EuroCC4SEE Workshop Belgrade, 20 May 2025

Data compression and parallelization in simulations of transient regimes

Jeremija Kovačević

Scientific Computing Laboratory Center for the Study of Complex Systems Institute of Physics Belgrade Belgrade, Serbia

Scientific Computing Laboratory

Funded by the European Union

European Research Council
Established by the European Commission

Institute of Physics Belgrade

Motivation

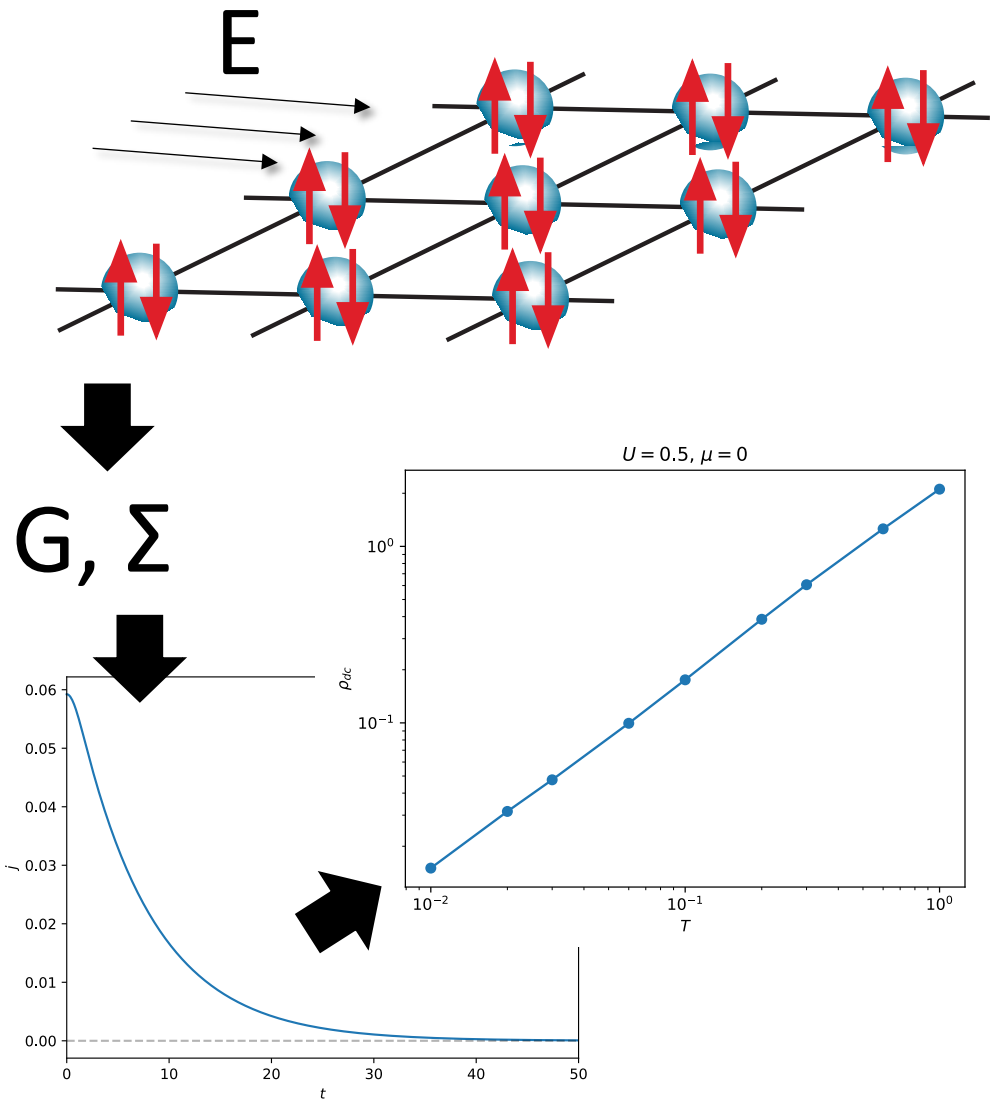
- Our goal is to understand and predict the physical and electronic properties of materials
- Central quantities of interest are the dynamical response functions, particularly electrical conductivity

$$X_i(\mathbf{r},t) = \int d^d r' \int dt'\, \chi_{ij}(\mathbf{r},t;\mathbf{r}',t') F'_j(\mathbf{r}',t')$$

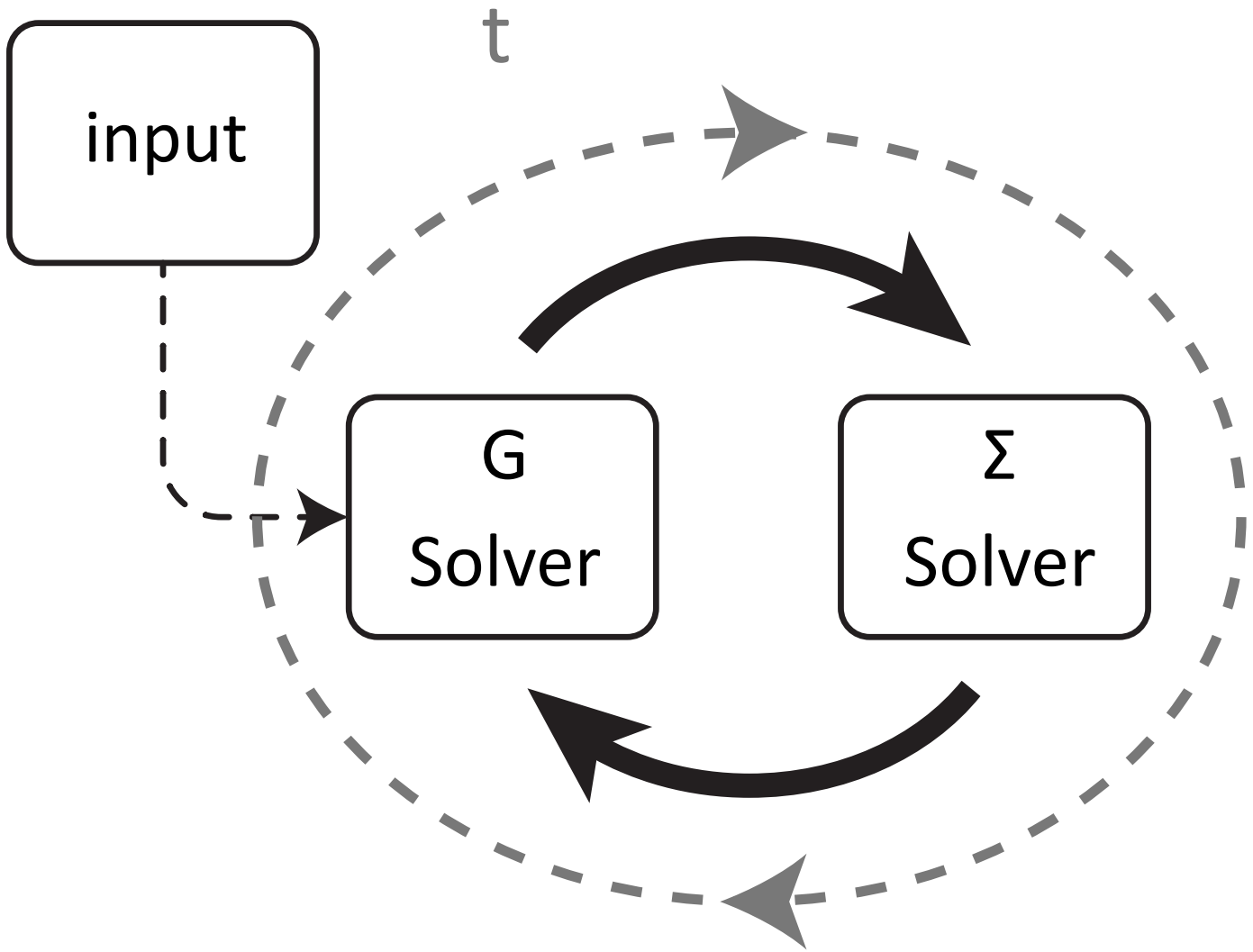
- Dynamical response functions encode the intrinsic properties of the system
- In equilibrium, calculate dynamical response function directly (no transient regimes)
- In Non-equilibrium, Probing the models with external fields and calculating the response

What we do

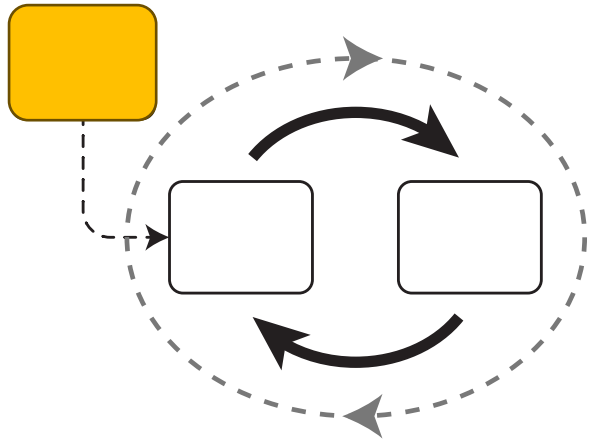
- Simulate an electron system on a lattice under a short electric pulse E
- Compute the Green's function $G_r(t,t')$: how electrons propagate over time
- Compute the self-energy $\Sigma_r(t,t')$: how interactions modify propagation
- Calculate current j from G
- Extract conductivity form the current
- Vary physical parameters (T, U, μ)
- Two natural domains: site space and k-space



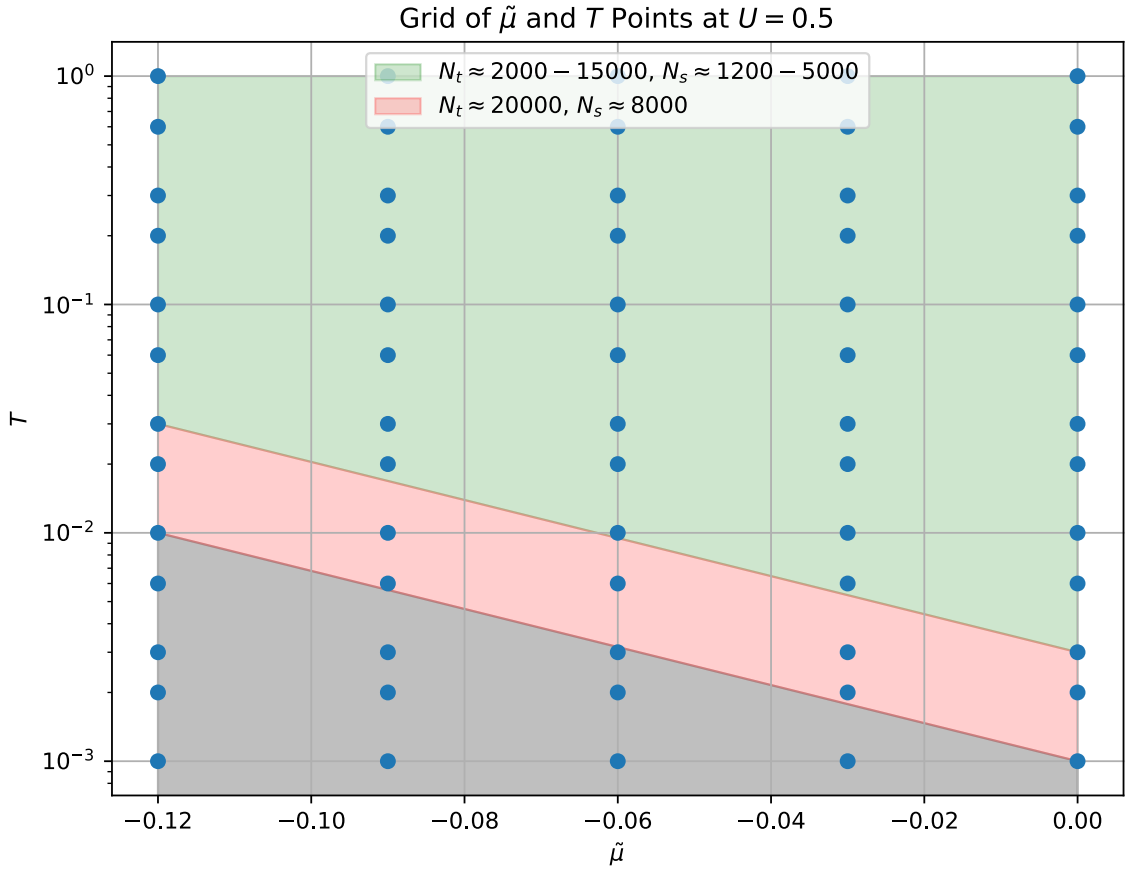
Algorithm



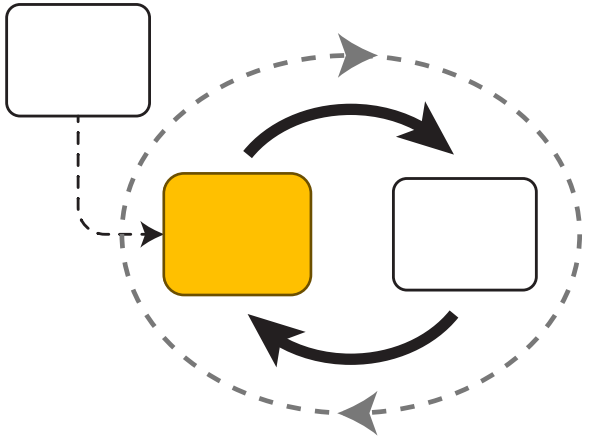
Input



- Temperature T
- Interaction strength U
- Doping level μ
- Lattice type
- Number of lattice sites N_s
- Number of time points N_t
- Time step discretization Δt
- Profile of the electric field pulse



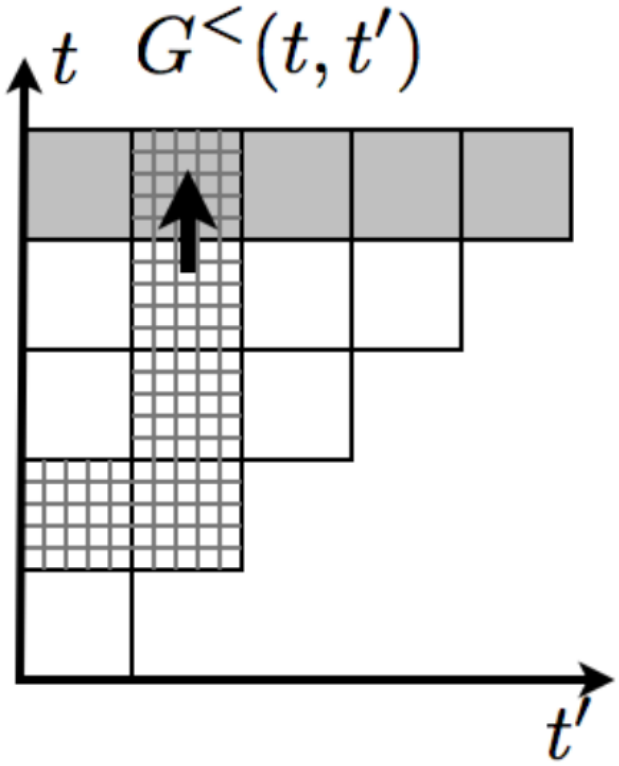
Calculating G



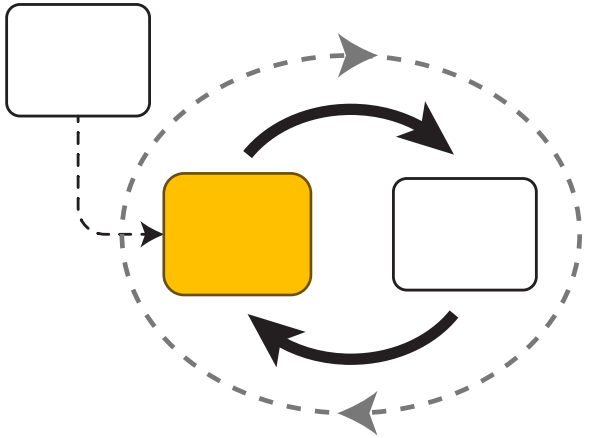
- Kadanoff-Baym equations (KBE)

$$i\partial_t G_{\mathbf{k}}(t,t') = \varepsilon_{\mathbf{k}}(t) G_{\mathbf{k}}(t,t') + \int_0^t \mathrm{d}\bar{t}\, \Sigma_{\mathbf{k}}(t,\bar{t}) G_{\mathbf{k}}(\bar{t},t')$$

- Self-consistent equations
- Time-stepping algorithm
- History integrals require knowledge of G and Σ for all times
- Embarrassingly parallel over k points
- NESSi (Computer Physics Communications 257, 107484 (2020))



Scaling of the KBE Solver

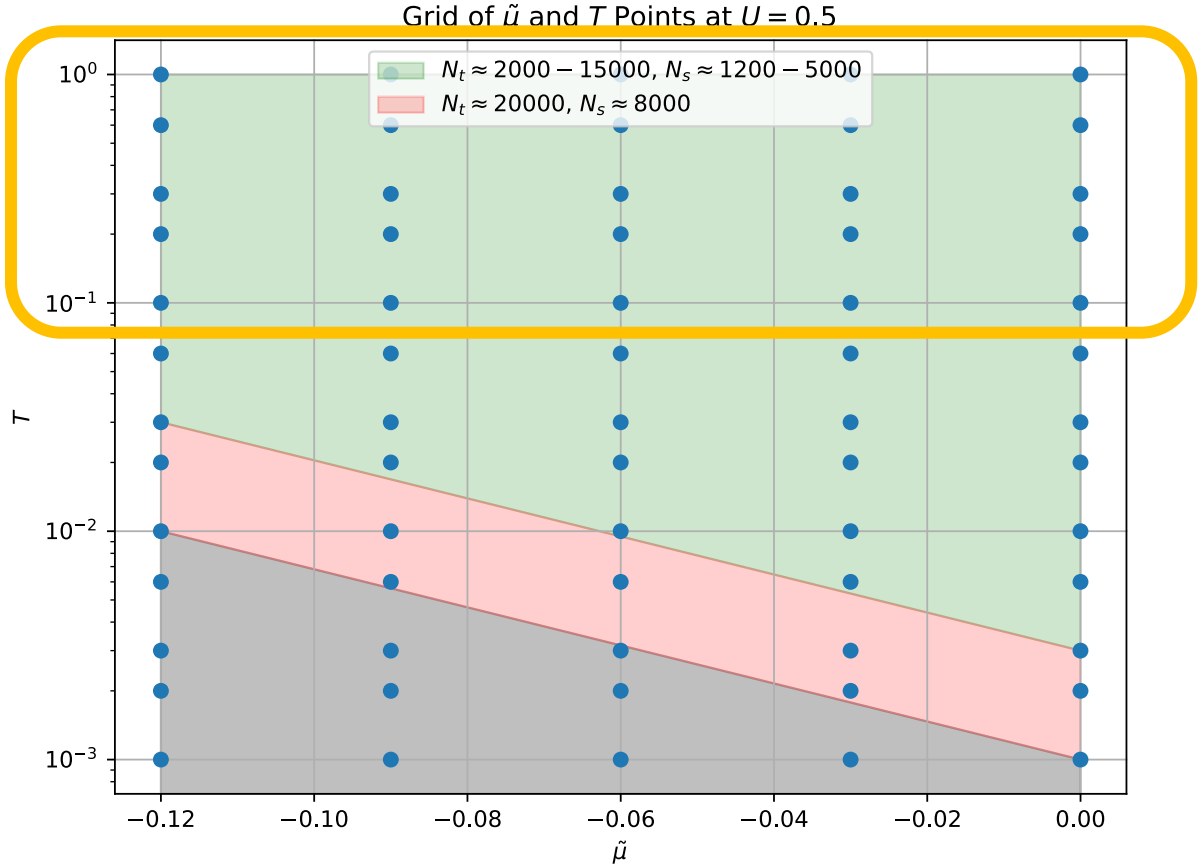


$$i\partial_t G_{\mathbf{k}}(t,t') = \varepsilon_{\mathbf{k}}(t) G_{\mathbf{k}}(t,t') + \int_0^t \mathrm{d}\bar{t}\, \Sigma_{\mathbf{k}}(t,\bar{t}) G_{\mathbf{k}}(\bar{t},t')$$

- Memory cost of G and Σ

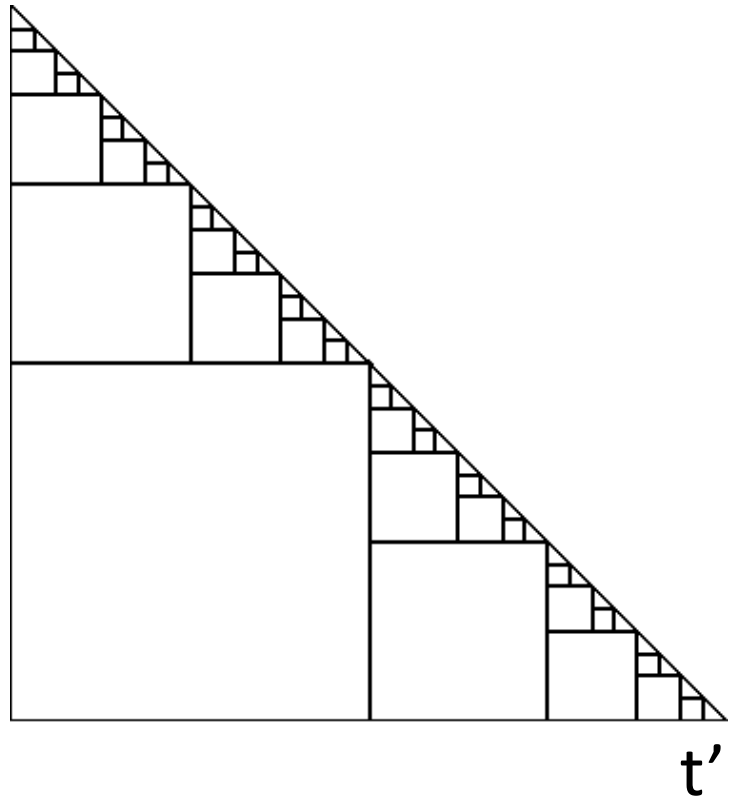
  $N_t = 10000$, $N_s = 5000$

  $2 \times N_t^2 \times N_s \times 16\mathrm{B} \approx 15\ \mathrm{TB}$
- Computational cost scales as N_t^3
- Large part of the parameter space is unavailable



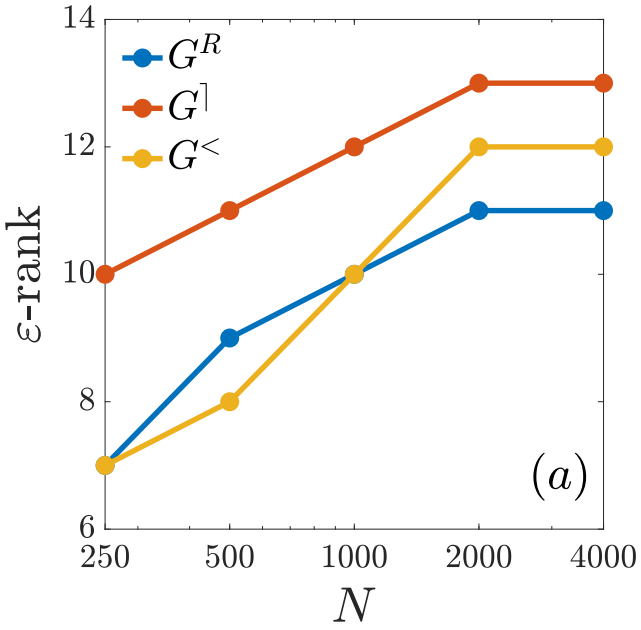
Data compression with SVD

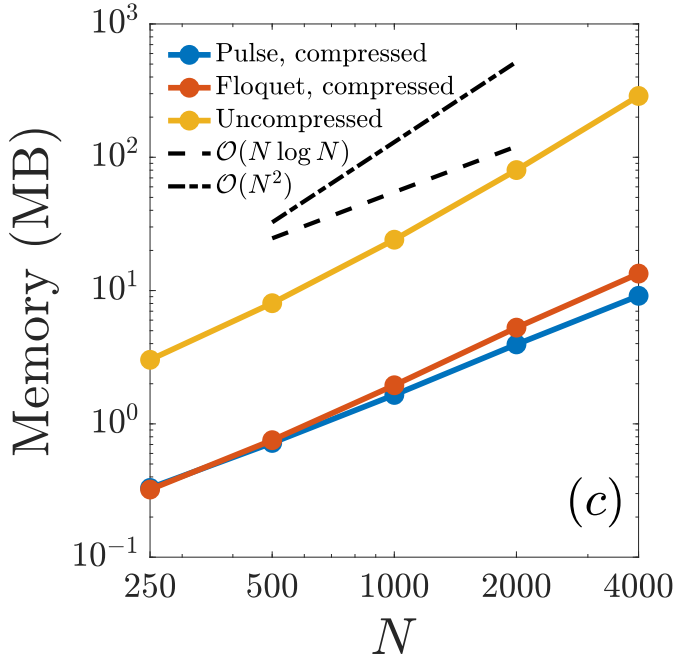
- Key Idea: Far from the diagonal there is less and less information
- In that region there should be a low-rank representation
- Hierarchically off-diagonal low-rank (hodlr)
- Small control parameter to control the scaling
- Matrix partition and perform SVD on each block
- We never save all the blocks, just their SVD representation



Data compression with SVD

- Ranks of blocks are much smaller than N and grow weakly with N
- Memory usage grows logarithmically
- History integrals are evaluated with SVDs which reduces costs
- Reduction of computational cost $N^3 \rightarrow N^2 \log(N)$
- Memory requirements $N^2 \rightarrow N \log(N)$
- Example: 5 months, 2.2TB → 26.5h, 3.8GB
- Lossy compression
- We don't know RAM usage ahead of time
- hodlr (SciPost Phys. 10, 091 (2021))





Self-Energy calculation

- KB solver operates naturally in k-space
- Cost of Σ calculation:

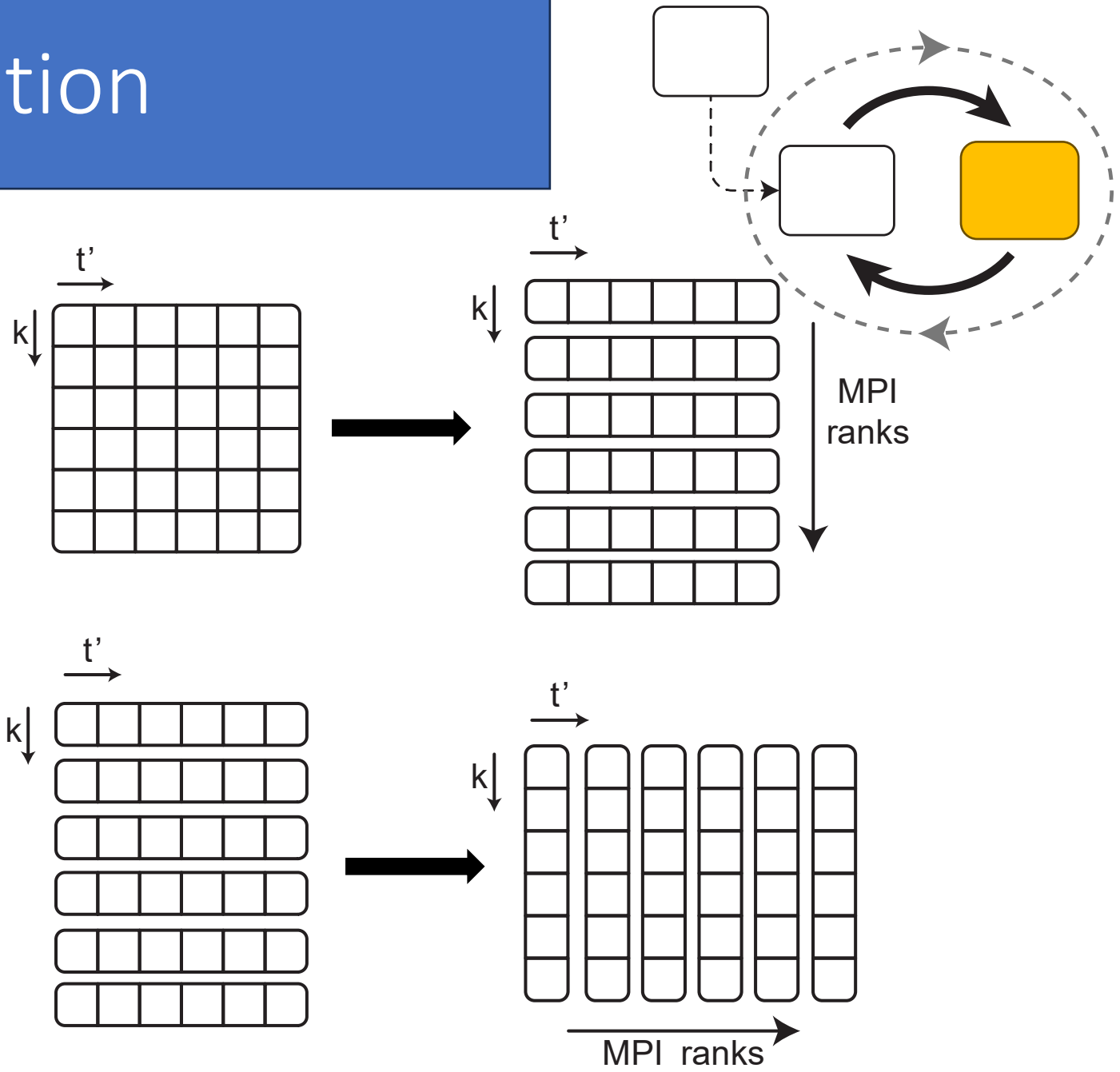
  $O(N_s^3)$ in k-space

  $O(N_s)$ in site space
- Cost of a Fourier transform:

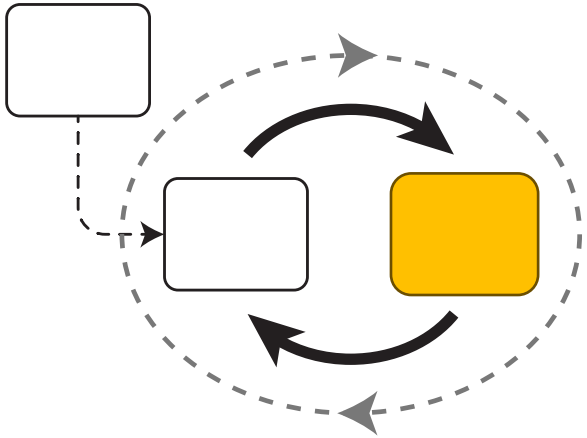
  $O(N_s \log(N_s))$
- A way to go:

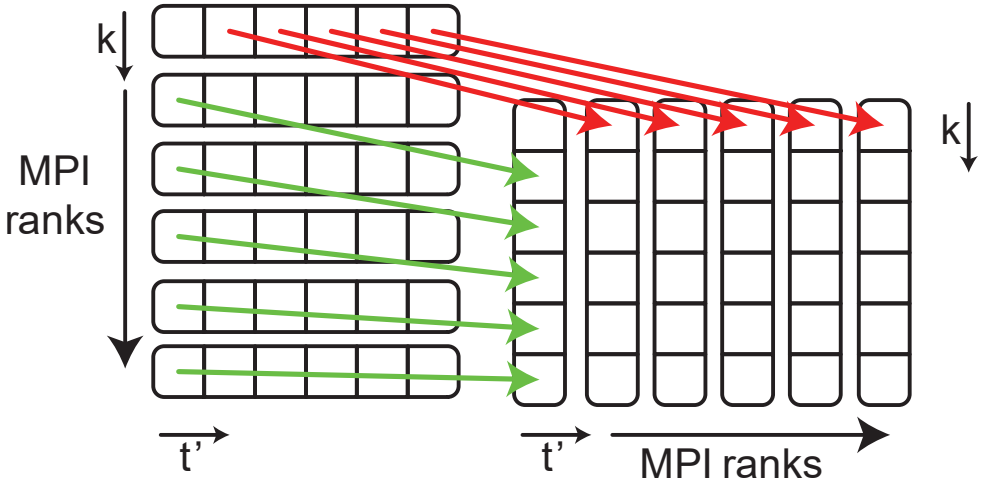
  $G_k \rightarrow G_r \rightarrow \Sigma_r \rightarrow \Sigma_k$

  $O(N_s \log(N_s))$



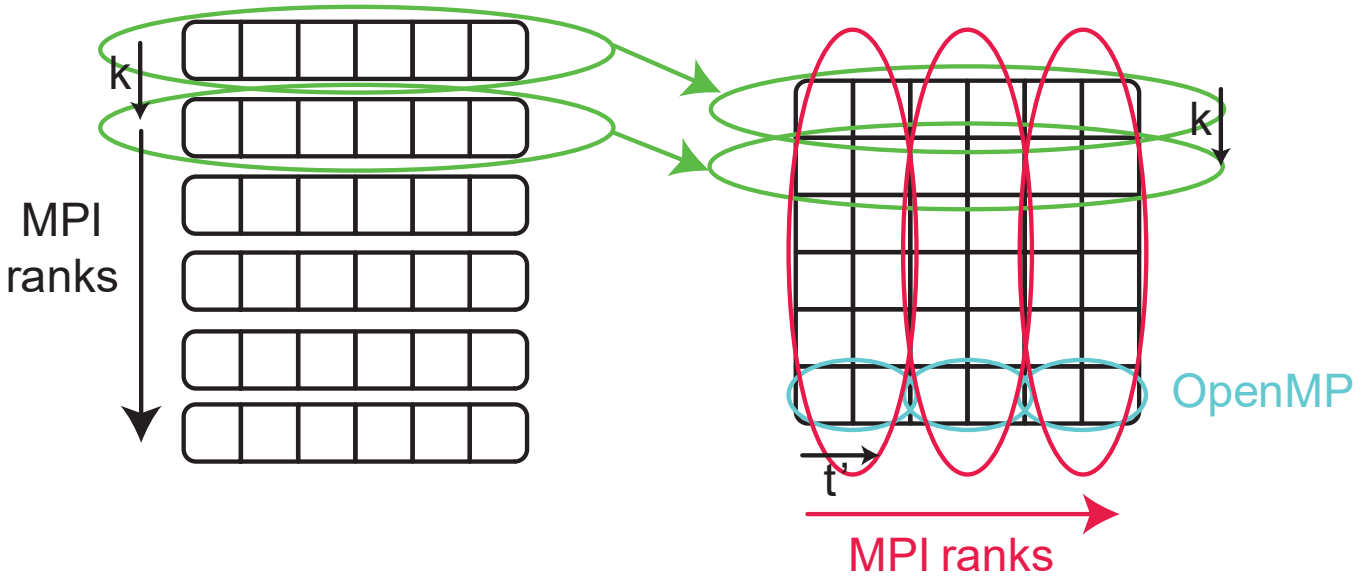
Parallelization strategies





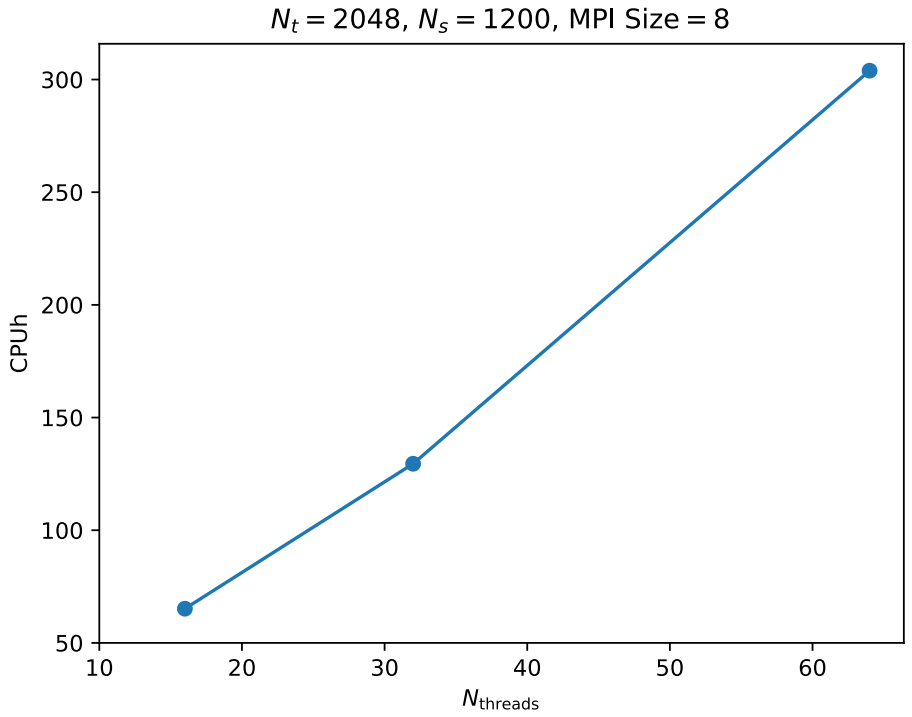
- MPI Send and Receive
- Simple to implement
- Too many communications

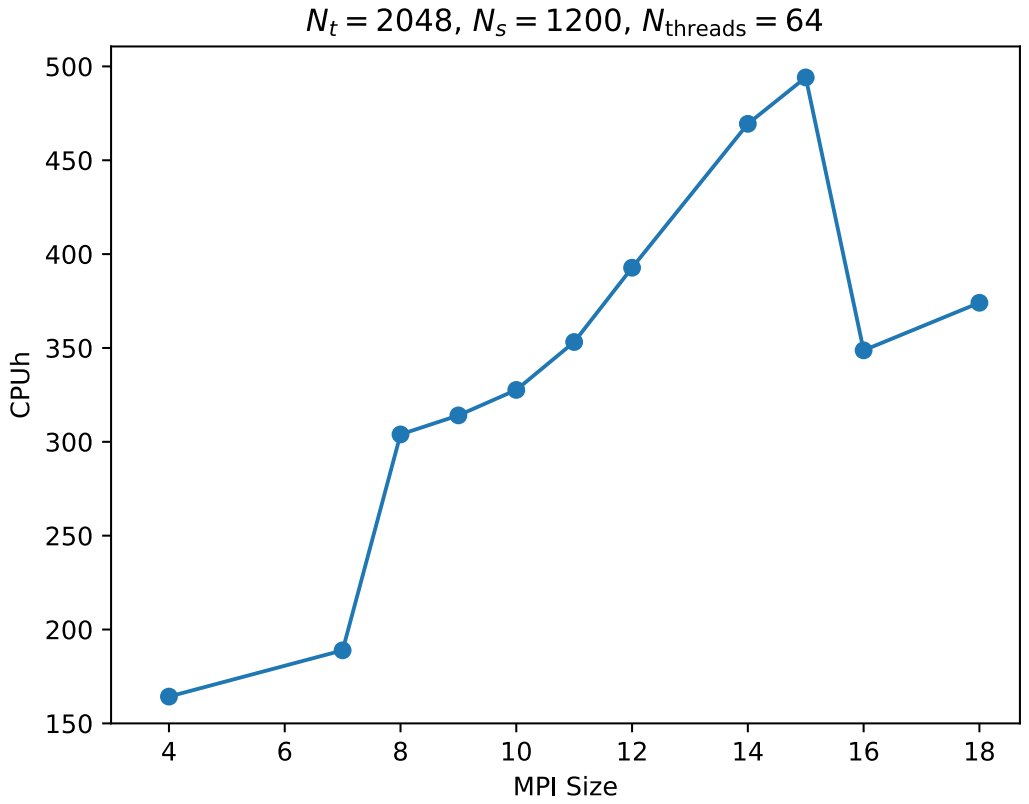
- MPI gather + OpenMP
- Harder to implement
- Less communications
- Better scaling



Scaling with MPI size and thread count

- Complicated scaling
- MPI communications become a bottleneck
- Threads are wasted





- Spawning threads can be costly

Summary and Conclusions

- We computed G and Sigma in the Kadanoff-Baym formalism to extract conductivity in the interacting electron system.
- Two-time nature of the problem leads to huge memory and computational costs.
- Our block-wise SVD compression reduces memory and computational complexity dramatically.
- Using FFT we switch between site and momentum space optimizing different parts of the calculation.
- Hybrid MPI+OpenMP offers complicated but better scaling options.