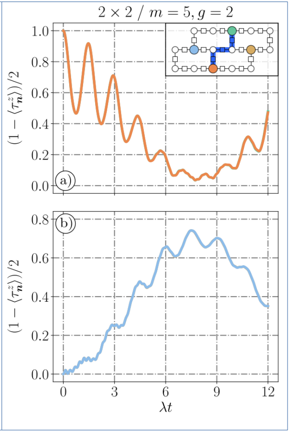
1-string initial state: Confined regime











- MPS simulation with M=128-256
- MPS can reach longer times
- Q-chip cannot resolve small vacuum fluctuations

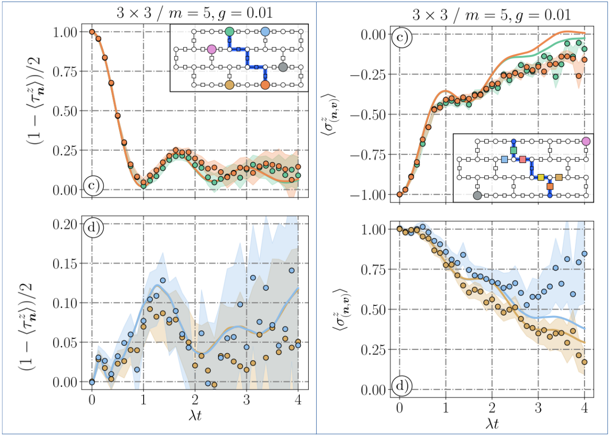
Periods from perturbation theory

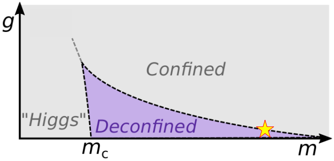
$$\lambda T_y = \frac{2\pi\lambda}{2g} \approx 1.57$$

$$J_b = -\frac{\lambda^2}{2g} + \frac{\lambda^2}{2g+4m} \approx -0.208\lambda$$

$$\lambda T_b = \frac{2\pi\lambda}{2|J_b|} = 15.1$$

1-string initial state: Deconfined regime

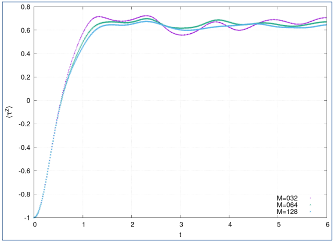




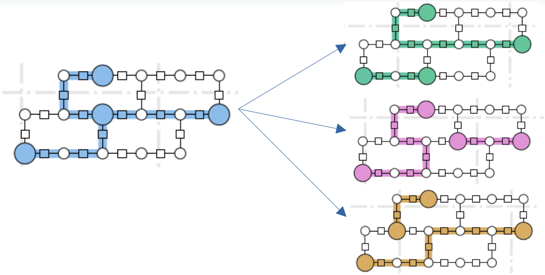
- Rapid relaxation to

$$\frac{1-\langle\tau^z_n\rangle}{2} \approx \frac{2}{\mathrm{vol}} \ (+\text{vac. fluct.})$$

High MPS dimension (M~512-1024) is necessary



String breaking



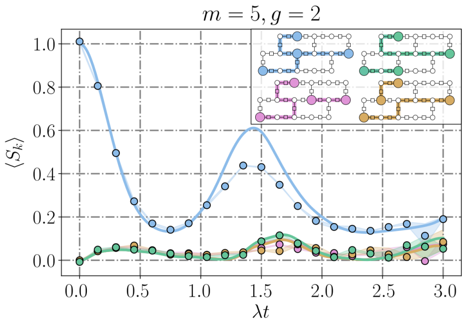
Probabilities of configurations?

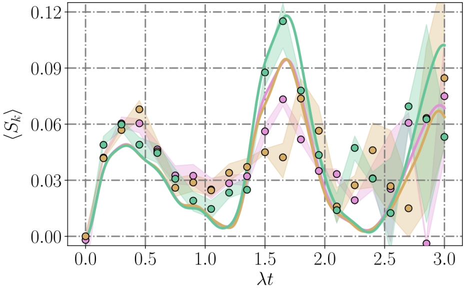
Impossible to measure (noise)

Correlators of 4 matter sites?

$$S_k = \prod_{n\in\mathcal{S}_k}\left(\frac{1-Z_n}{2}\right)$$

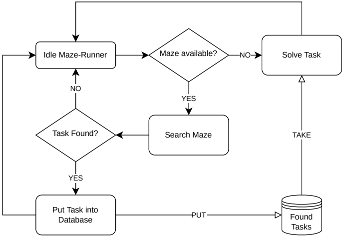
Matter sites of the configuration



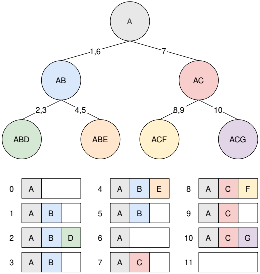


Novel algorithmic solutions & parallelization A.Menczer,ÖL(2023)

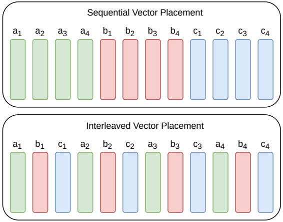
Life Cycle of a Maze-Runner Thread.



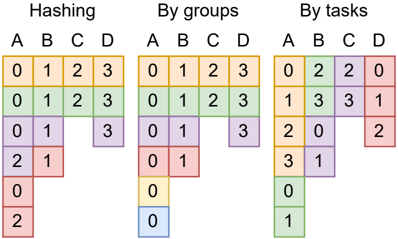
Graph theory based memory management



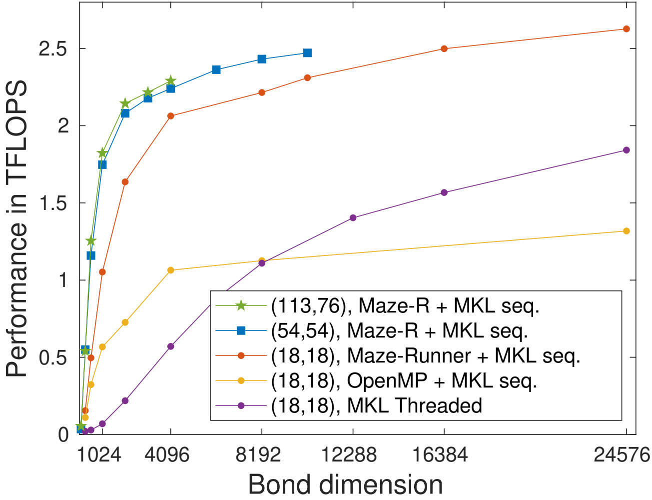
Strided Batched operations via data localization



Execution via hierarchy of tasks

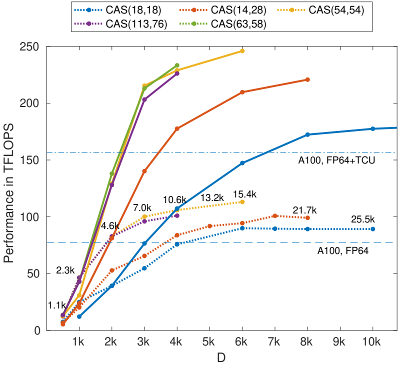


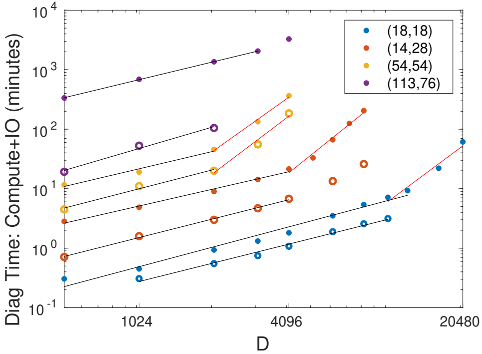
CPU only limit (for CAS(113,76) $\dim \mathcal{H} = 2.88 \times 10^{36}$)



Performance measured in TFLOPS for the F_2 and FeMoco chemical systems for CAS(18,18) and CAS(54,54) orbitals spaces, respectively, as a function of the DMRG bond dimension on a dual Intel(R) Xeon(R) Gold 5318Y CPU system with 2×24 physical cores running at 2.10 Ghz.

Quarter petaflops on a single node $\sim$ 10000x speedup; $D^3 \rightarrow D$





- NVIDIA DGX H100: 80x speedup wrt a single node with 128 cores

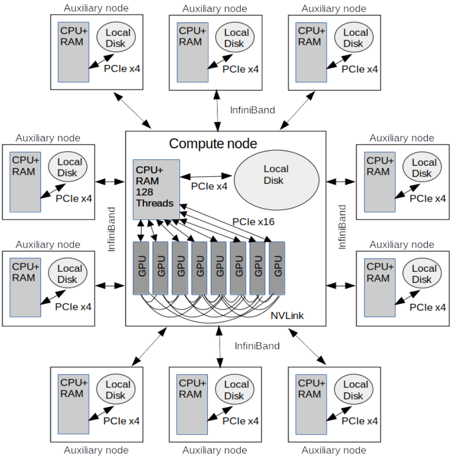
Testing performance up to $\sim$ 250 TFLOPS in collab with NVIDIA and SandboxAQ Menczer,Damme,Rask,Huntington,Hammond,Xantheas,Ganahl,ÖL

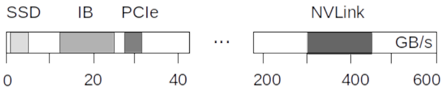
- New model to utilize NVIDIA D2D links. A. Menczer ÖL (unpublished 2023)

- Combination of our MPI and GPU kernels: full replacement of *boost library*, asynchronous IO, multiNode-multiGPU

$\rightarrow$ petascale computing. A. Menczer ÖL (unpublished 2023-2024)

Cost optimized ”green” TNS Menczer, ÖL 2024





- DGX-H100 costs 100 USD/hour on Google Cloud

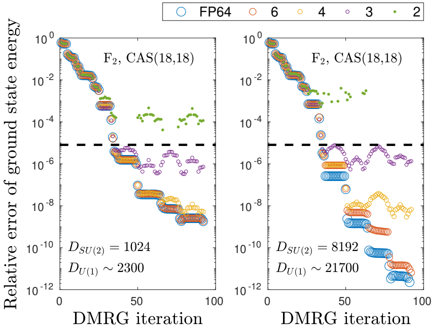
- Schematic plot of hardware topology illustrating the various communication channels (arrows), such as host to host (H2H), host to device (H2D) and device to host (D2H), and device to device (D2D), i.e., InfiniBand, PCI-E, and NVLink, accordingly.

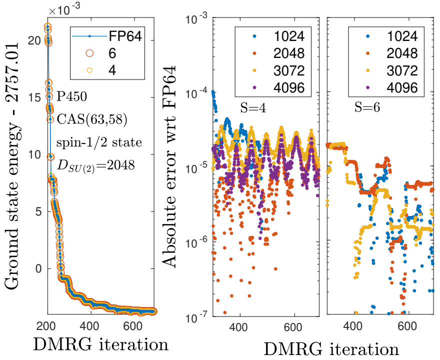
- The compute node is a very powerful and expensive unit surrounded by one or more cheap auxiliary nodes with minimal computational capacity, but with substantial amount of RAM

Mixed precision ab initio TNS methods adapted for NVIDIA Blackwell technology via emulated FP64 arithmetic

J. Gunnels, C. Brower, S. R. Bernabeu, J. Hammond, S. Xantheas, M. Ganahl, A. Menczer, Ö.L.

- Results obtained on DGX-B200 single node utilizing the Ozaki scheme
- Results obtained via early access utilizing a pre-release cuBLAS binary, and the data is subject to change.
- mantissa bit setting $\{15, 23, 31, 39, 47, 55\}$ for $S = 2, 3, 4, 5, 6, 7$ slices.

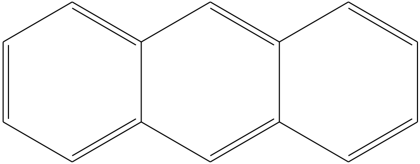




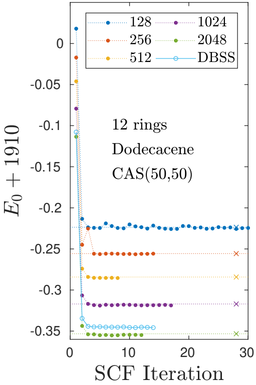
- Chemical accuracy, 1.6mHa, can be reached with 4, 6 slices

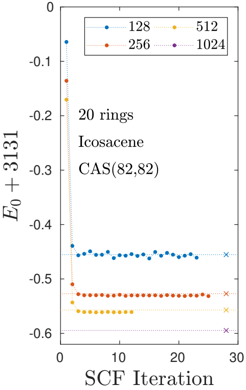
Self Consistent Field DMRG on NVIDIA DGX-H100/B200

ÖL, Menczer,Ganyecz,Kapas,Werner,Hammond,Xantheas,Ganahl,Neese (JCTC 2025)



Polycyclic aromatic hydrocarbons

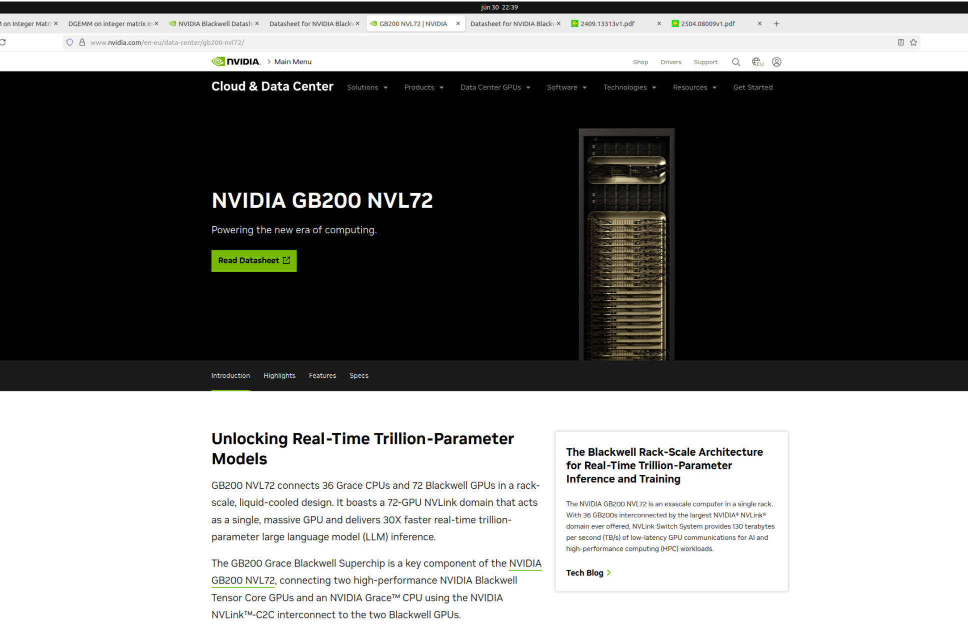




- For Heptacene, CAS(30,30), 25 DMRGSCF iterations with $D = 256$ using other codes took $\sim$ 7 days
  $\rightarrow \sim$ 3.8 hours with our hybrid DMRG + ORCA
- Dodecacene CAS(50,50)
- Full orbital space:
  328 electrons on 840 orbitals.
- Wall time: 13.3 hours ($D = 512$).

- Icosacene CAS(82,82)
- Full orbital space:
  536 electrons on 1368 orbitals.
- Wall time: $\sim$ 1.2 days ($D = 512$).
- Use DMRG-TCC (DLPNO).

New TNS benchmarks for quantum computing ???



Conclusion and near future on GH200, MI300, B200 etc

- Tensor topologies together with proper basis representations are important for efficient data sparse representaion of the wavefunction
- Global and local mode optimization for tree-like TNS provides black-box tool to reduce computatinal complexity
- Long time evolution with adaptive mode transformation is a promising direction in collab. Eisert, Lübich
- Combination of TNS with other (conventional) methods can exploit benefits of the individual methods, DMRG-RAS-X, DMRG-SCF, ...
- Massive Parallelization multiNode-multiGPU → exascale computation
- Mixed precision TNS on specialized new hardware with lower energy consumtion
- → Simulation of realistic material properties in collab. Riverlane, Furukawa

Supports: Hungarian Academy of Sciences, the Hungarian National Research, Development and Innovation Office TKP2021-NVA-04, Quantum Information National Laboratory of Hungary, Alexander von Humboldt Foundation (Germany), Hans Fischer Senior Fellowship programme (IAS-TUM, Germany), SPEC, DOE, (PNNL, USA)