Large Scale DMRG Calculations

ETH zürich
UNIVERSITÉ DE GENÈVE

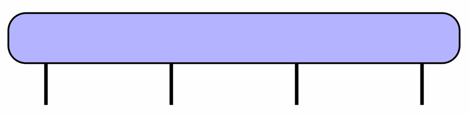
M. Dolfi[1], A. Kantian[2], A. Kosenkov[1], B. Bauer[3], T. Giamarchi[4], M. Troyer[1]

1 Theoretische Physik, ETH Zurich, Switzerland
2 Nordita, KTH Royal Institute of Technology and Stockholm University, Sweden
3 Station Q, Microsoft Research, Santa Barbara, USA
4 DPMC-MaNEP, University of Geneva, Switzerland

NORDITA The Nordic Institute for Theoretical Physics
Microsoft Research

DMRG in a nutshell

full wavefunction

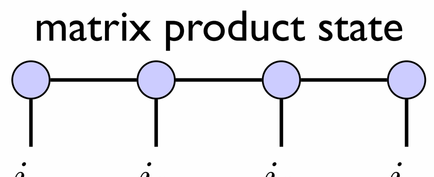


$$|\Psi\rangle = \sum c(i_1, i_2, i_3, i_4)\, |i_1 i_2 i_3 i_4\rangle$$

- exponential number of parameters

$$|\mathcal{H}| = d^L$$

matrix product state



$$|\Psi\rangle = \sum A_1(i_1) \cdots A_4(i_4)\, |i_1 i_2 i_3 i_4\rangle$$

- product over (MxM) matrices

$$|\mathcal{H}| = LM^2 d$$

- stores only a finite entanglement

$$S \propto \log M$$

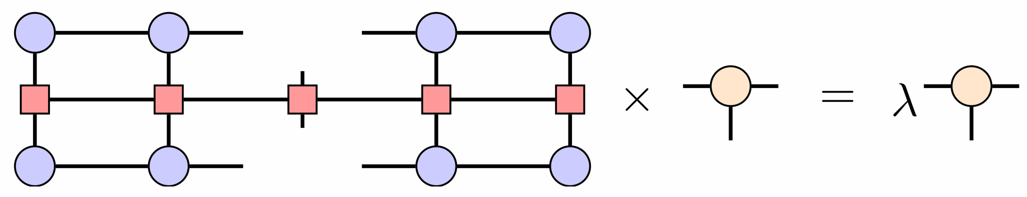
variational principle
for every value of the parameters in the matrices: $E \geq E_0$

variational iterative optimization

1. solve for the best tensor A_i keeping all others fixed

$$\frac{\partial}{\partial A_i^*}\left(\langle\psi|\hat{H}|\psi\rangle - \lambda\left[\langle\psi|\psi\rangle - 1\right]\right) = 0$$

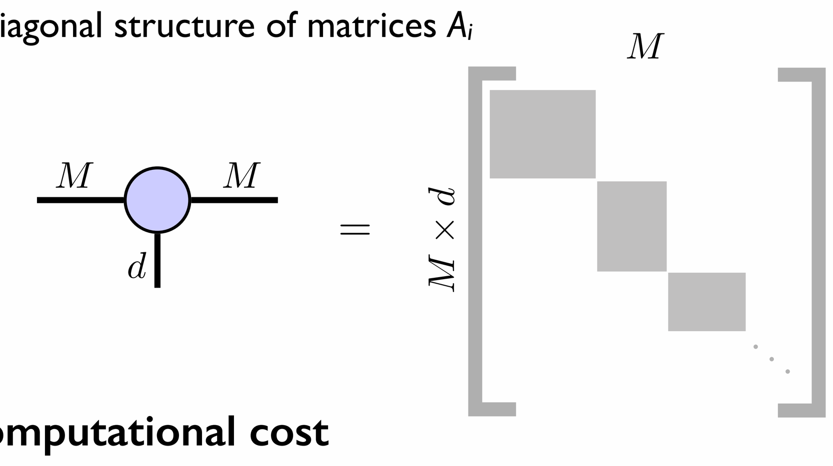
- this is an eigenproblem: solve for the lowest eigenvector



2. move to the next site $i+1$
3. repeat sweeping back and forth until convergence

symmetries

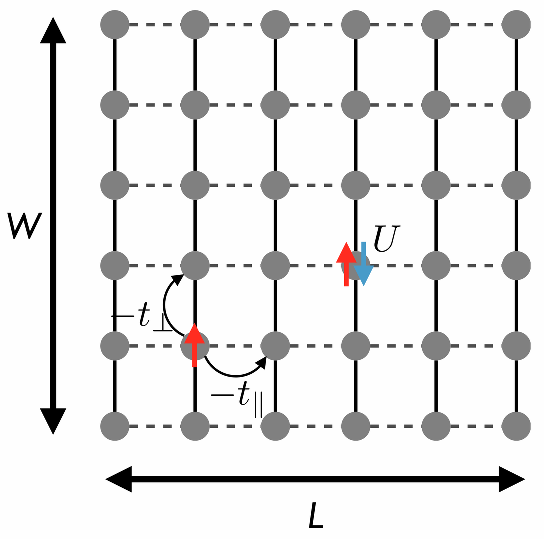
- many models conserve multiple quantum numbers
  - particle number, spin, etc.
- block-diagonal structure of matrices A_i



main computational cost

- matrix-matrix multiplications $O(M^3 d^2)$
- singular value decomposition $O(M^3 d^3)$

Hubbard model with DMRG



$-t_\parallel$ hopping along a chain

$-t_\perp$ hopping between chains

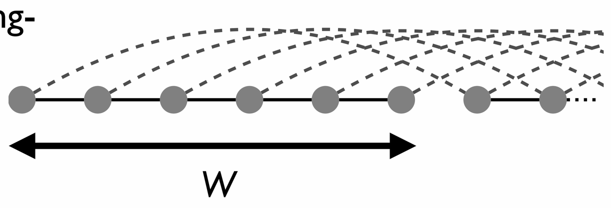
U interaction

$$\hat{H} = -t_\parallel \sum_{i,j,\sigma}\left[\hat{c}^\dagger_{(i,j),\sigma}\hat{c}_{(i+1,j),\sigma} + \text{H.c.}\right] - t_\perp \sum_{i,j,\sigma}\left[\hat{c}^\dagger_{(i,j),\sigma}\hat{c}_{(i,j+1),\sigma} + \text{H.c.}\right] + U\sum_{i,j}\hat{n}_{(i,j),\uparrow}\hat{n}_{(i,j),\downarrow}$$

the model

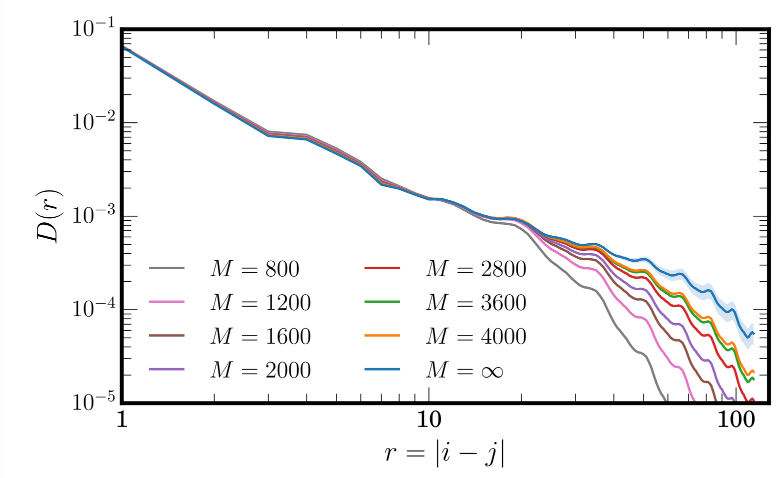
- toy model for High-T_c superconductors
  - Copper / Iron oxydes
- implemented experimentally with ultra-cold gases
- very challenging low-temperature phase diagram
  - proposal of many phases
  - disagreement between numeric methods

- lattice mapped to a chain with long-range interactions
- expected a ground-state entanglement scaling as



$$S \propto W$$

- need of exponentially large MPS bond dimension M



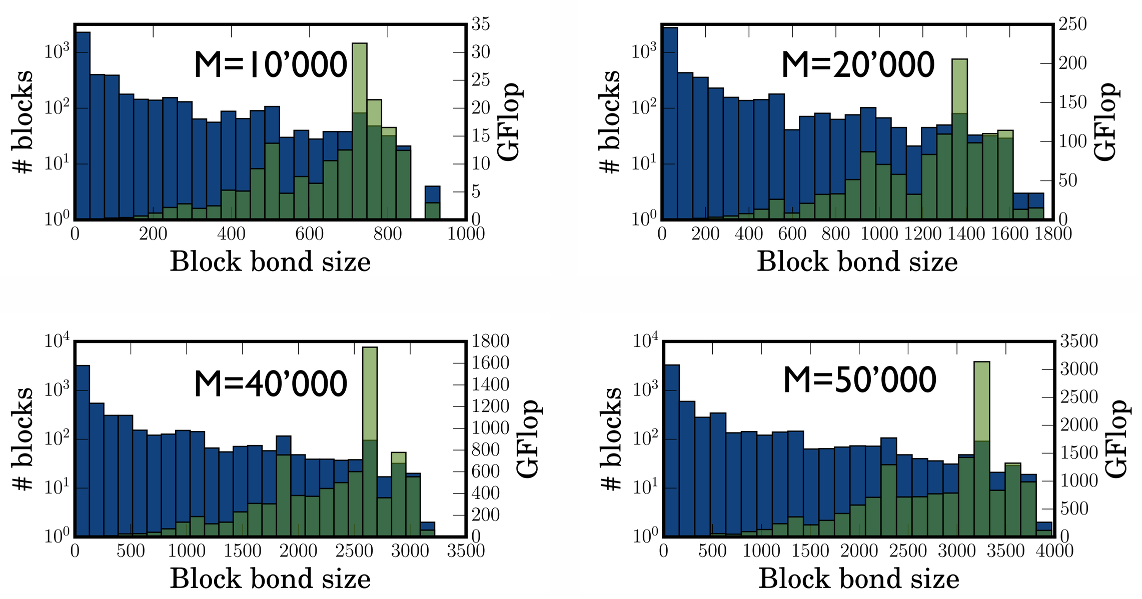
pair correlation on 2x128 ladder

- in quasi-1D models the slowest decaying correlation function characterizes the phase of the system
- dominant superconductivity for $D(r) = r^\mu$ with $\mu < 1$
  - underestimated in previous studies with limited M

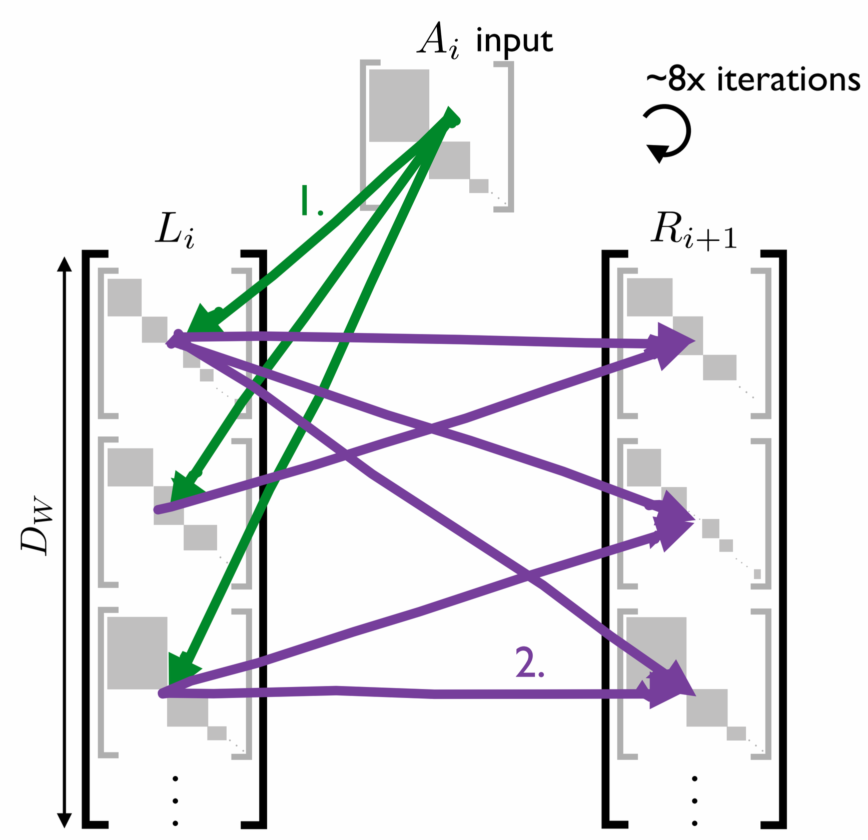
Parallel 2D DMRG

resources scaling $\propto LWM^2d$ $\propto LWD_WM^2$

L	W	D_W	M	d	MPS Size [GB]	Number of blocks	Sparsity	Scratch space [TB]
8	10	42	10'000	4	8	64	2.00%	2.8
8	10	42	20'000	4	31	76	2.00%	10.2
8	10	42	40'000	4	113	84	1.80%	37.1
8	10	42	50'000	4	172	88	1.80%	56.4



- very sparse block matrices
- quickly needing more resources than what is available on a single node
- many small matrices, also when M is very large
- main computation time spent in medium-large matrices



multi-level memory parallelism

- reflects algorithm structure

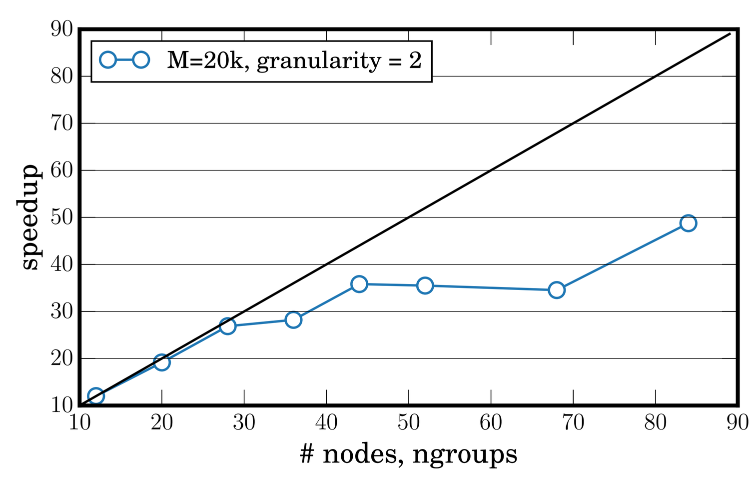
A. **ngroups :** distribute block-matrices of L_i and R_{i+1} in ngroups each

B. **granularity :** size of a group distribute dense matrices within the group using a round-robin distribution based on the matrix size

computational parallelism

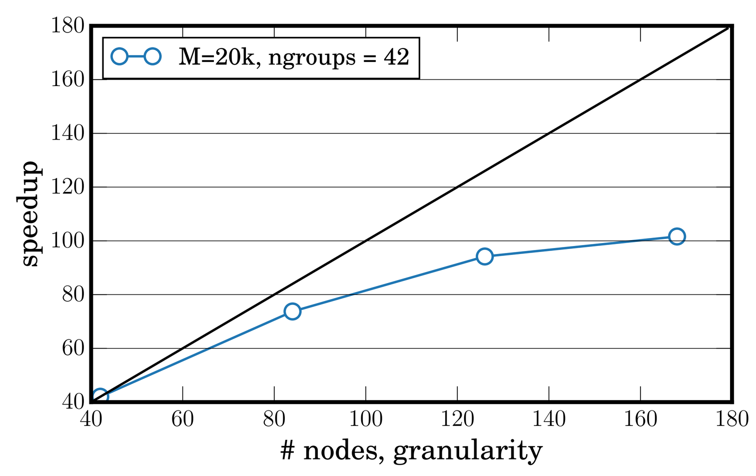
- group operations

1. broadcast A_i to all groups and compute many parallel gemm
2. complex reduction:
   a. send L_i blocks to group of R_{i+1}
   b. multiple parallel gemm
   c. sum all blocks among groups



scaling number of groups : 6 - 42

- number of groups limited by D_W
- the complex reduction causes communication overhead and imbalance workload



scaling granularity : 1 - 6

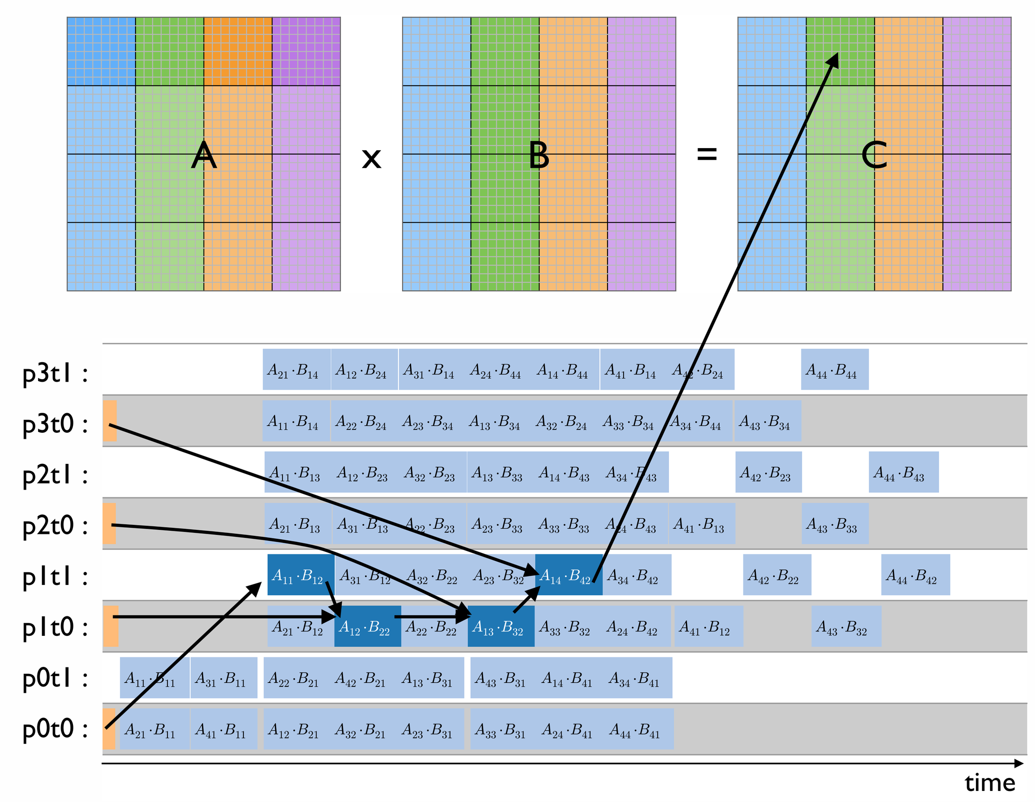
- higher granularity introduces communication overhead, but allows for more nodes (needed because of memory constrains)

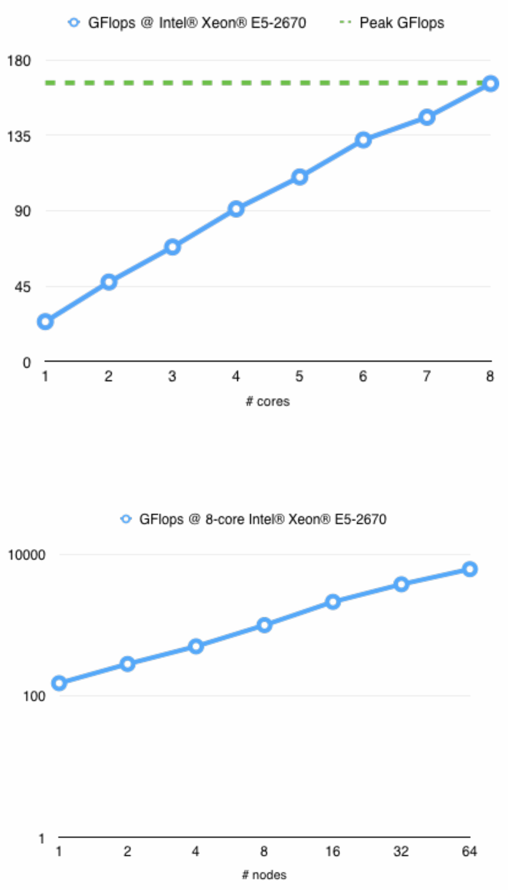
Parallelization framework – Ambient

```
ambient::scope::const_iterator where(int i, int j){
  return (ambient::scope::begin() + j % ambient::scope::size());
}

int main(){
   ambient::numeric::tiles< ambient::numeric::matrix<double> >
     a(N,N), b(N,N), c(N,N);

   for(int i = 0; i < a.mt; i++)
   for(int j = 0; j < a.nt; j++){
     ambient::actor proc(where(i,j));
     ambient::numeric::fill_random(a.tile(i,j));
     ambient::numeric::fill_random(b.tile(i,j));
   }
   for(int k = 0; k < a.nt; k++)
   for(int j = 0; j < c.nt; j++)
   for(int i = 0; i < c.mt; i++){
     ambient::actor proc(where(i,j));
     ambient::numeric::gemm_fma(a.tile(i,k), b.tile(k,j), c.tile(i,j));
   }
   ambient::sync();
   return 0;
}
```

- linear algebra containers
  - objects versioning
  - BLAS / LAPACK callbacks
- kernels
  - automatic task-queue
  - automatic dependency tracking
- parallel scopes
  - pick *actors* / MPI processes
  - validity defined by C++ RAII
- sync
  - launch parallel execution
  - MPI + OpenMP / Intel® Cilk™

dataflow parallelism





ambient.comp-phys.org

DMRG : White. *Phys. Rev. Lett.* **69**, 2863 (1992), U. Schollwöck, *Ann. Phys. (N.Y).* **326**, 96–192 (2011).

2D DMRG : E. M. Stoudenmire and S. R. White, *Annu. Rev. Condens. Matter Phys.* **3**, 111–128 (2012).

Our code: M. Dolfi, B. Bauer, S. Keller, A. Kosenkov, T. Ewart, A. Kantian, T. Giamarchi, and M. Troyer, *Comput. Phys. Commun.* **185**, 3430–3440 (2014).

Hubbard Ladders : Noack, White, Scalapino, *Physica C*, **270**, 281 (1996), Noack, Bulut, Scalapino, Zacher. *Phys. Rev. B* **56**, 7162 (1997).