An Algebraic Sparsified Nested Dissection Algorithm using Low-Rank Approximations

Léopold Cambier*, Chao Chen*, Eric Darve*
Erik Boman†, Sivasankaran Rajamanickam†, Raymond Tuminaro†

* Stanford University (ICME/ME)
† Sandia National Lab

Mathias Conference, Serris, France
October 22, 2018

Sparse Linear Systems

We want to solve

$$Ax = b$$

When A of size $N \times N$ is large, SPD and sparse

Bottleneck in many physical applications

Sparse Linear Systems

Direct methods
(Sparse) LU
+ Pre-Ordering (ND)

- Very Robust
- Very Accurate
- Generic
- (Very) Costly

Incomplete Factorizations
Incomplete LU / Sparsification methods / ...

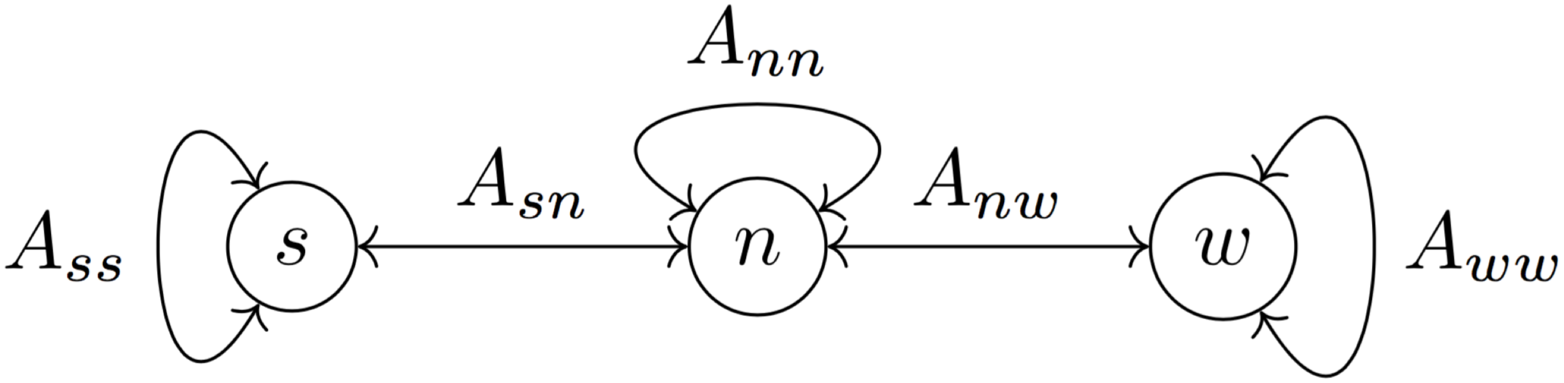
- Tunable Accuracy
- Tunable Cost
- (Fairly) Generic
- Direct method or Preconditioner

Iterative methods
CG / GMRES
+ Custom Preco

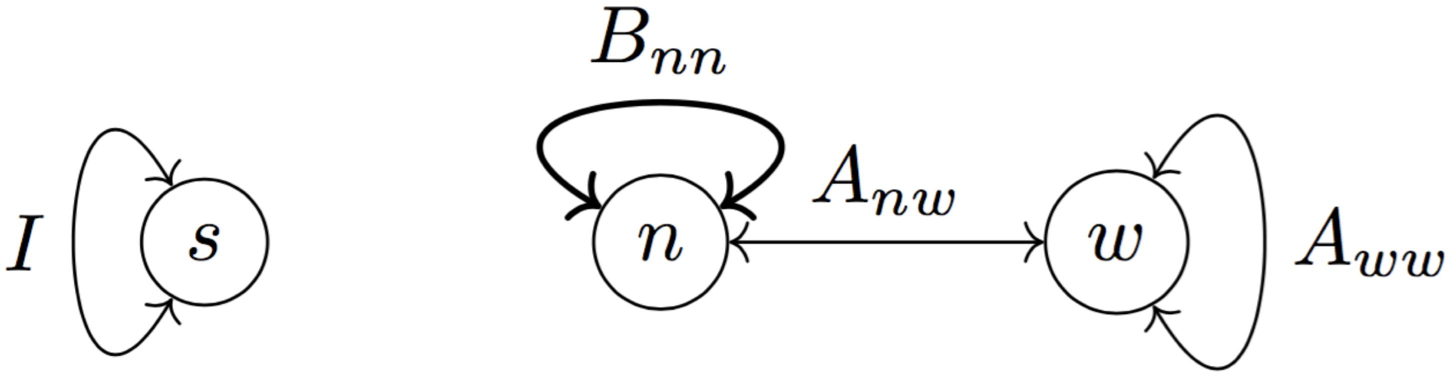
- Cheap
- (Very) Specific
- Complex Convergence
- Need Domain knowledge

Sparse Linear Systems

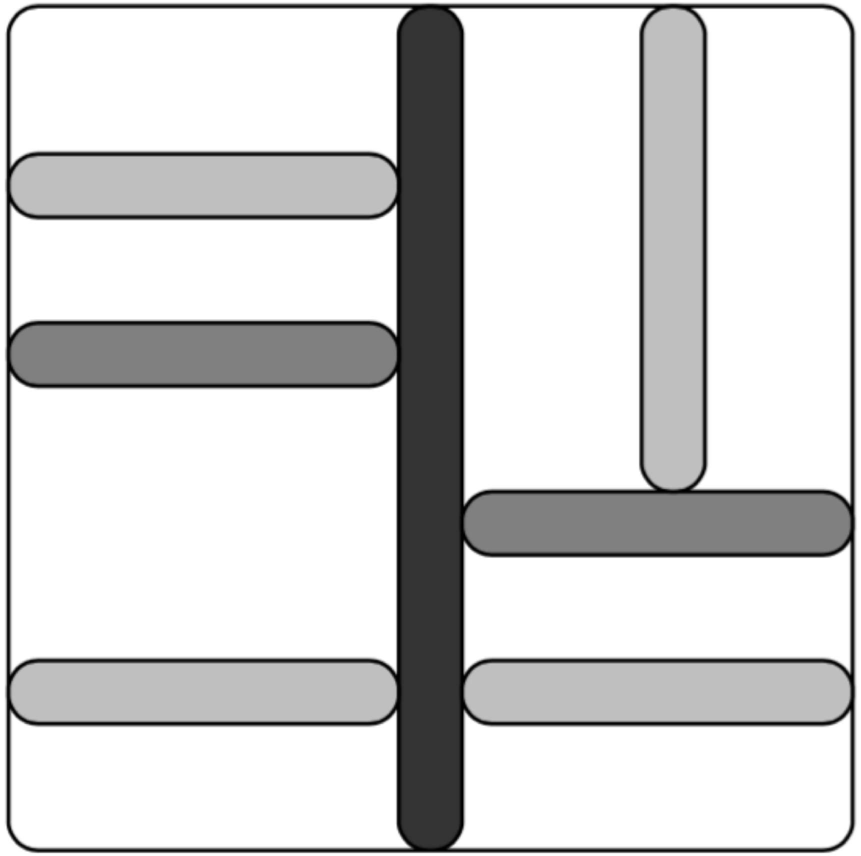
$$\begin{bmatrix} A_{ss} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix}$$

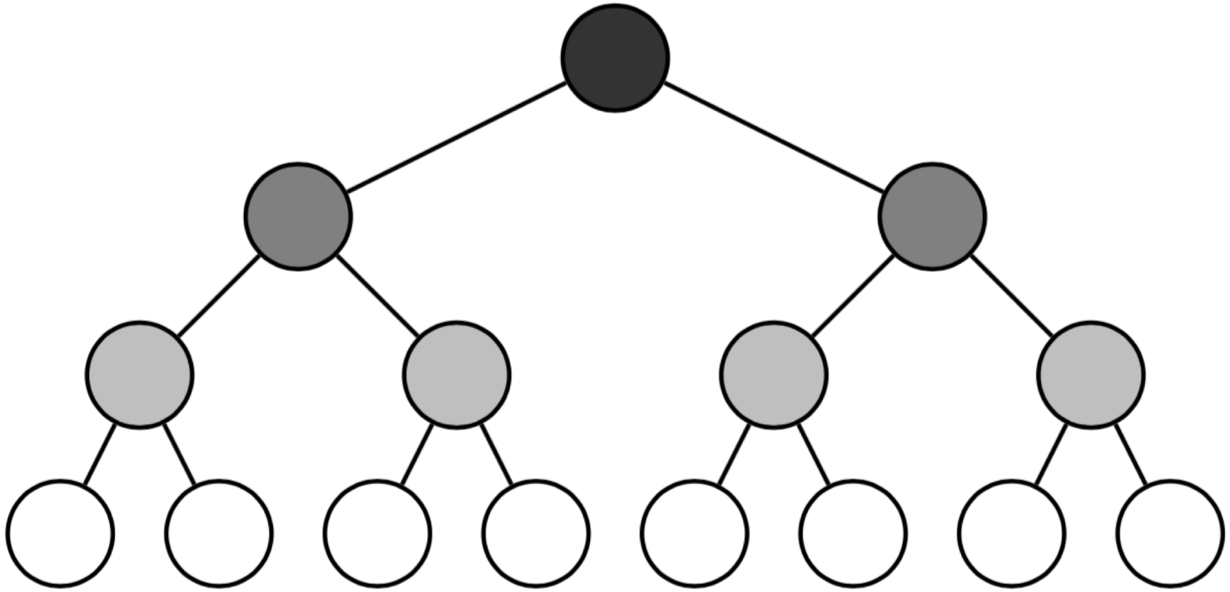


$$\begin{bmatrix} I & & \\ & A_{ww} & A_{wn} \\ & A_{nw} & B_{nn} \end{bmatrix}$$



Nested Dissection



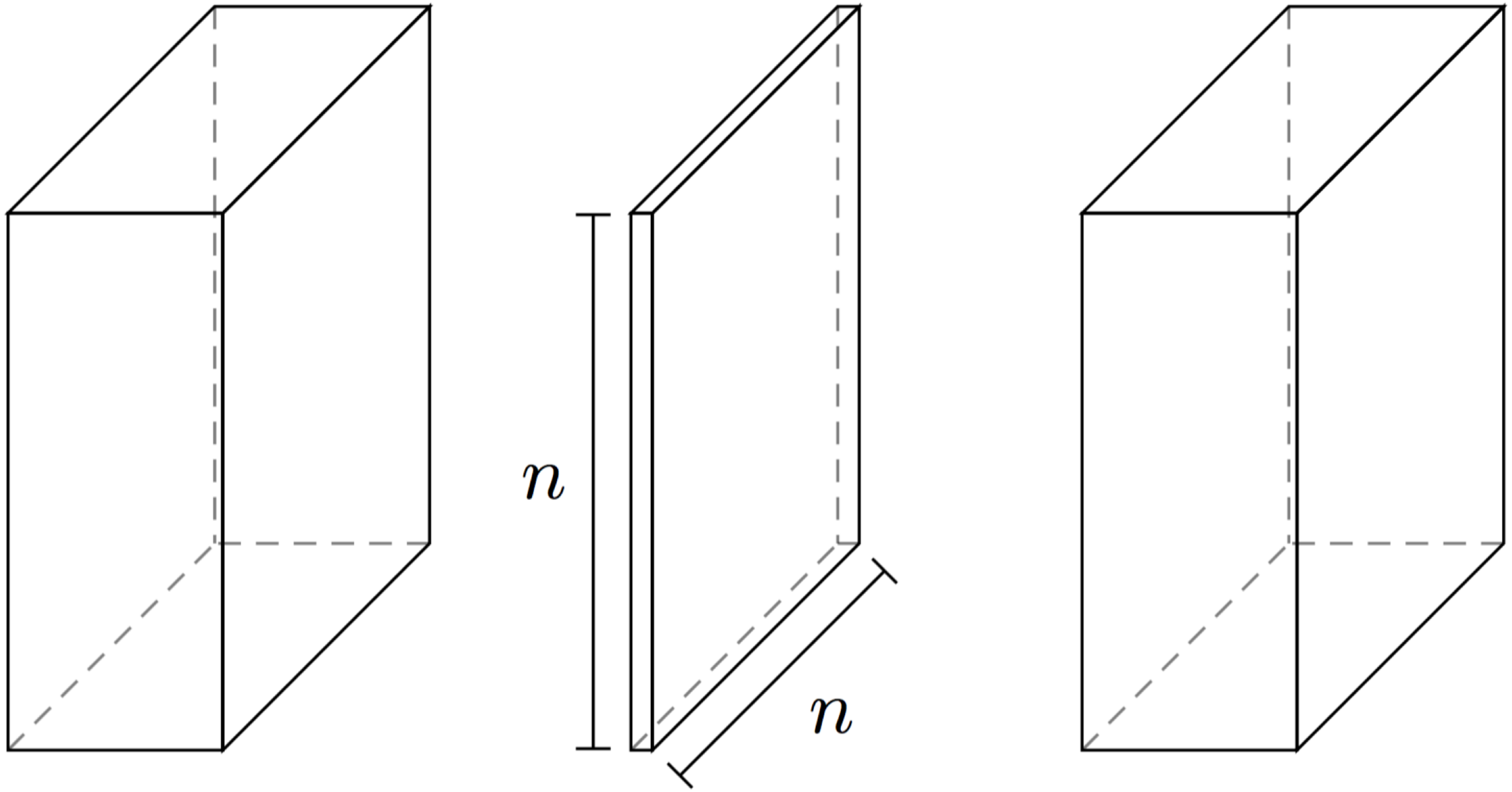


Nested Dissection

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

Separator: n^2

Cost: $n^{2 \cdot 3} = N^2$

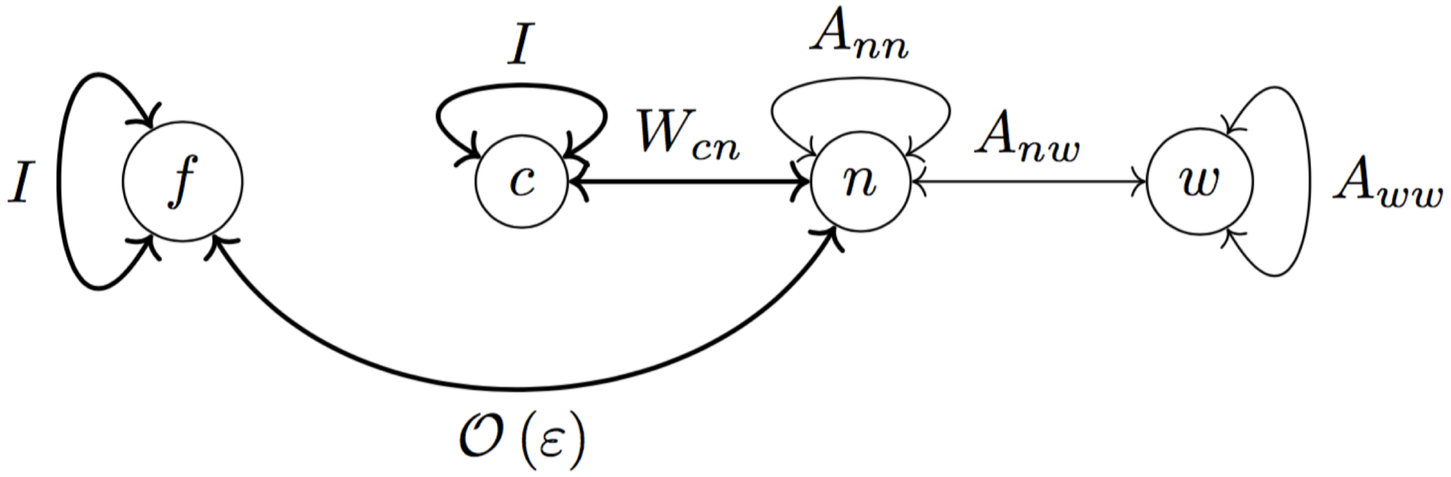


Sparsification

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

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

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

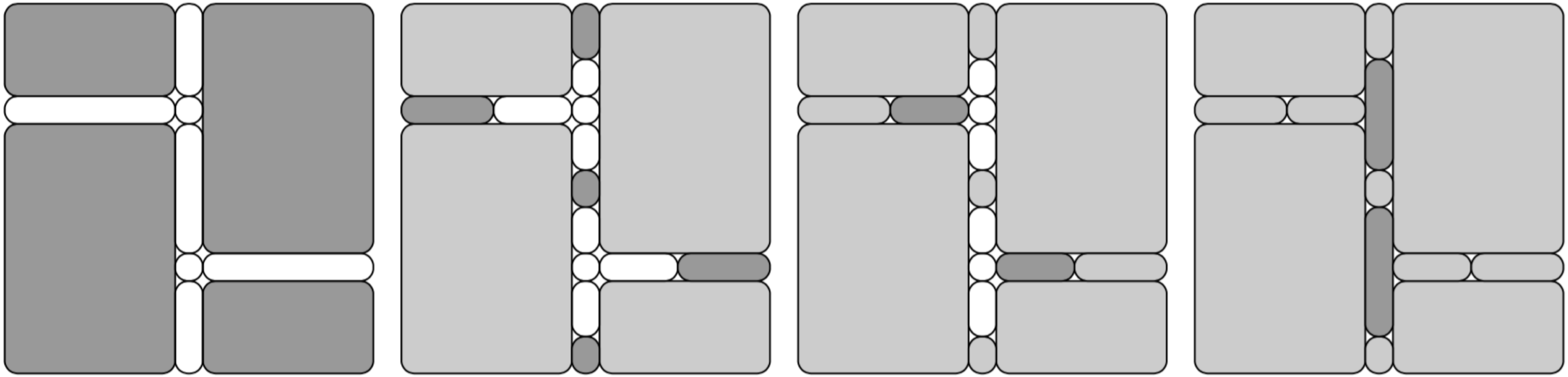


Sparsified Nested Dissection

For level ℓ, from leaves to top

Eliminate interiors at level ℓ

Sparsify (Scale & Compress) interfaces at level ℓ



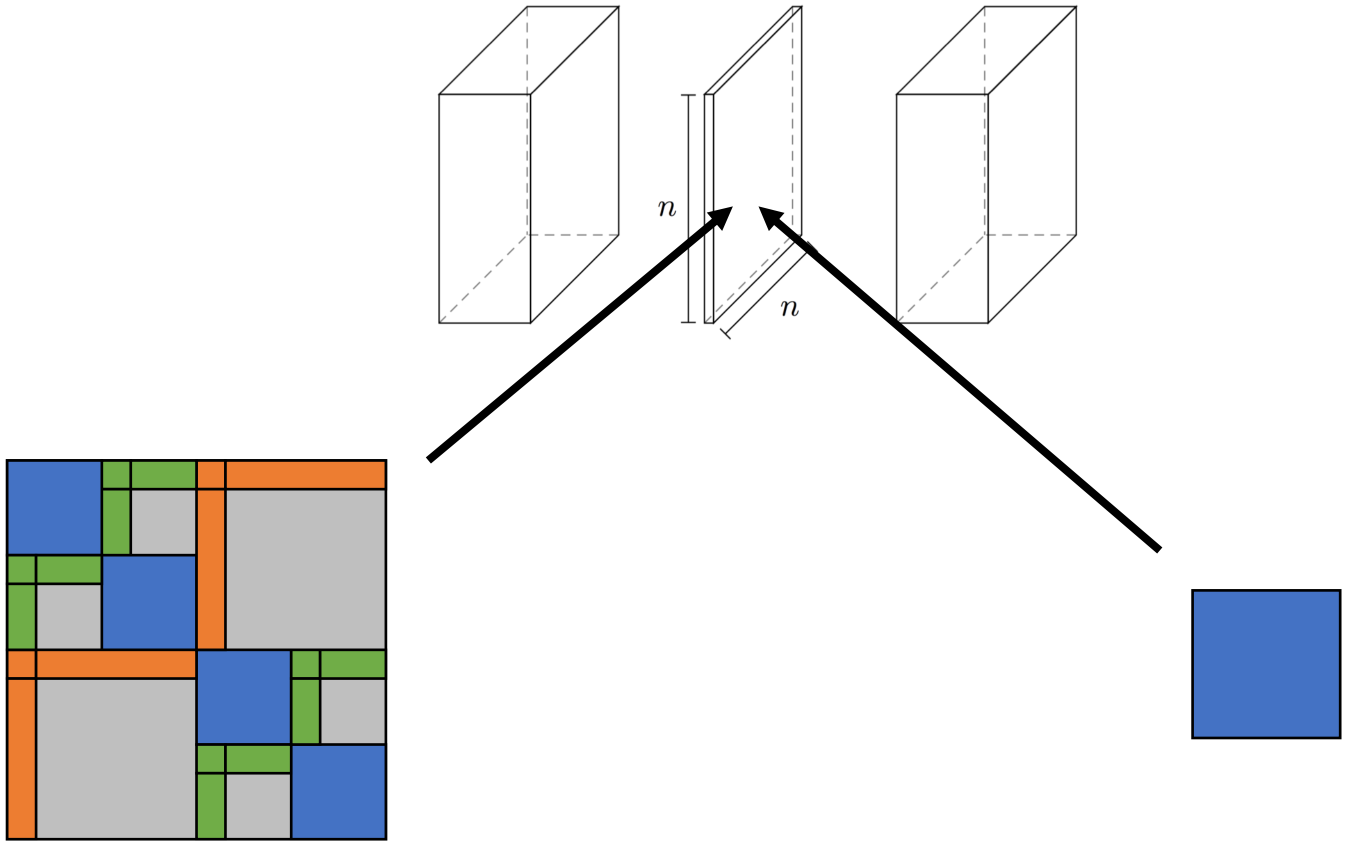
Sparsified Nested Dissection

We effectively build a preconditioner M such that

$$M^\top A M \approx I \qquad\qquad M = \prod_{\ell} \left(\prod_{s} Q_s^\top \prod_{s} E_s \right)$$

We then use M as a preconditioner for CG or GMRES.

Different from fast-algebra techniques



Low-Rank Compression: three variants

$$\begin{bmatrix} \boldsymbol{A_{ss}} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix} \qquad A_{ns} = [A_{nc} \quad A_{nc}T_{cf}] + \varepsilon$$

OR

$$\begin{bmatrix} \boldsymbol{I} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix} \qquad \begin{aligned} A_{ns} &= [A_{nc} \quad A_{nc}T_{cf}] + \varepsilon \\ A_{sn} &= Q_{sc}W_{cn} + \varepsilon \end{aligned}$$

Low-Rank Compression: three variants

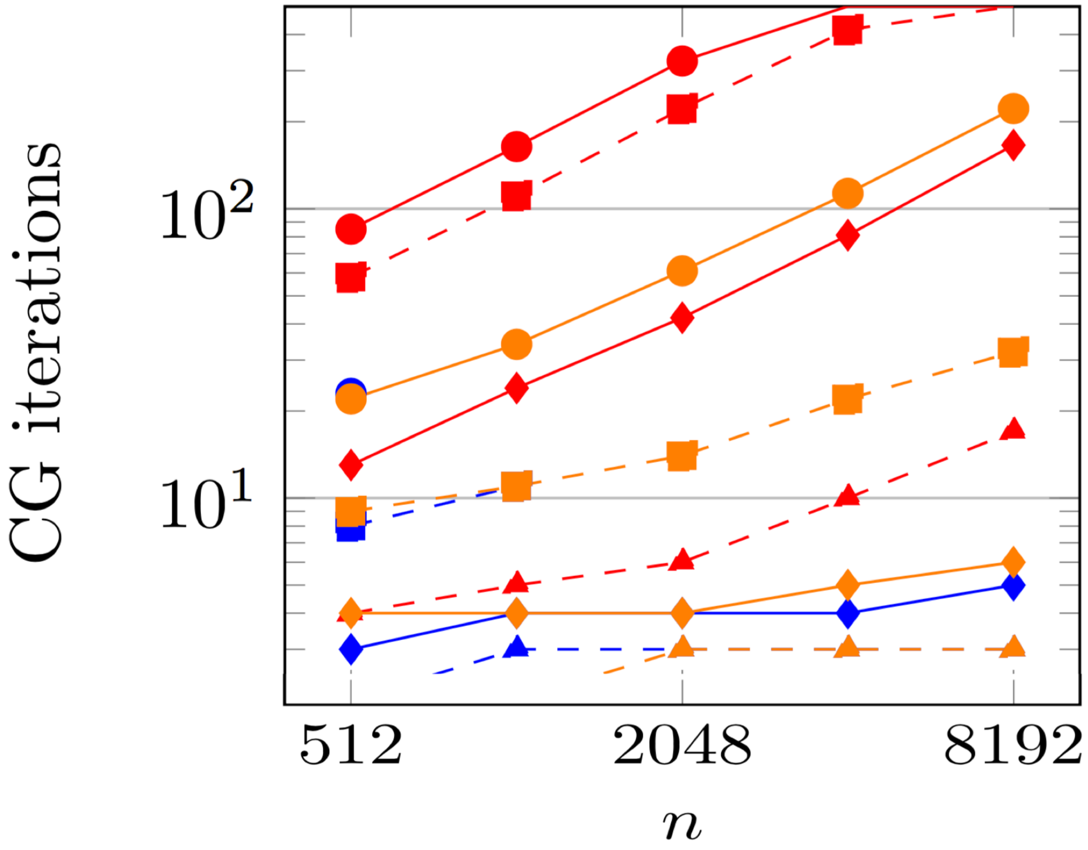
$$\begin{bmatrix} \boldsymbol{I} & & A_{sn} \\ & A_{ww} & A_{wn} \\ A_{ns} & A_{nw} & A_{nn} \end{bmatrix} \qquad A_{sn} = Q_{sc}W_{cn} + \boxed{Q_{sf}W_{fn}}$$

$O(\varepsilon)$

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

Schur Complement over (n,n)

Low-Rank Compression: three variants



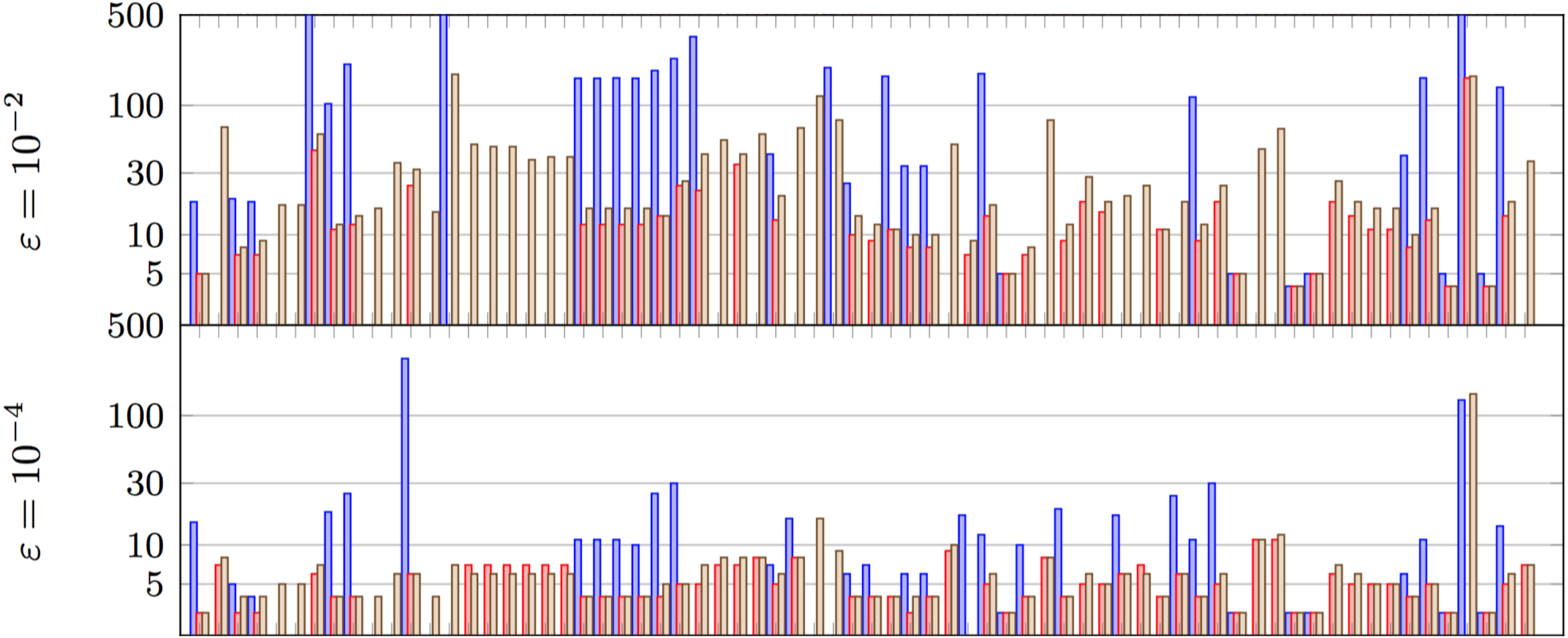
Interpolative, no scaling

Interpolative, with scaling

Orthogonal, with scaling

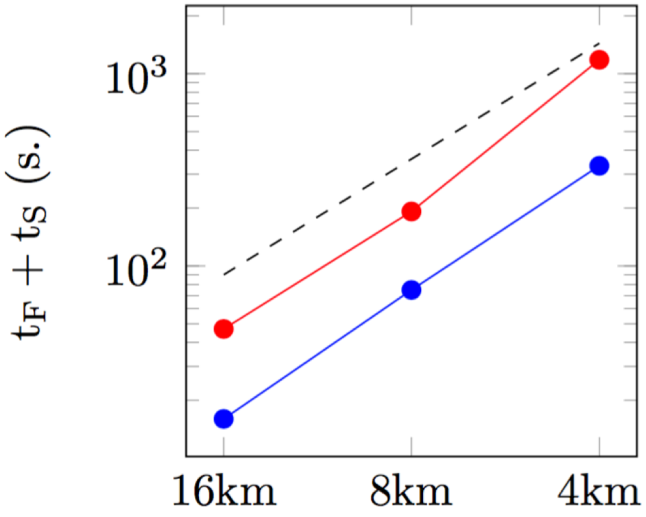
$$\varepsilon = 10^{-6} \rightarrow 10^{-1}$$

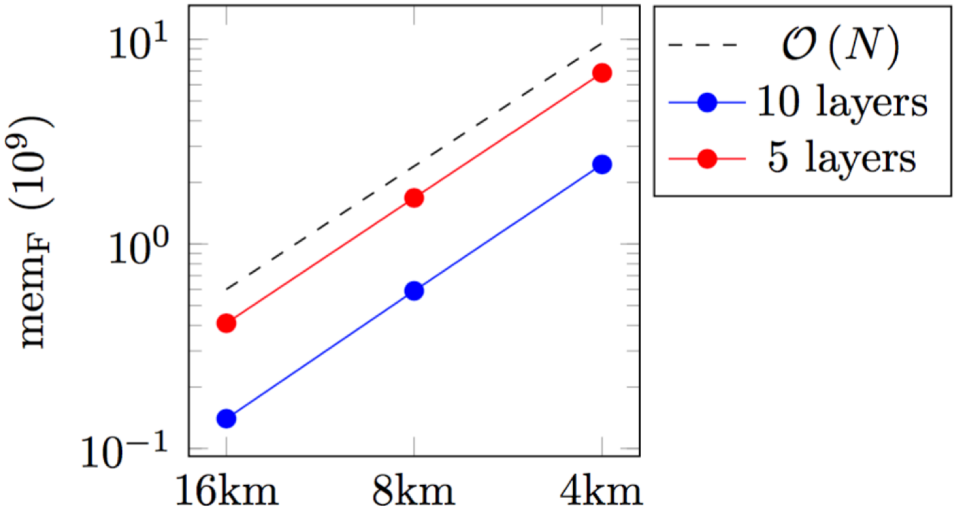
All SPD problems from SuiteSparse

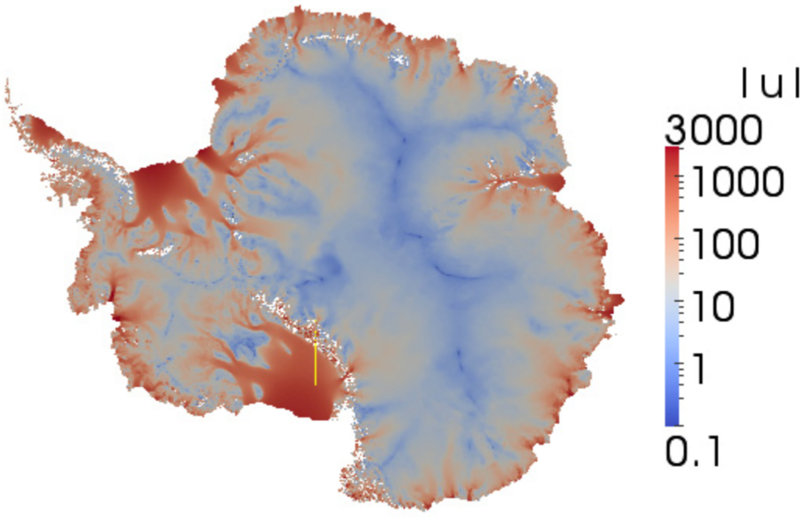


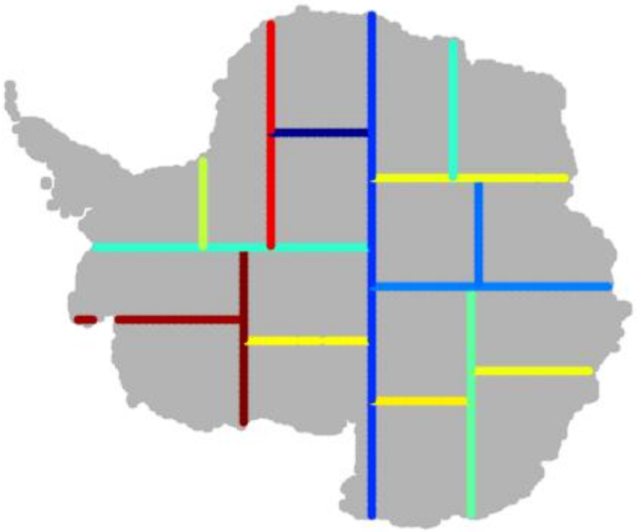
A very ill-conditioned problem

N	spaND t_F (s.)	t_S (s.)	n_{CG}	size_{Top}	mem_F (10^9)	Direct $t_F + t_S$ (s.)	ILU $t_F + t_S$ (s.) (GMRES)
5 layers							
629 544 (16 km)	13	3	7	76	0.14	22	150 (45)
2 521 872 (8 km)	55	20	8	89	0.59	206	801 (76)
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	524 (90)
4 623 432 (8 km)	148	44	8	148	1.68	710	
18 509 480 (4 km)	798	384	10	159	6.86	—	



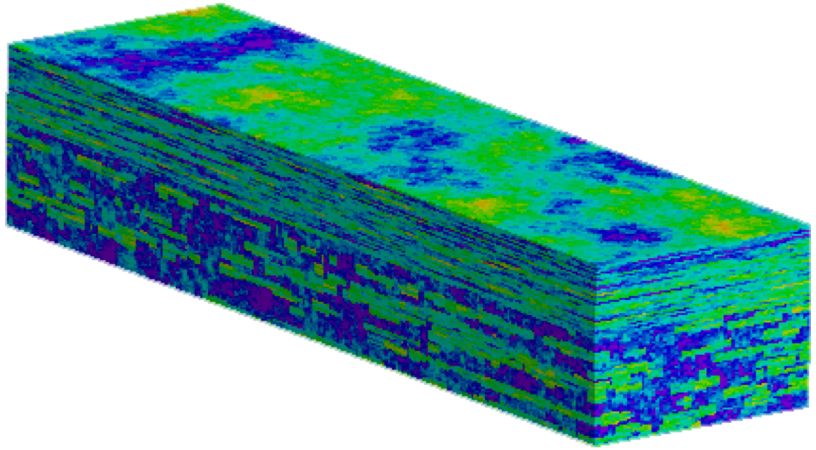


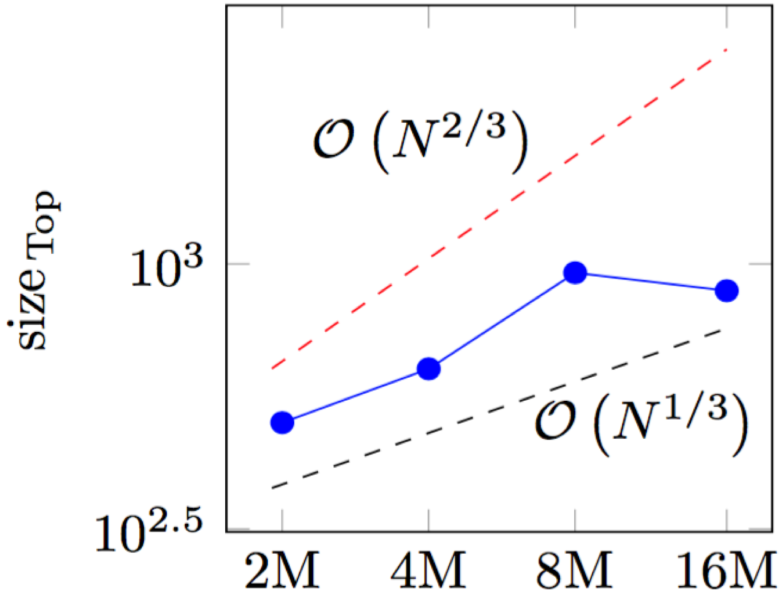


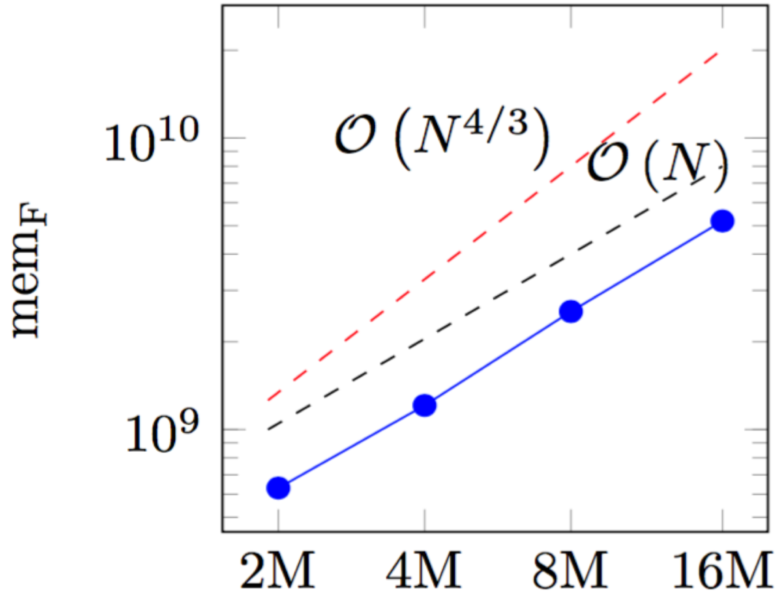


A ’cubic-like’ problem

n	$N = n^3$	spaND t_F (s.)	t_S (s.)	n_{CG}	$size_{Top}$	mem_F (10^9)	Direct. $t_F + t_S$ (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	—

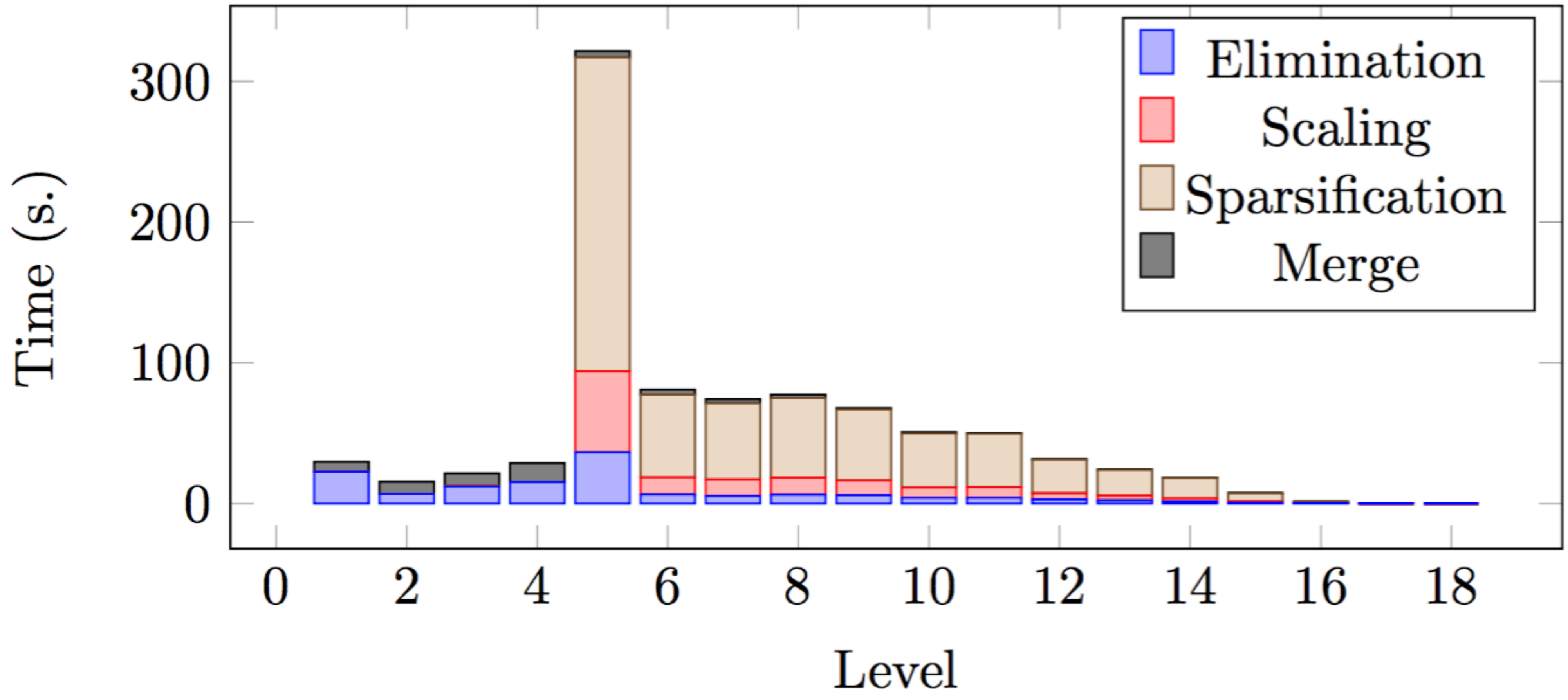






Top separator bloc would be 32 GB otherwise!

Profiling: RRQR takes (lots) of time



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
  - **Total Innovation Fellowship**