An Algebraic Sparsified Nested Dissection Algorithm using Low-Rank Approximations

Léopold Cambier* (lcambier@stanford.edu), Chao Chen*, Eric Darve*

Erik Boman†, Sivasankaran Rajamanickam†, Raymond Tuminaro†

* Stanford University (ICME/ME), † Sandia National Laboratory

Conference on Fast Direct Solvers, Purdue University, November 9, 2018

† Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525

Linear Systems

We want to solve $Ax = b$

Direct Methods

(Sparse) LU + Ordering

- Robust
- Accurate
- Generic
- Costly

Approximate Factorizations

ILU / ND + H-algebra …

- Tunable accuracy
- Tunable cost
- Relatively generic

Iterative methods

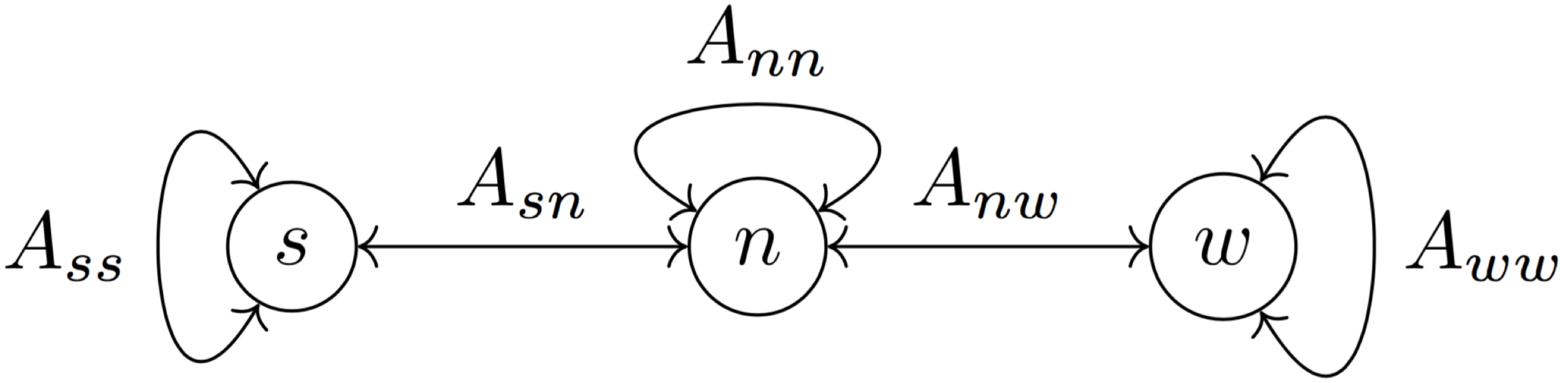
+ Custom Preconditioner

- Cheap
- Specific

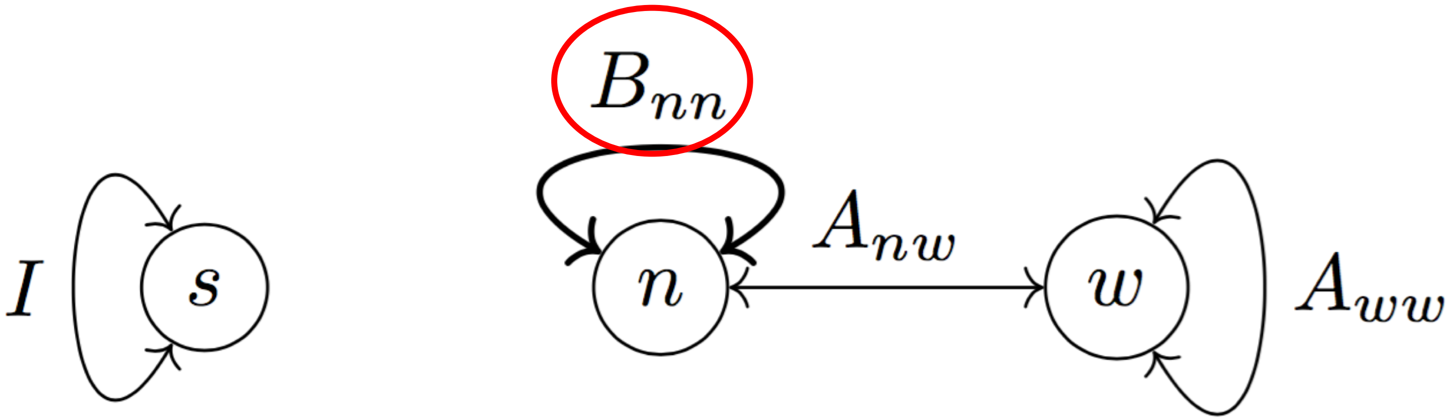
Our approach is heavily inspired by K. L. Ho and L. Ying, Hierarchical interpolative factorization for elliptic operators: differential equations, Communications on Pure and Applied Mathematics, 69 (2016), pp. 1415–1451

Sparse Linear Systems

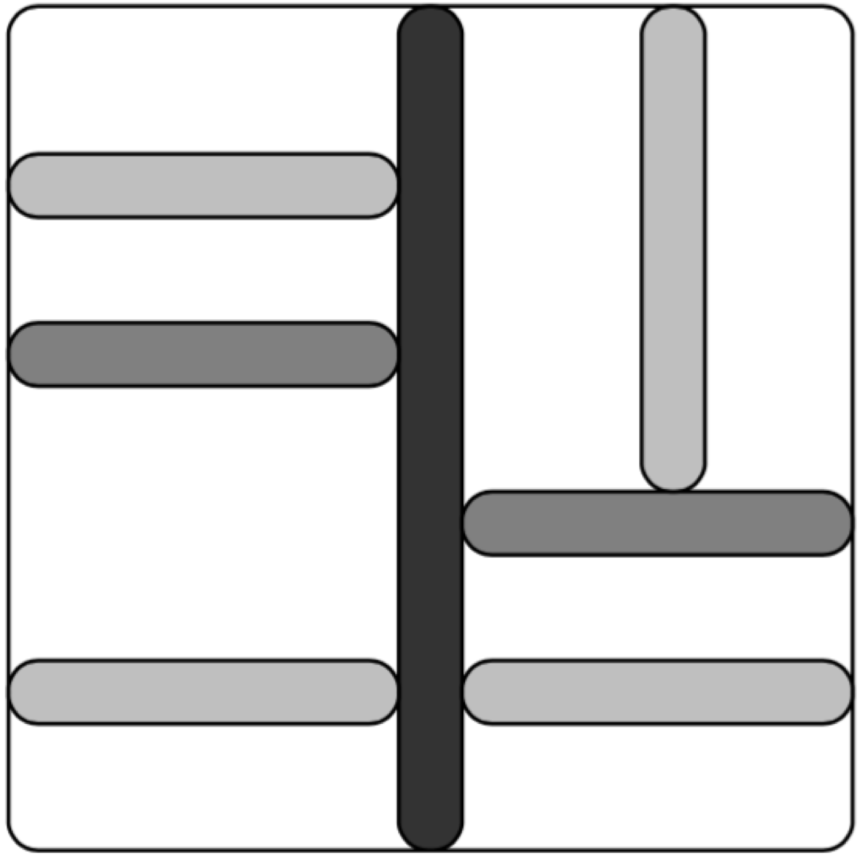
$$\begin{bmatrix} A_{ss} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

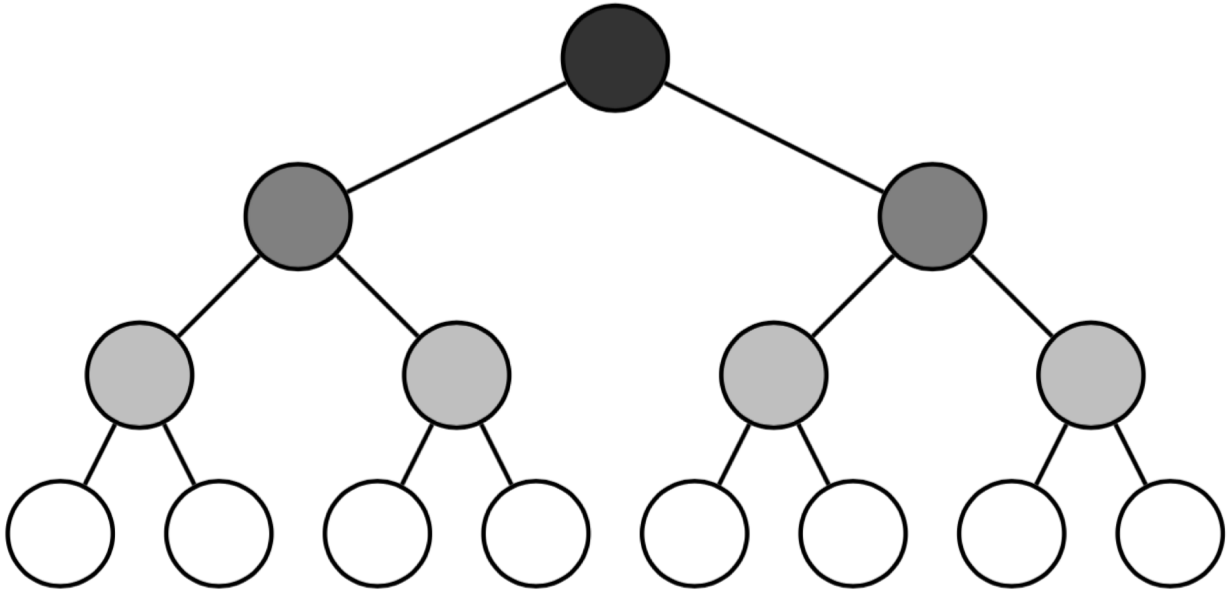


$$\begin{bmatrix} I & & \\ & A_{ww} & A_{wn} \\ & A_{nw} & B_{nn} \end{bmatrix}$$



Nested Dissection

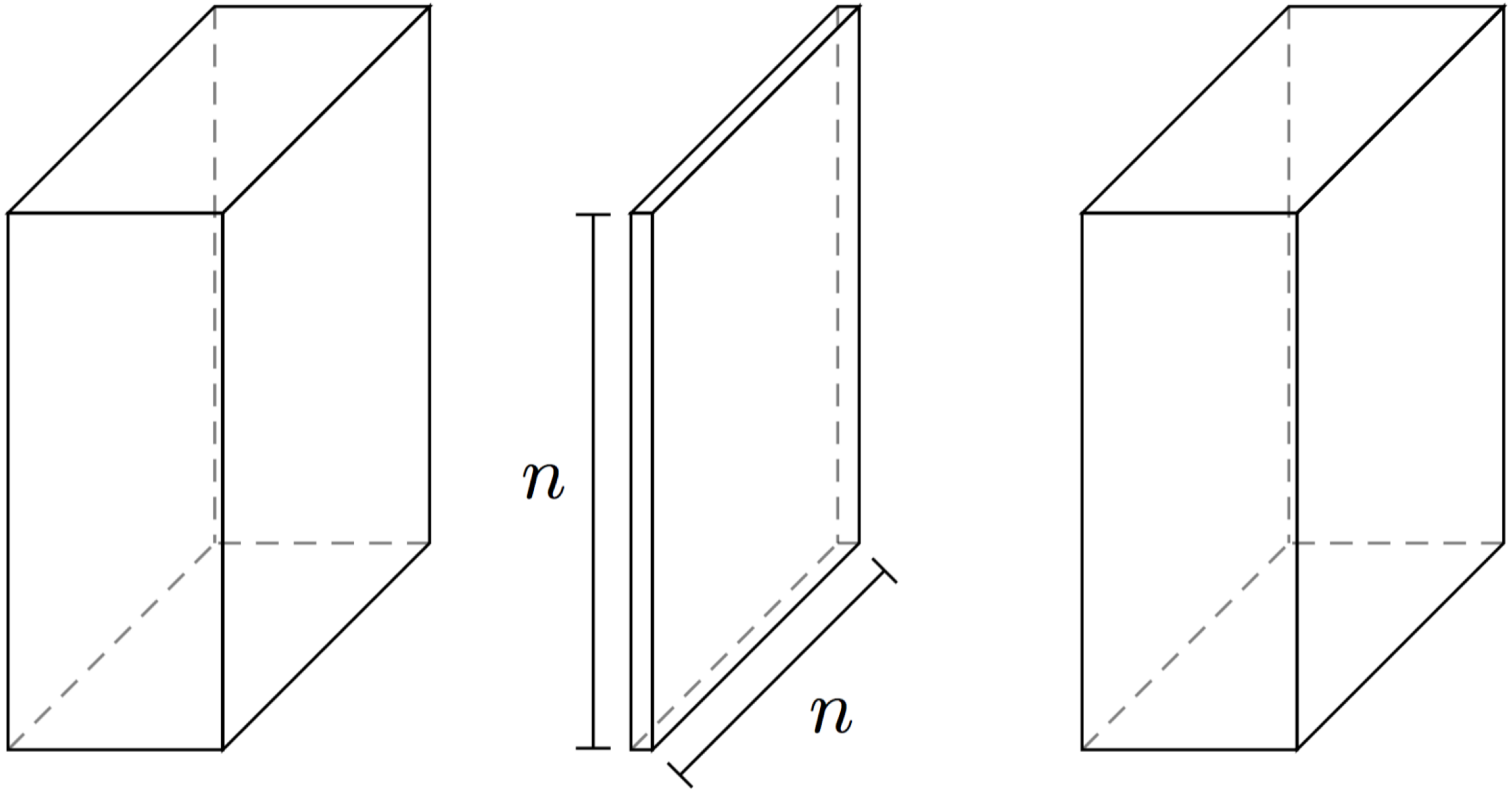




Nested Dissection

Issue: separators are small, but still too big on typical 3D problems

$N = n^3$
Separator: n^2
Fact. cost: $n^{2 \cdot 3} = N^2$



Sparsification I

(1) We start with

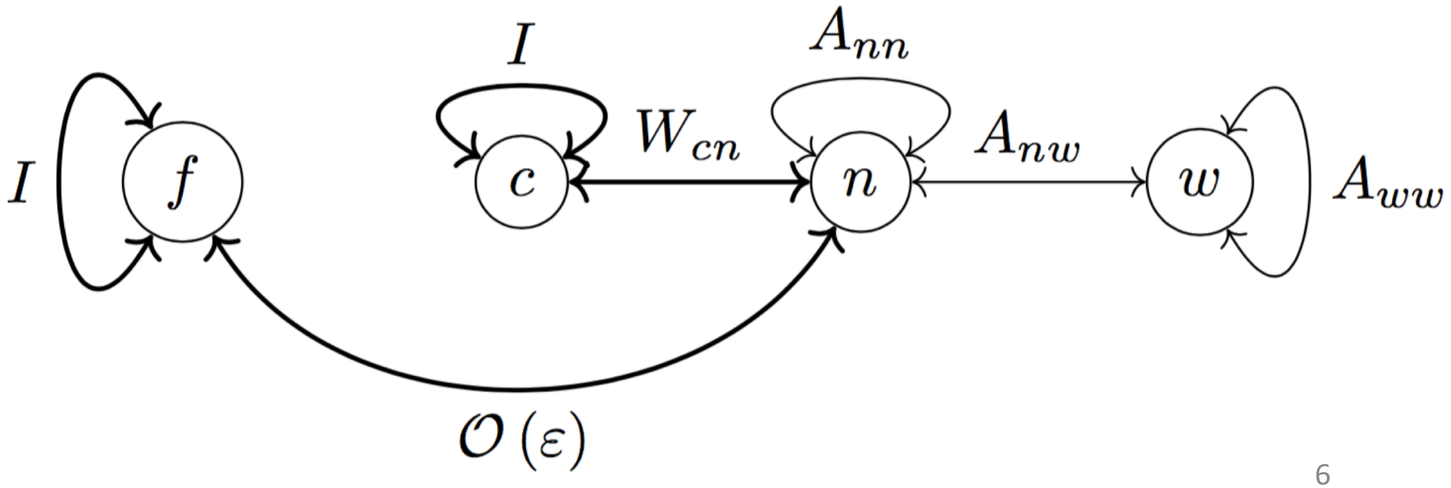
$$\begin{bmatrix} I & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

(2) We then approximate

$$A_{sn} = Q_{sc} W_{cn} + \varepsilon$$
$$Q^\top s = f \cup c$$

(3) We end up with

$$\begin{bmatrix} I & & & \varepsilon \\ & I & & W_{cn} \\ & & A_{ww} & A_{wn} \\ \varepsilon & W_{cn}^\top & A_{nw} & A_{nn} \end{bmatrix}$$



Sparsification I

$$\underbrace{\underbrace{\begin{bmatrix} L_{ss}^{-1} & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} A_{ss} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix} \begin{bmatrix} L_{ss}^{-\top} & & \\ & I & \\ & & I \end{bmatrix}}_{\begin{bmatrix} Q^{\top} & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} I & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix} \begin{bmatrix} Q & & \\ & I & \\ & & I \end{bmatrix}}}_{\begin{bmatrix} I & & & \varepsilon \\ & I & & W_{cn} \\ & & A_{ww} & A_{wn} \\ \varepsilon & W_{cn}^{\top} & A_{nw} & A_{nn} \end{bmatrix}}$$

Sparsification II

(1) We start with

$$\begin{bmatrix} A_{ss} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

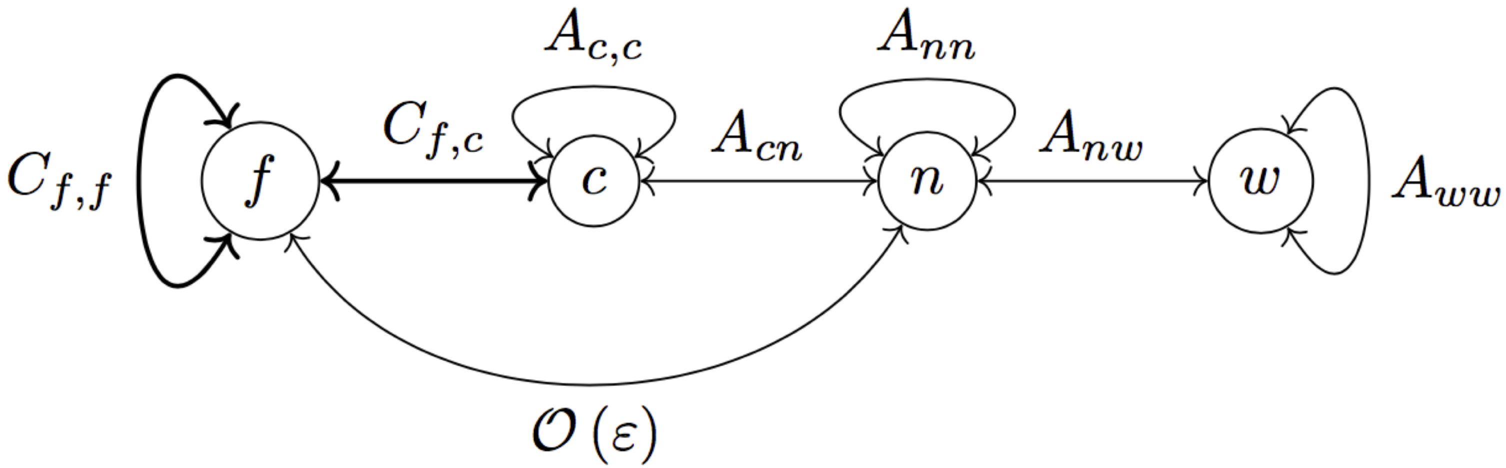
(2) We then approximate

$$A_{sn} = \begin{pmatrix} T_{fc} \\ I \end{pmatrix} A_{cn} + \varepsilon$$

$$s = f \cup c$$

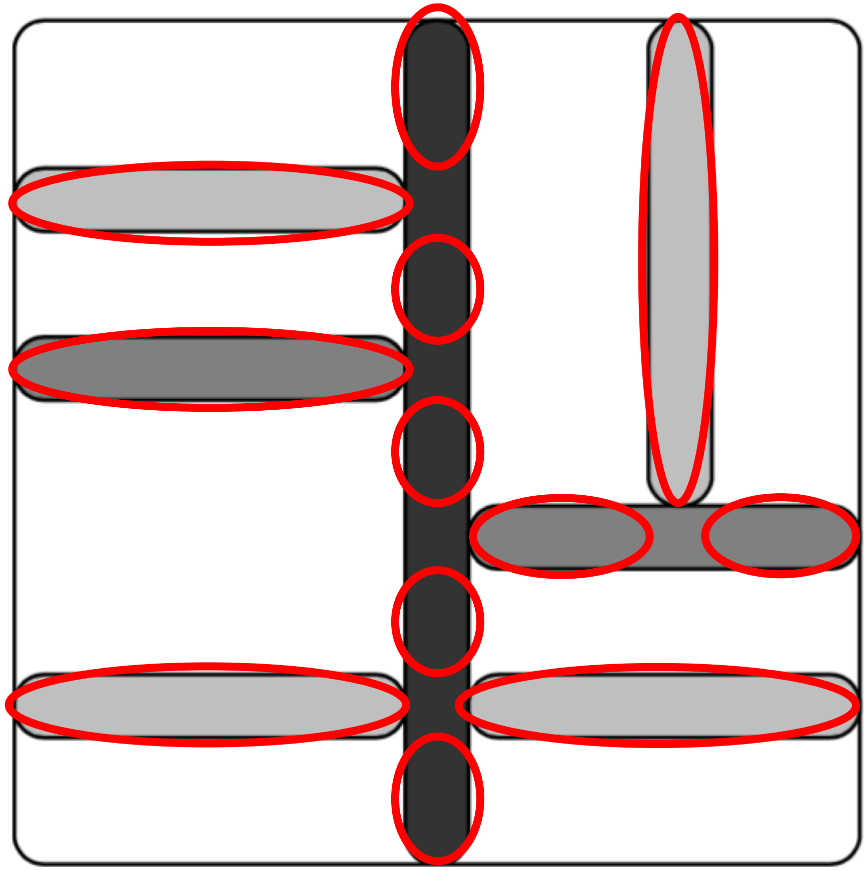
(3) We end up with

$$\begin{bmatrix} C_{ff} & C_{fc} & & \varepsilon \\ C_{cf} & A_{cc} & & A_{cn} \\ & & A_{ww} & A_{wn} \\ \varepsilon & A_{nc} & A_{nw} & A_{nn} \end{bmatrix}$$

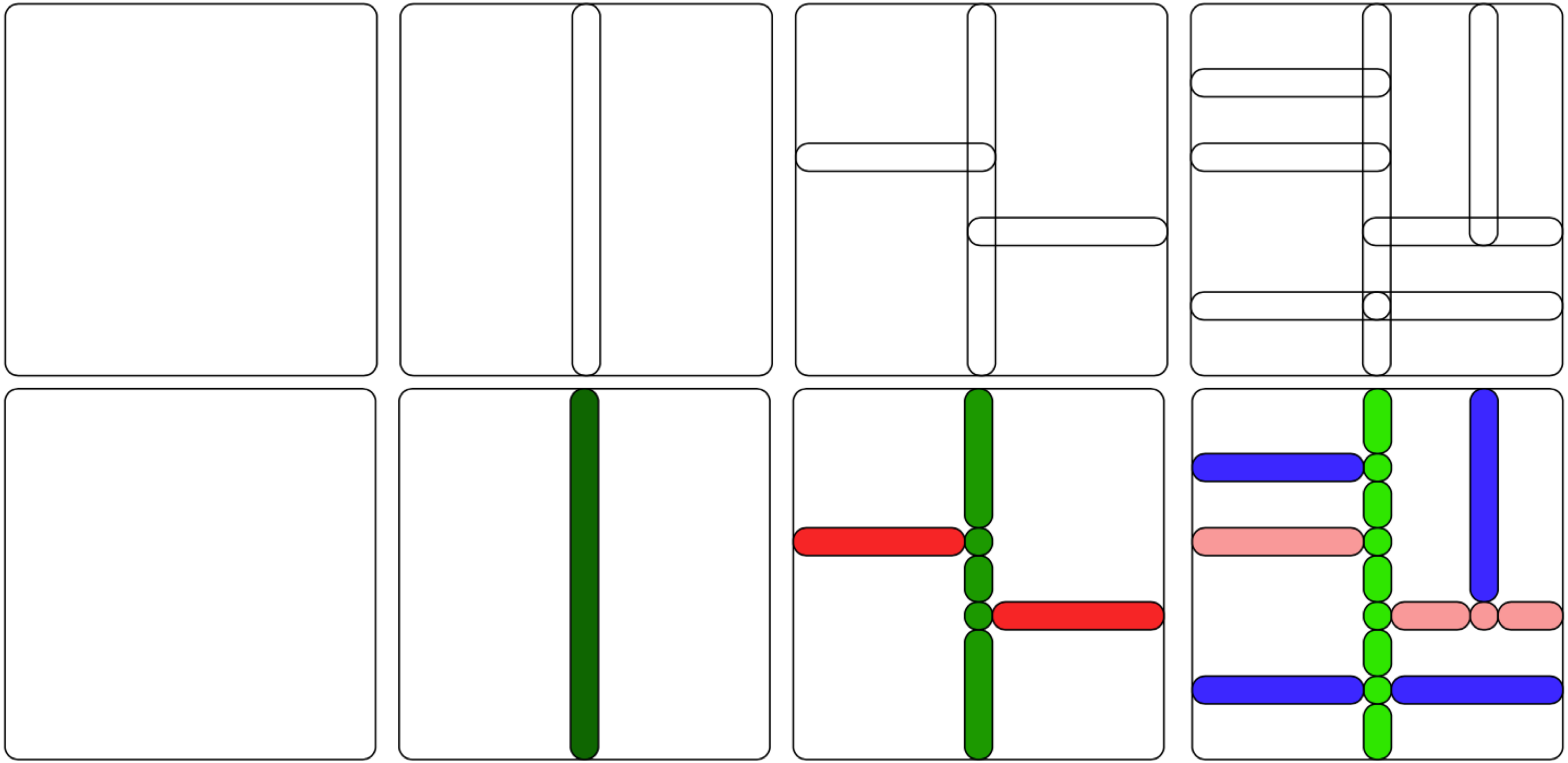


What do we sparsify?

Interfaces between eliminated-interiors



How do we find those interfaces? Coloring

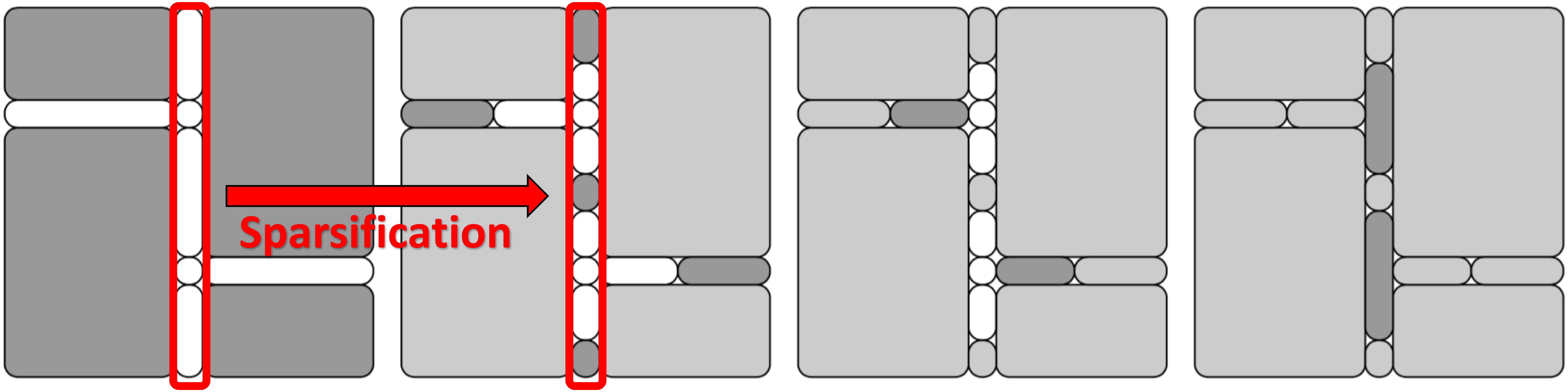


Sparsified Nested Dissection

For level ℓ, from leaves to top

Eliminate interiors at level ℓ

(Scale &) Sparsify interfaces at level ℓ



Sparsified Nested Dissection

We effectively build a preconditioner P such that

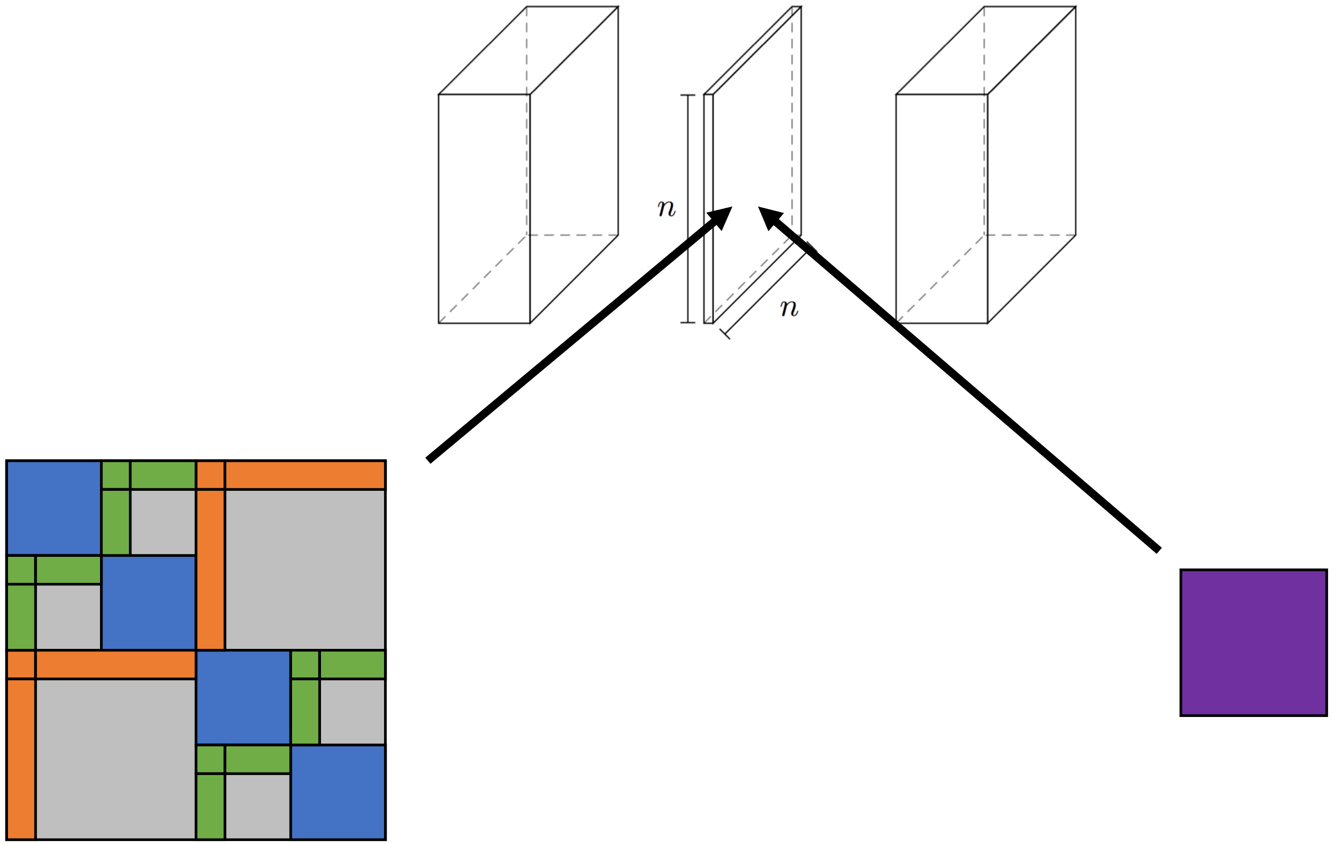
$$P^\top A P \approx I + \varepsilon$$

Where P is a sequence (product) of

- Eliminations
- (Scalings)
- Sparsifications

We then use P as a preconditioner for CG

Different from fast-algebra techniques



Sparsification I & II

$$\begin{bmatrix} \boldsymbol{A_{ss}} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

$$A_{sn} = \binom{T_{fc}}{I} A_{cn} + \varepsilon$$

OR

$$\begin{bmatrix} \boldsymbol{I} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

$$A_{sn} = \binom{T_{fc}}{I} A_{cn} + \varepsilon$$

$$A_{sn} = Q_{sc} W_{cn} + \varepsilon$$

Orthogonal, with scaling, stays SPD

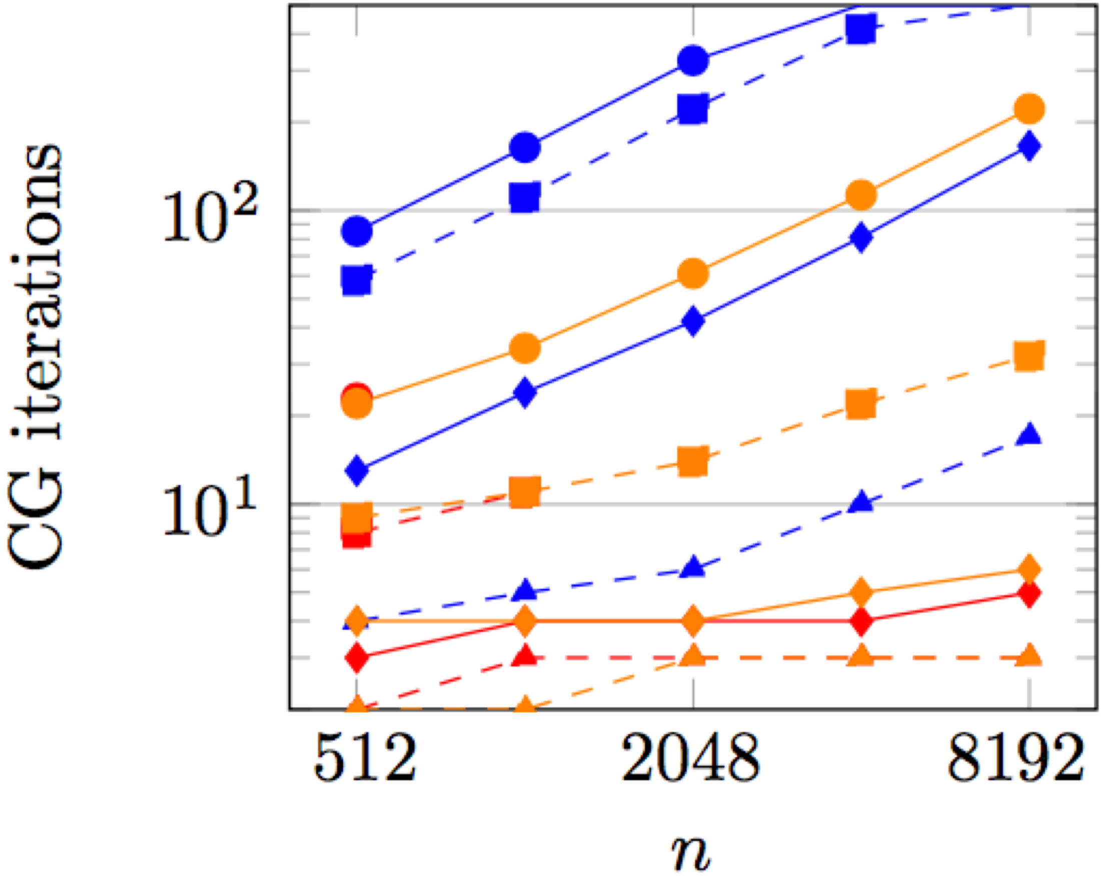
$$\begin{bmatrix} \boldsymbol{I} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

$$A_{sn} = Q_{sc} W_{cn} + \underbrace{\boxed{Q_{sf} W_{fn}}}^{O(\varepsilon)}$$

$$\boxed{S_{nn} = A_{nn} - W_{cn}^{\top} W_{cn}} - W_{cf}^{\top} W_{cf}$$

Approximate Schur Complement over (n,n)

Low-Rank Compression: three variants (2D Laplacians)



Interpolative, no scaling

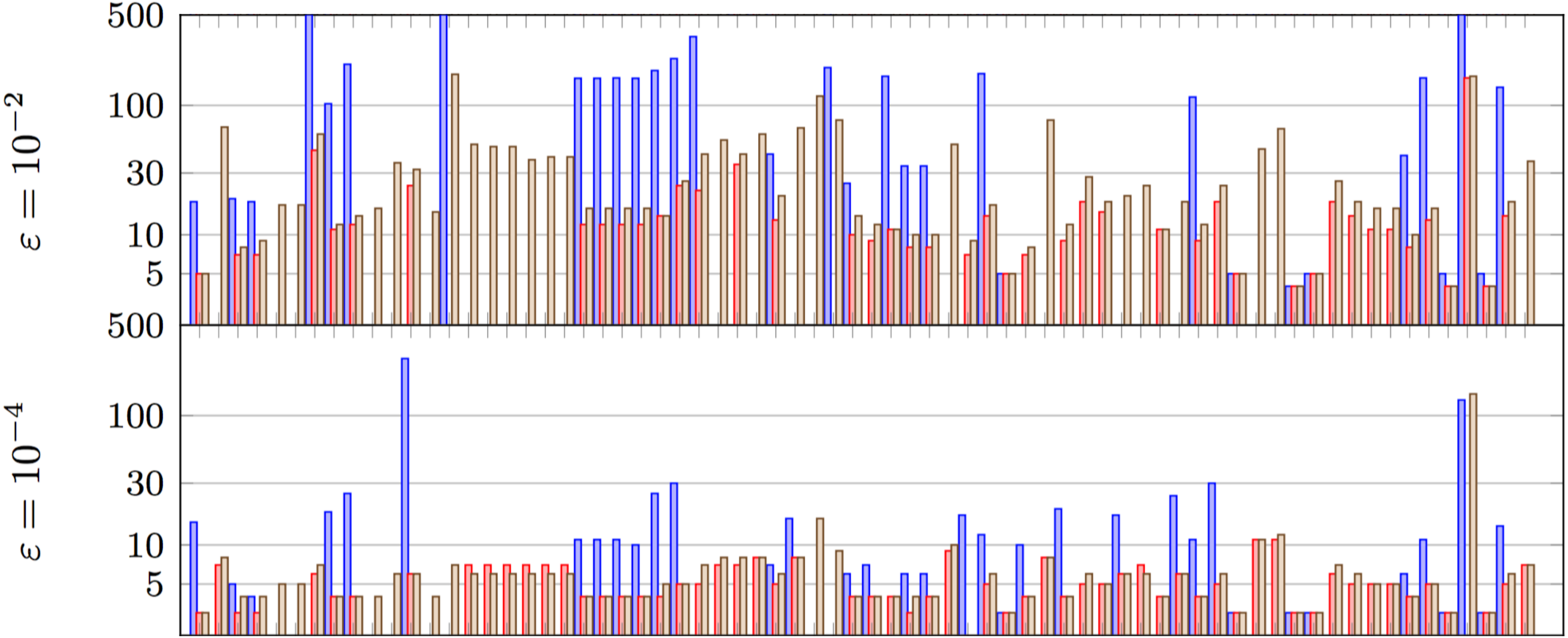
Interpolative, with scaling

Orthogonal, with scaling

$$\varepsilon = 10^{-1} \rightarrow 10^{-6}$$

SPD problems from SuiteSparse

Interpolative, no scaling
Interpolative, with scaling
Orthogonal, with scaling



SPD problems from SuiteSparse

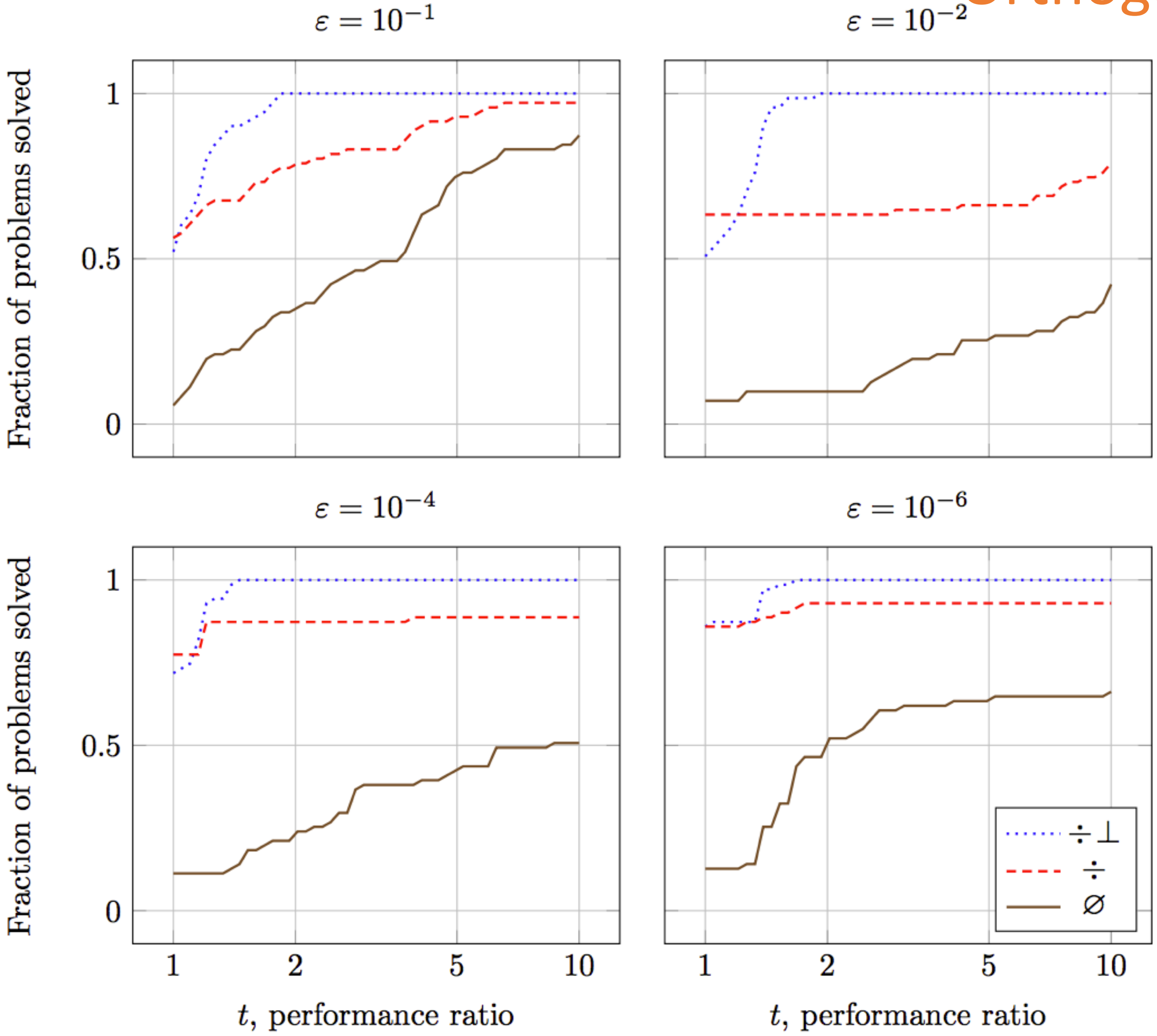
Interpolative, no scaling

Interpolative, with scaling

Orthogonal, with scaling

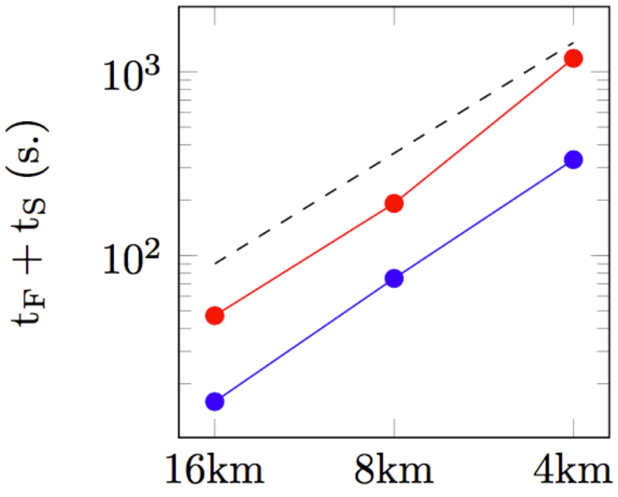
Perf ratio(t) =

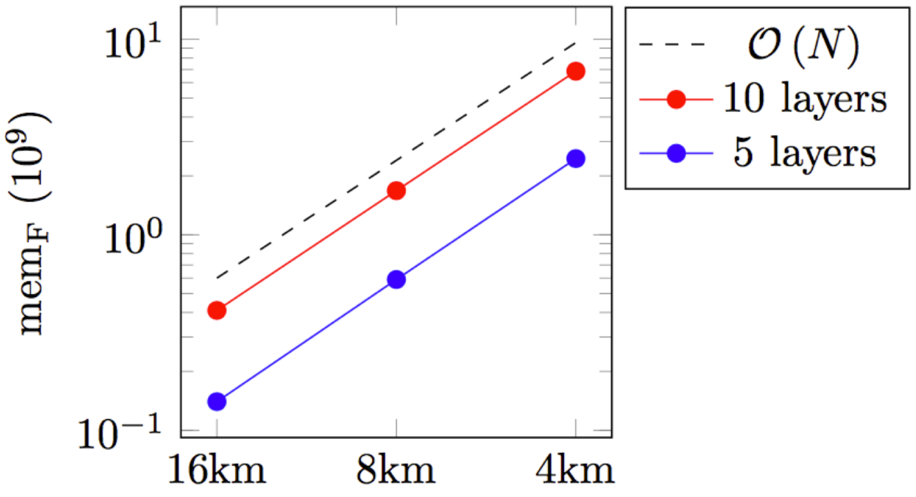
$$\frac{\#\left\{p \in P \,\middle|\, \frac{CG_{pv}}{CG_p^*} \leq t\right\}}{\#P}$$

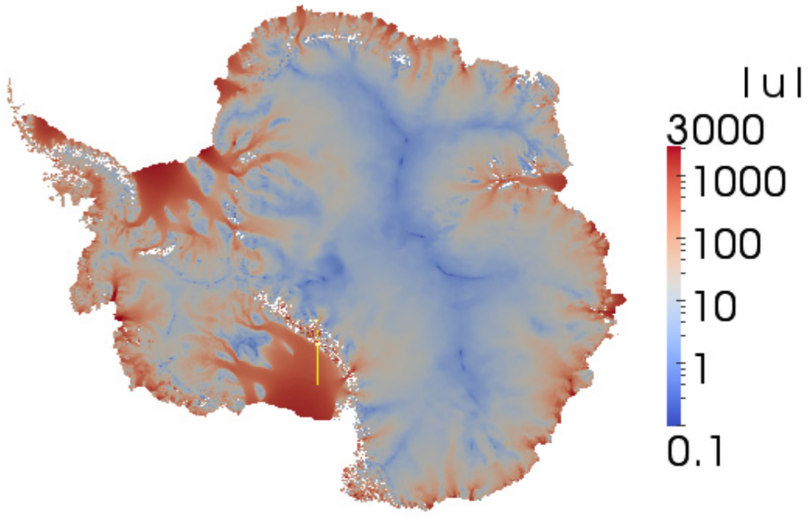


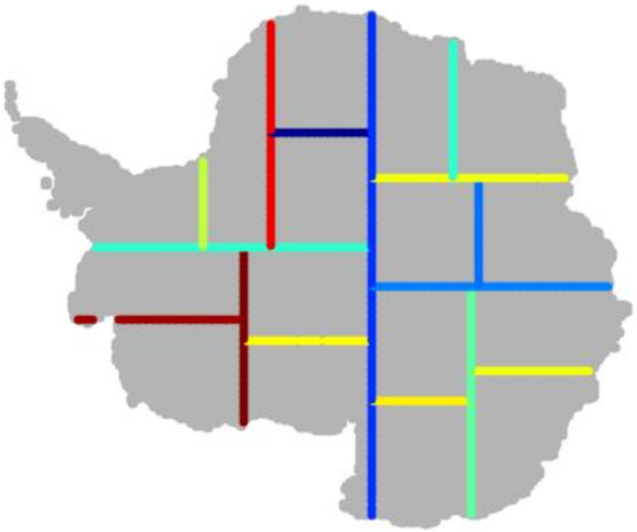
Ice-Sheet modeling $\kappa(A) > 10^{11}$

N	spaND t_F (s.)	t_S (s.)	n_{CG}	size_{Top}	mem_F (10^9)	Direct $t_F + t_S$ (s.)
5 layers						
629 544 (16 km)	13	3	7	76	0.14	22
2 521 872 (8 km)	55	20	8	89	0.59	206
10 096 080 (4 km)	217	115	10	100	2.45	1578
10 layers						
1 154 164 (16 km)	39	8	7	136	0.41	90
4 623 432 (8 km)	148	44	8	148	1.68	710
18 509 480 (4 km)	798	384	10	159	6.86	—



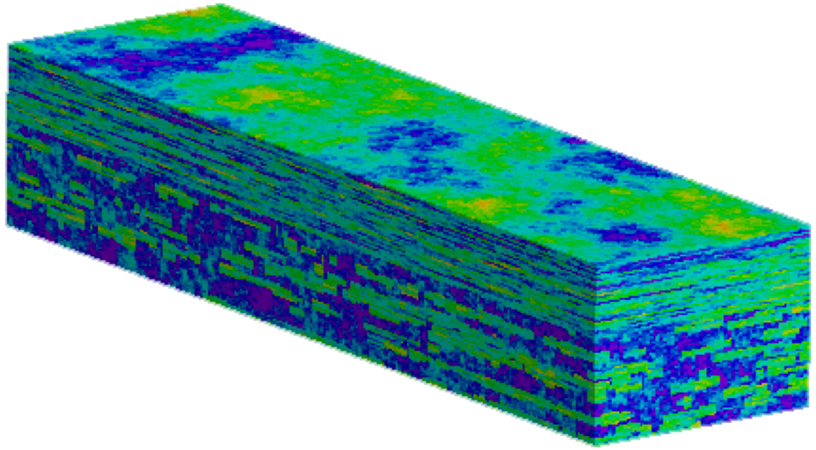


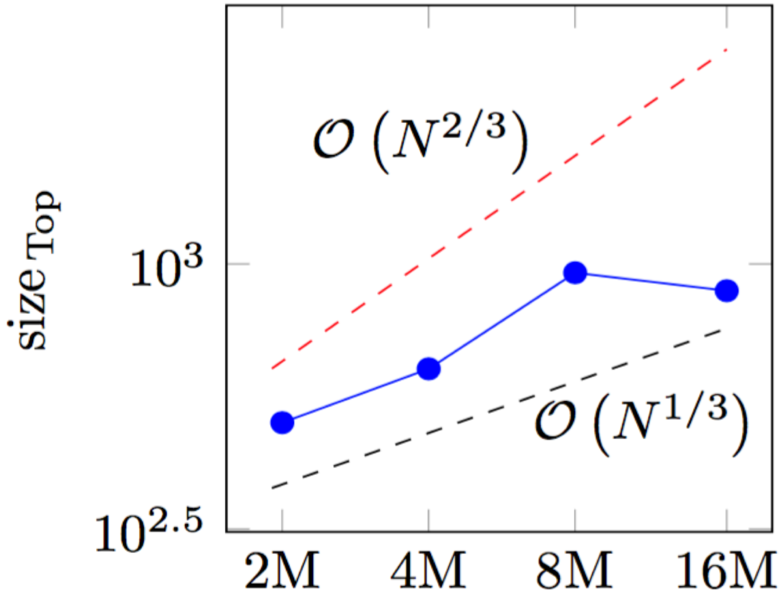


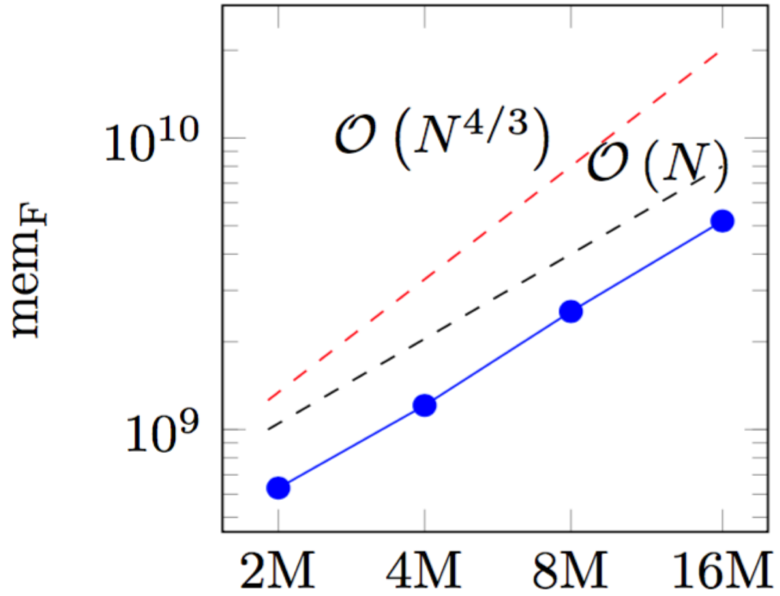


The SPE problem

n	$N = n^3$	spaND $\mathrm{t_F}$ (s.)	$\mathrm{t_S}$ (s.)	$\mathrm{n_{CG}}$	$\mathrm{size_{Top}}$	$\mathrm{mem_F}$ (10^9)	Direct. $t_F + \mathrm{t_S}$ (s.)
128	2 097 152	61	23	12	502	0.63	686
160	4 096 000	175	46	13	634	1.21	—
200	8 000 000	287	158	16	962	2.54	—
252	16 003 008	963	369	16	890	5.19	—



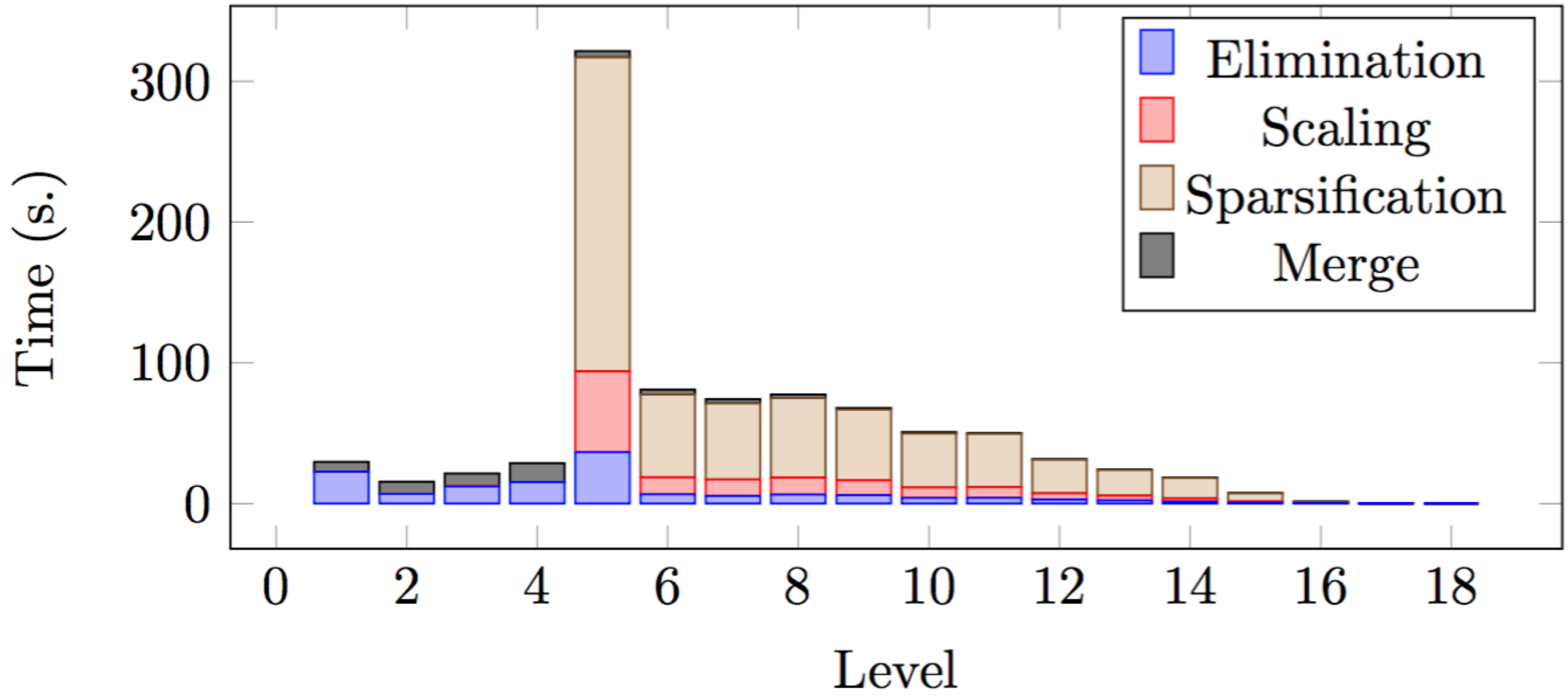




Top separator block would be 32 GB without the sparsification!

Profiling: main cost is RRQR

Optimum is to skip sparsification for levels 1 to 4



Acknowledgements & Funding

- References:
  - K. L. Ho and L. Ying, Hierarchical interpolative factorization for elliptic operators: differential equations, Communications on Pure and Applied Mathematics, 69 (2016), pp. 1415–1451
- Funding:
  - U.S. DOE NNSA under Award Number DE-NA0002373-1, LDRD research grant from Sandia National Laboratories