Towards the scalable inversion of structured matrices with standard admissibility conditions

Jack Poulson   Yingzhou Li   Lexing Ying

Department of Mathematics
Stanford University

KAUST Seminar
Thuwal, Saudi Arabia
October 28, 2014

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Overview

1. Why do Green's functions from certain second-order elliptic equations have numerically low-rank long-range interactions? *(Nested applications of Poincaré and Caccioppoli inequalities)*
2. Why is general dense $\mathcal{H}$-matrix factorization not effectively parallelizable? *(Long critical path)*
3. Why is Newton-Schulz inversion tempting? *(Each iteration is $\mathcal{H}$-matrix composition, which can be fast and parallel)*
4. Why is a good initial guess for Newton-Schulz needed? *(Otherwise, interior iterates will not be cheaply approximable)*
5. What is a reasonable initial guess that can be computed scalably? *(An inverse of a weakly-admissible $\mathcal{H}^2$-matrix)*
6. Some results on scalable $\mathcal{H}$-matrix composition and a discussion of future work

Outline

Low-rank approximations

Why $\mathcal{H}$-matrix factorization is problematic

The promise of Newton-Schulz inversion

The need for a good initial guess

Parallel weakly-admissible $\mathcal{H}^2$ inversion

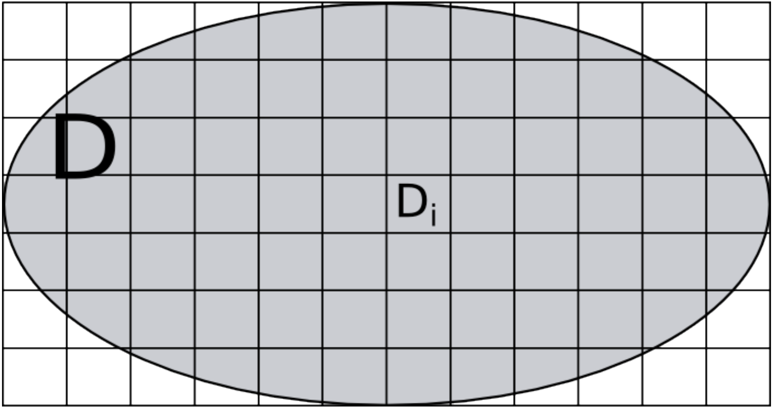
Some distributed $\mathcal{H}$-matrix composition results

Future work

Piecewise constant approximation bound via Poincaré

[Bebendorf/Hackbusch-2002]
For convex $D \subset \mathbb{R}^d$, use Poincaré over convex covering $\{D_i\}_{i=0}^{\ell^d}$:

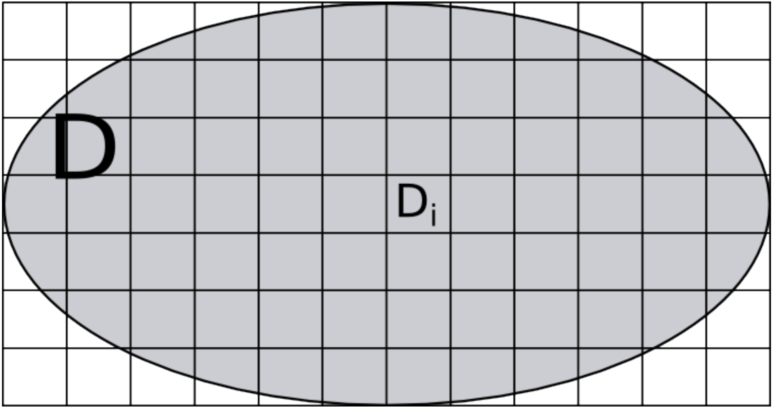
$$\|u - \bar{u}\|_{L^2(D_i)} \leq \frac{\mathrm{diam}(D_i)}{\pi} \|\nabla u\|_{L^2(D_i)}, \quad u \in H^1(D).$$



Piecewise constant approximation bound via Poincaré

[Bebendorf/Hackbusch-2002]
For convex $D \subset \mathbb{R}^d$, use Poincaré over convex covering $\{D_i\}_{i=0}^{\ell^d}$:

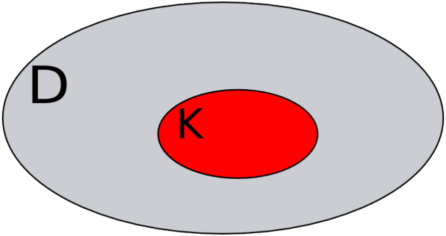
$$\|u - \bar{u}\|_{L^2(D)} \leq \frac{2^{1/d}\mathrm{diam}(D)}{\pi\ell}\|\nabla u\|_{L^2(D)}, \quad u \in H^1(D).$$



Bounding the gradient via Caccioppoli

[Bebendorf/Hackbusch-2002]
Let $u \in H_1(D)$, $D \subset \Omega$, be $C_0^\infty(D)$-weakly L-harmonic with respect to $Lv = -\sum_{i,j=1}^{d} \partial_j (c_{ij} \partial_i v)$, where $C = C(x) = (c_{ij})_{ij}$, $c_{ij} \in L^\infty(\Omega)$, is symmetric with condition number bound κ_C.
Then, for $K \subset D$,

$$\|\nabla u\|_{L^2(K)} \leq \frac{4\sqrt{\kappa_C}}{\mathrm{dist}(K, \partial D)} \|u\|_{L^2(D)}.$$

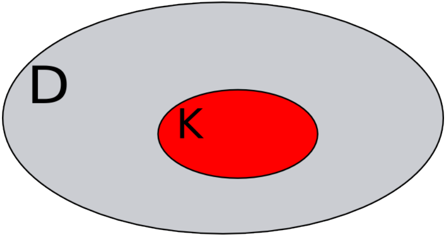


Combined with the previous result, we can bound the piecewise constant approximation of u in K with respect to $\|u\|_{L^2(D)}$.

Bounding the gradient via Caccioppoli

[Bebendorf/Hackbusch-2002]
Let $u \in H_1(D)$, $D \subset \Omega$, be $C_0^\infty(D)$-weakly L-harmonic with respect to $Lv = -\sum_{i,j=1}^{d} \partial_j (c_{ij} \partial_i v)$, where $C = C(x) = (c_{ij})_{ij}$, $c_{ij} \in L^\infty(\Omega)$, is symmetric with condition number bound κ_C.
Then, for $K \subset D$,

$$\|\nabla u\|_{L^2(K)} \leq \frac{4\sqrt{\kappa_C}}{\text{dist}(K, \partial D)} \|u\|_{L^2(D)}.$$



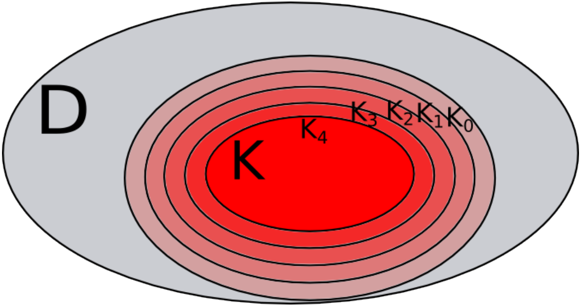
Combined with the previous result, we can bound the piecewise constant approximation of u in K with respect to $\|u\|_{L^2(D)}$.

Exponential convergence via nested Caccioppoli

[Bebendorf/Hackbusch-2002]
Poincaré combined with Caccioppoli of K w.r.t. D yields the ℓ^d-dim. approx:

$$\|u-\bar{u}\|_{L^2(K)} \leq \frac{\gamma\,\mathrm{diam}(K)}{\mathrm{dist}(K,\partial D)\ell}\|u\|_{L^2(D)}, \quad \gamma \equiv \frac{2^{2+1/d}\sqrt{\kappa_C}}{\pi},$$

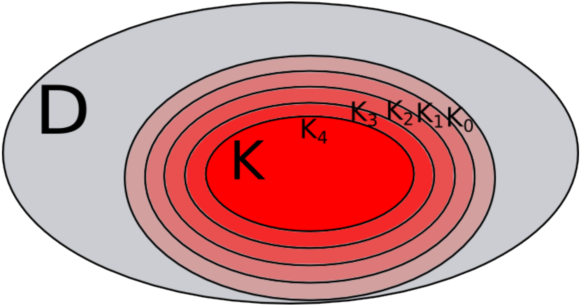
where $u \in H^1(D)$ is weakly L-harmonic. But consider repeated application of inequality over decreasing expansions of K...



Exponential convergence via nested Caccioppoli

[Bebendorf/Hackbusch-2002]
Let $r_j = (1 - j/i)\text{dist}(K, \partial D)$ and $K_j = \{x \in D : \text{dist}(x, K) \leq r_j\}$.
Setting $\rho \equiv \text{dist}(K, \partial D)/\text{diam}(K)$, $u_0 \equiv u$, and
$u_{j+1} \equiv u_j|_{K_{j+1}} - \bar{u}_j$, then we have the $(j+1)\ell^d$-dim. approx. error:

$$\|u_{j+1}\|_{L^2(K_{j+1})} \leq \frac{i\gamma(1+2\rho)}{\rho\ell}\|u_j\|_{L^2(K_j)} \leq \left(\frac{i\gamma(1+2\rho)}{\rho\ell}\right)^j \|u\|_{L^2(K_0)}$$



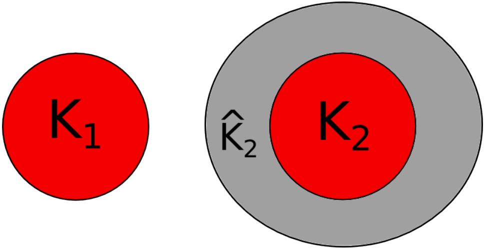
Approximating the Green's function

[Bebendorf/Hackbusch-2002]
Let $K_1, K_2 \subset \Omega$ such that $\mathrm{dist}(K_1, K_2) \geq \rho\, \mathrm{diam}(K_2) > 0$. Then, for any $\epsilon \in (0, 1)$, there is an approximation

$$G_k(x, y) = \sum_{i=1}^{k} u_i(x) v_i(y) \text{ with } k = O(\log(\frac{1}{\epsilon})^{d+1}), \text{ and}$$

$$\|G(x, \cdot) - G_k(x, \cdot)\|_{L^2(K_2)} \leq \epsilon \|G(x, \cdot)\|_{L^2(\hat{K}_2)}, \text{ for all } x \in K_1,$$

where $\hat{K}_2 \equiv \{y \in \Omega : \mathrm{dist}(y, K_2) \leq \frac{\rho}{2} \mathrm{diam}(K_2)\}$.



Outline

Low-rank approximations

Why $\mathcal{H}$-matrix factorization is problematic

The promise of Newton-Schulz inversion

The need for a good initial guess

Parallel weakly-admissible $\mathcal{H}^2$ inversion

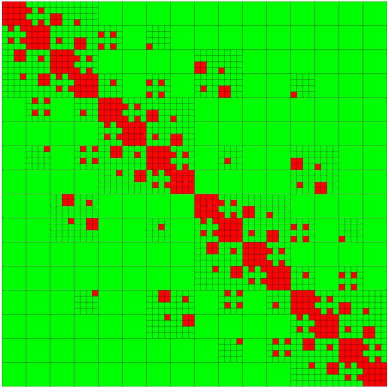
Some distributed $\mathcal{H}$-matrix composition results

Future work

Factoring (dense) $\mathcal{H}$-matrices

Four-level 2D $\mathcal{H}$-matrix with standard admissibility:

$$\min(\mathrm{diam}(K_1), \mathrm{diam}(K_2)) \leq \rho\, \mathrm{dist}(K_1, K_2)$$



- Critical path generally $\Omega(N)$; (approximate) factorization requires $O(N \lg^2 N)$ work, so $O(\lg^2 N)$ parallel speedup...
- Multifrontal techniques boil sparse factorization down to dense interface problems [Grasedyck et al.-2009,Xia et al.-2009]

Outline

Newton-Schulz inversion

Applying Newton's method to

$$f(X) = X^{-1} - A$$

yields the iteration

$$X_{k+1} := (2I - X_k A) X_k$$

Suggested for parallel $\mathcal{H}$-matrix inversion by [Kriemann-2004] as alternative to $\mathcal{H}$-matrix factorization.

Ideally $O(\lg(\kappa(A)))$ approximate $\mathcal{H}$-matrix compositions are required, each involving $O(N \lg^2 N)$ operations

$\mathcal{H}$-matrix composition is highly parallelizable, and so, if the $\mathcal{H}$-matrix approximations are all valid, the problem is solved

Newton-Schulz inversion

Applying Newton's method to

$$f(X) = X^{-1} - A$$

yields the iteration

$$X_{k+1} := (2I - X_k A)X_k$$

Suggested for parallel $\mathcal{H}$-matrix inversion by [Kriemann-2004] as alternative to $\mathcal{H}$-matrix factorization.

Ideally $O(\lg(\kappa(A)))$ approximate $\mathcal{H}$-matrix compositions are required, each involving $O(N \lg^2 N)$ operations

$\mathcal{H}$-matrix composition is highly parallelizable, and so, if the $\mathcal{H}$-matrix approximations are all valid, the problem is solved

Newton-Schulz inversion

Applying Newton's method to

$$f(X) = X^{-1} - A$$

yields the iteration

$$X_{k+1} := (2I - X_k A) X_k$$

Suggested for parallel $\mathcal{H}$-matrix inversion by [Kriemann-2004] as alternative to $\mathcal{H}$-matrix factorization.

Ideally $O(\lg(\kappa(A)))$ approximate $\mathcal{H}$-matrix compositions are required, each involving $O(N \lg^2 N)$ operations

$\mathcal{H}$-matrix composition is highly parallelizable, and so, if the $\mathcal{H}$-matrix approximations are all valid, the problem is solved

Newton-Schulz inversion

Applying Newton's method to

$$f(X) = X^{-1} - A$$

yields the iteration

$$X_{k+1} := (2I - X_k A) X_k$$

Suggested for parallel $\mathcal{H}$-matrix inversion by [Kriemann-2004] as alternative to $\mathcal{H}$-matrix factorization.

Ideally $O(\lg(\kappa(A)))$ approximate $\mathcal{H}$-matrix compositions are required, each involving $O(N \lg^2 N)$ operations

$\mathcal{H}$-matrix composition is highly parallelizable, and so, if the $\mathcal{H}$-matrix approximations are all valid, the problem is solved

Outline

Low-rank approximations

Why $\mathcal{H}$-matrix factorization is problematic

The promise of Newton-Schulz inversion

The need for a good initial guess

Parallel weakly-admissible $\mathcal{H}^2$ inversion

Some distributed $\mathcal{H}$-matrix composition results

Future work

Truncated Newton-Schulz

For large classes of equations (see, e.g., [Hackbusch/Bebendorf-2003]), A and A^{-1} are known to be representable as $\mathcal{H}$-matrices with low ranks.

But what about the intermediate iterates?

$$X_{k+1} := (2I - X_k A) X_k$$

Approximate matrix iterations studied in [Hackbusch/Khoromskij/Tyrtyshnikov-2007]

Their conclusion: Intermediate iterates typically **NOT** representable, but with a sufficiently good initial guess, $\mathcal{H}$-matrix Newton-Schulz converges.

An unanswered question: How should one efficiently construct the initial guess in parallel?

Truncated Newton-Schulz

For large classes of equations (see, e.g., [Hackbusch/Bebendorf-2003]), A and A^{-1} are known to be representable as $\mathcal{H}$-matrices with low ranks.

But what about the intermediate iterates?

$$X_{k+1} := (2I - X_k A) X_k$$

Approximate matrix iterations studied in [Hackbusch/Khoromskij/Tyrtyshnikov-2007]

Their conclusion: Intermediate iterates typically **NOT** representable, but with a sufficiently good initial guess, $\mathcal{H}$-matrix Newton-Schulz converges.

An unanswered question: How should one efficiently construct the initial guess in parallel?

Truncated Newton-Schulz

For large classes of equations (see, e.g., [Hackbusch/Bebendorf-2003]), A and A^{-1} are known to be representable as $\mathcal{H}$-matrices with low ranks.

But what about the intermediate iterates?

$$X_{k+1} := (2I - X_k A) X_k$$

Approximate matrix iterations studied in [Hackbusch/Khoromskij/Tyrtyshnikov-2007]

Their conclusion: Intermediate iterates typically **NOT** representable, but with a sufficiently good initial guess, $\mathcal{H}$-matrix Newton-Schulz converges.

An unanswered question: How should one efficiently construct the initial guess in parallel?

Truncated Newton-Schulz

For large classes of equations (see, e.g., [Hackbusch/Bebendorf-2003]), A and A^{-1} are known to be representable as $\mathcal{H}$-matrices with low ranks.

But what about the intermediate iterates?

$$X_{k+1} := (2I - X_k A) X_k$$

Approximate matrix iterations studied in [Hackbusch/Khoromskij/Tyrtyshnikov-2007]

Their conclusion: Intermediate iterates typically **NOT** representable, but with a sufficiently good initial guess, $\mathcal{H}$-matrix Newton-Schulz converges.

An unanswered question: How should one efficiently construct the initial guess in parallel?

Truncated Newton-Schulz

For large classes of equations (see, e.g., [Hackbusch/Bebendorf-2003]), A and A^{-1} are known to be representable as $\mathcal{H}$-matrices with low ranks.

But what about the intermediate iterates?

$$X_{k+1} := (2I - X_k A) X_k$$

Approximate matrix iterations studied in [Hackbusch/Khoromskij/Tyrtyshnikov-2007]

Their conclusion: Intermediate iterates typically **NOT** representable, but with a sufficiently good initial guess, $\mathcal{H}$-matrix Newton-Schulz converges.

An unanswered question: How should one efficiently construct the initial guess in parallel?

Outline

Low-rank approximations

Why $\mathcal{H}$-matrix factorization is problematic

The promise of Newton-Schulz inversion

The need for a good initial guess

Parallel weakly-admissible $\mathcal{H}^2$ inversion

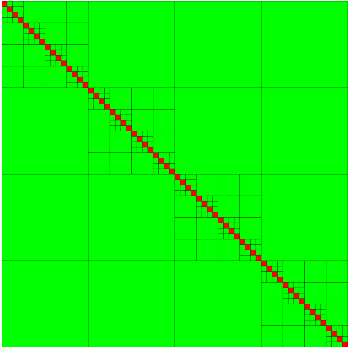
Some distributed $\mathcal{H}$-matrix composition results

Future work

Parallel weakly-admissible $\mathcal{H}^2$-matrix factorization

Weakly-admissible $\mathcal{H}^2$-matrices can be scalably factored in parallel (see, e.g., [Wang et al.-2012]) via ULV decompositions [Chandrasekaran/Gu/Pals-2006].

$K_1 \cap K_2 = \varnothing$



Weak admiss. and shared bases are crucial for the parallelization!

Is the rank needed for initializing a convergent Newton-Schulz sufficiently lower than required for accurate direct inversion?

Outline

Bulk synchronous $\mathcal{H}$-matrix composition

Can vastly generalize three-stage $\mathcal{H}$-matrix/vector multiplication scheme of [Kriemann-2004] to handle composition.[1]

Main idea: Aggressively combine phases of all $\mathcal{H}$-matrix application suboperations, resulting in $O(1)$ communication/computation phases.

[1] Cf. [Izadi-2012]

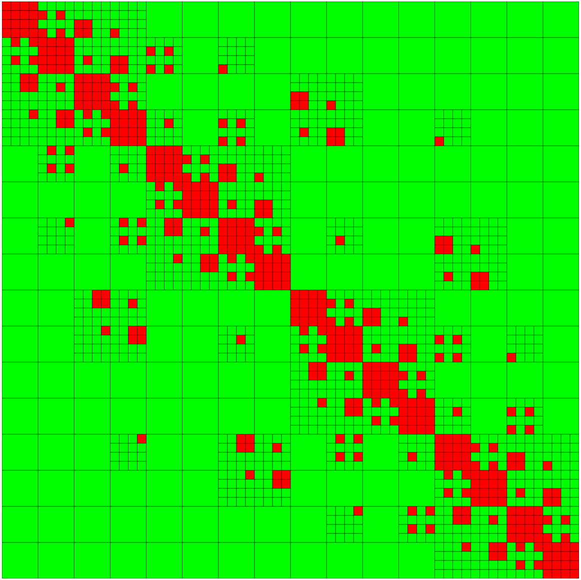
Bulk synchronous $\mathcal{H}$-matrix composition

Can vastly generalize three-stage $\mathcal{H}$-matrix/vector multiplication scheme of [Kriemann-2004] to handle composition.[1]

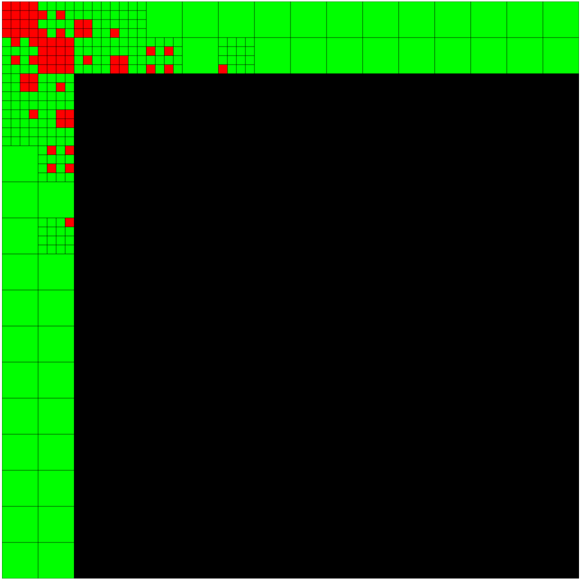
Main idea: Aggressively combine phases of all $\mathcal{H}$-matrix application suboperations, resulting in $O(1)$ communication/computation phases.

[1] Cf. [Izadi-2012]

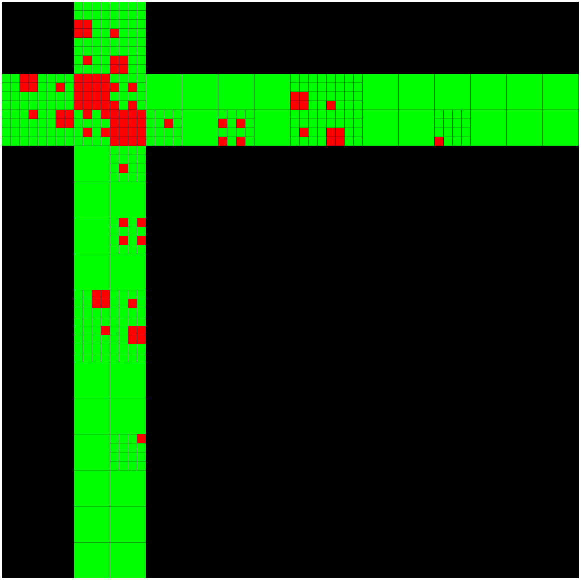
2D standard admissibility



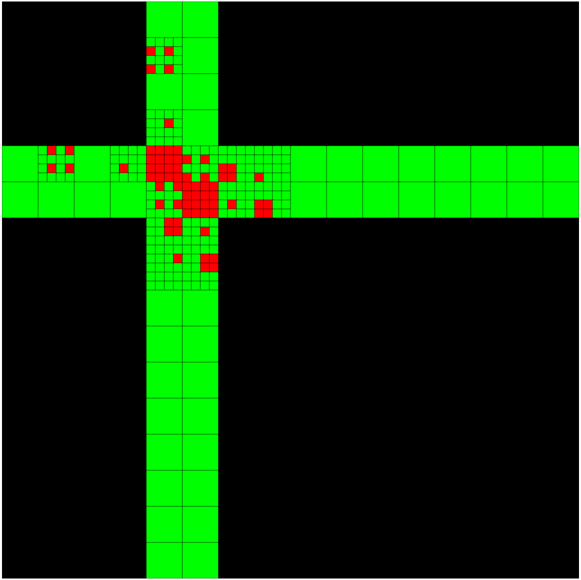
2D standard admissibility (Process 0/8)



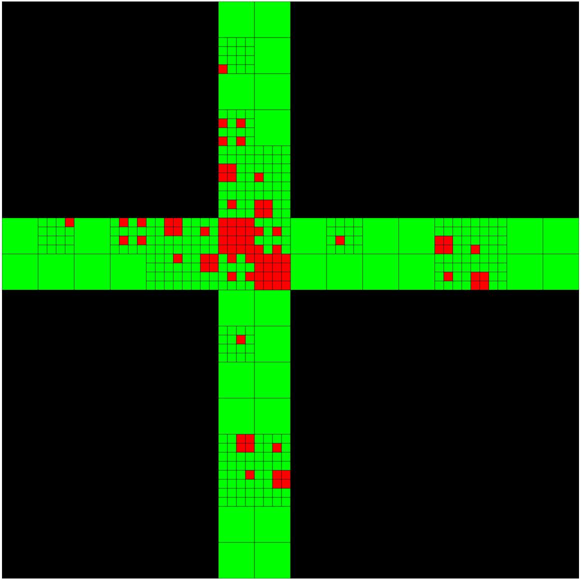
2D standard admissibility (Process 1/8)



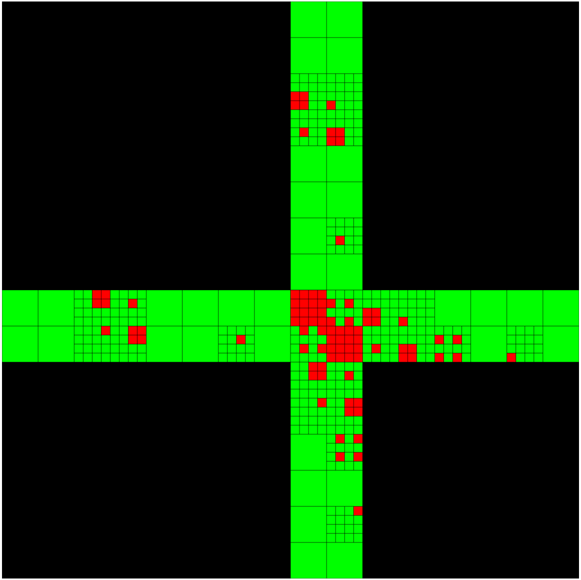
2D standard admissibility (Process 2/8)



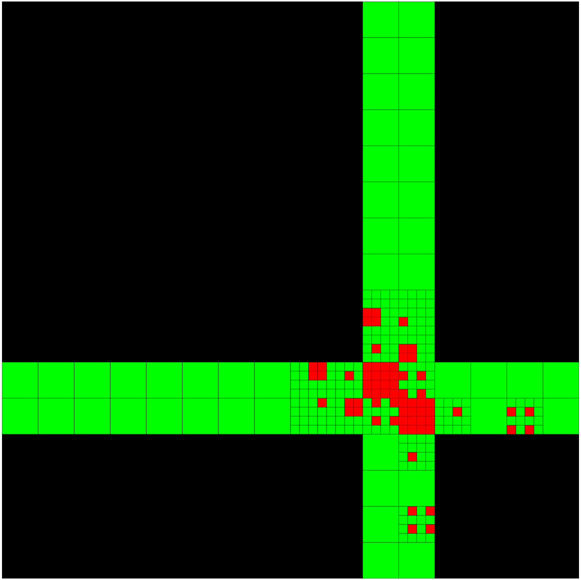
2D standard admissibility (Process 3/8)



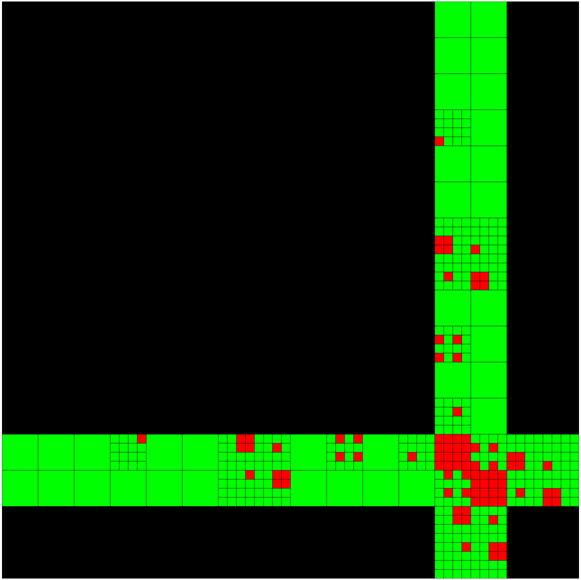
2D standard admissibility (Process 4/8)



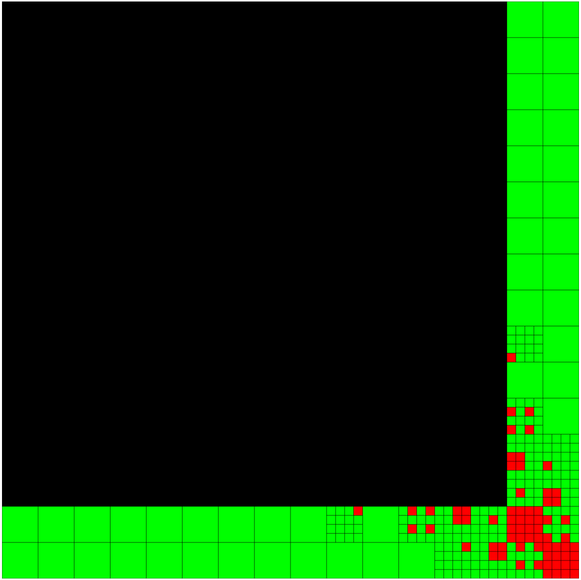
2D standard admissibility (Process 5/8)



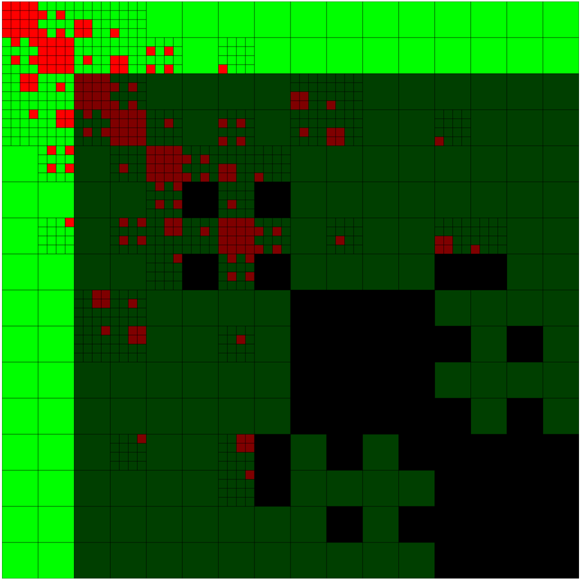
2D standard admissibility (Process 6/8)



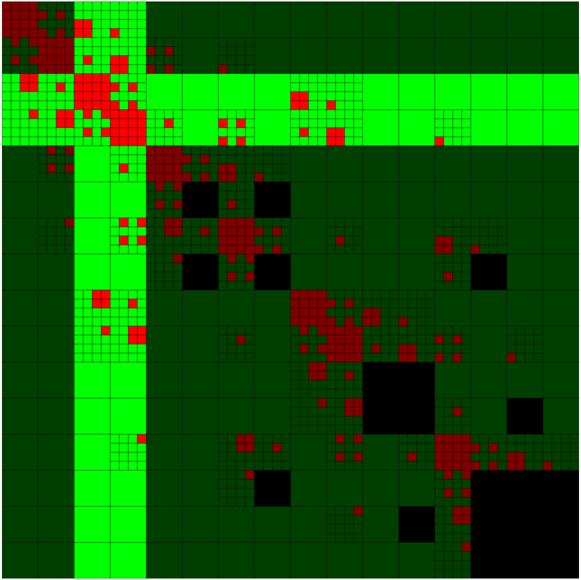
2D standard admissibility (Process 7/8)



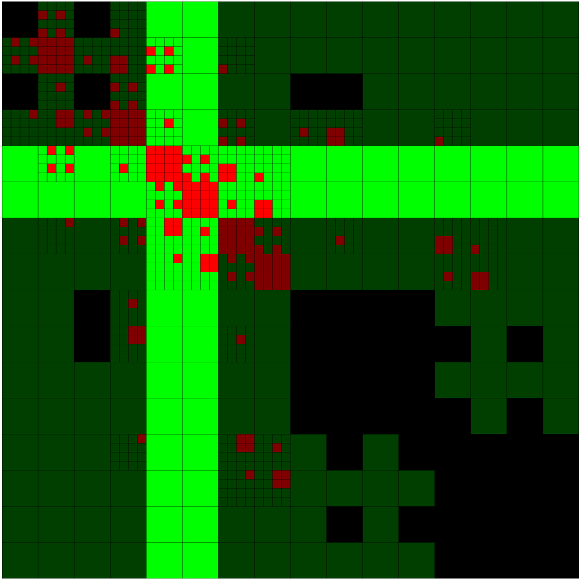
Ghosted 2D standard admissibility (Process 0/8)



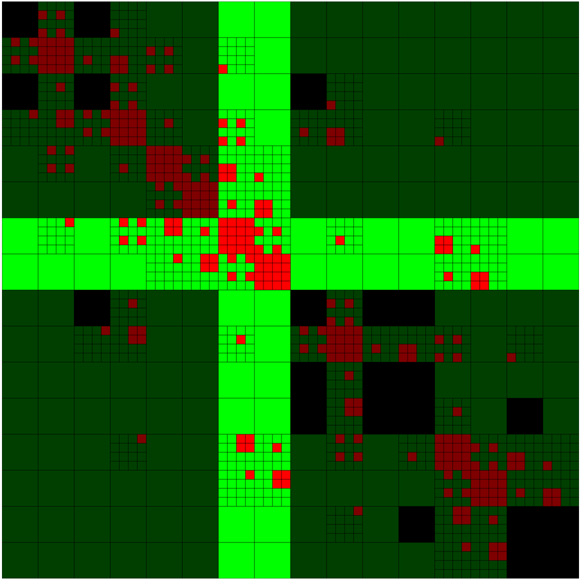
Ghosted 2D standard admissibility (Process 1/8)



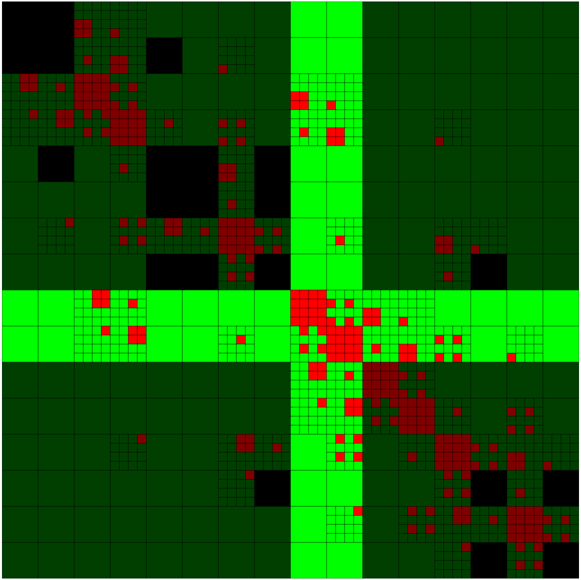
Ghosted 2D standard admissibility (Process 2/8)



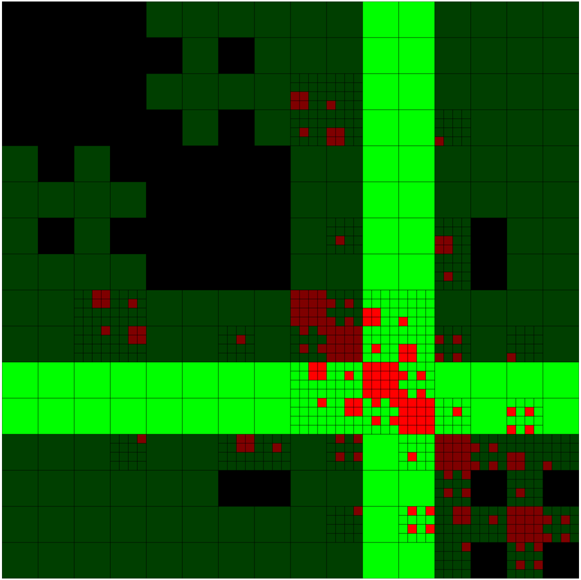
Ghosted 2D standard admissibility (Process 3/8)



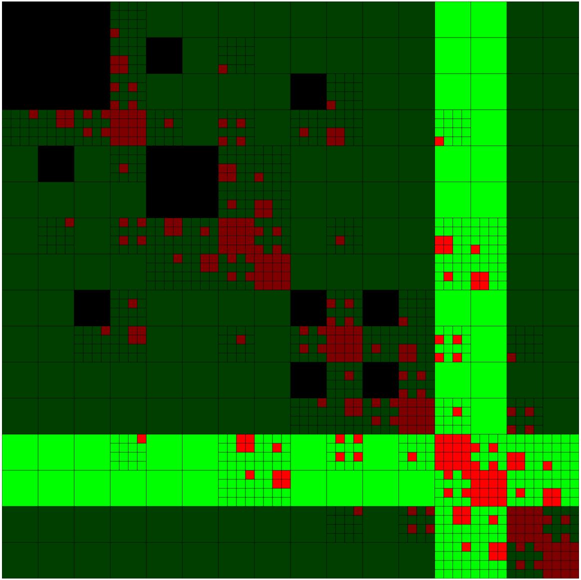
Ghosted 2D standard admissibility (Process 4/8)



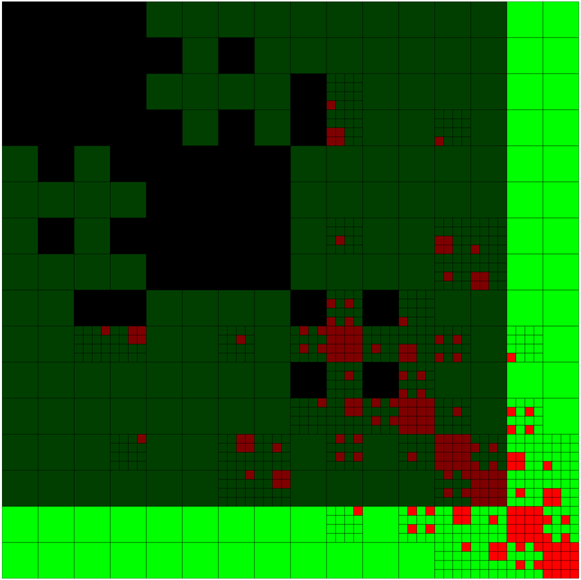
Ghosted 2D standard admissibility (Process 5/8)



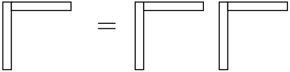
Ghosted 2D standard admissibility (Process 6/8)



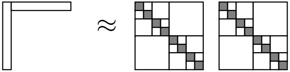
Ghosted 2D standard admissibility (Process 7/8)



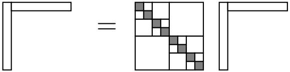
Communication phases for low-rank accumulation



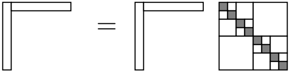
(a) Type FFF



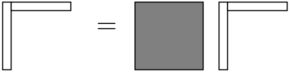
(b) Type FHH



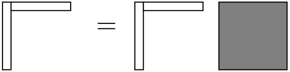
(c) Type FHF



(d) Type FFH



(e) Type FDF



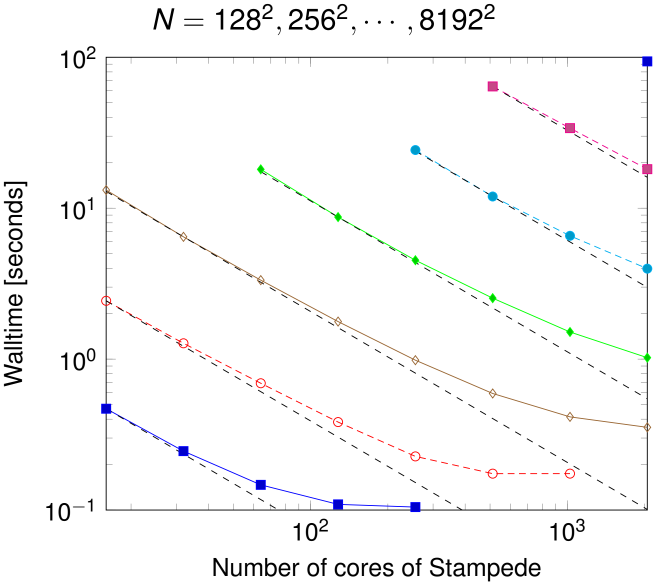
(f) Type FFD

Type FFF	Reduce $\rightarrow$ Pass $\rightarrow$ Broadcast
Type FHH	Reduce $\rightarrow$ Pass $\rightarrow$ Broadcast $\rightarrow$ Reduce $\rightarrow$ Pass $\rightarrow$ Broadcast
Type FHF	Reduce $\rightarrow$ Pass $\rightarrow$ Broadcast
Type FFH	Reduce $\rightarrow$ Pass $\rightarrow$ Broadcast
Type FDF	Pass
Type FFD	Pass

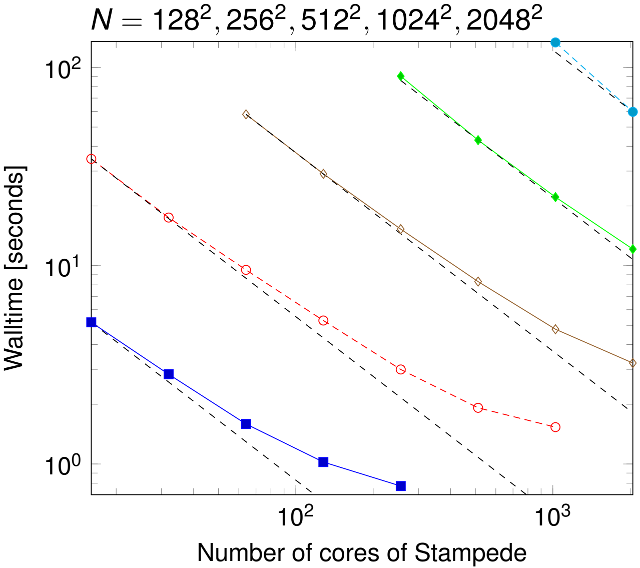
Cost analysis

	Method	γ	α	β
Full	Householder QR	$O(r^2 \frac{N}{p} \lg^3 N)$	$O(r \lg^2 p)$	$O(r^2 \lg^4 p + r^2 \frac{N}{p})$
	TSQR	$O(r^3 \lg^5 p + r^2 \frac{N}{p} \lg^3 N)$	$O(\lg p)$	$O(r^2 \lg^4 p + r^2 \frac{N}{p})$
	CholeskyQR (SVD)	$O(r^3 \lg^4 p + r^2 \frac{N}{p} \lg^3 N)$	$O(\lg p)$	$O(r^2 \lg^4 p + r^2 \frac{N}{p})$
k levels	Householder QR	$O(kr^2 \frac{N}{p} \lg^2 N)$	$O(r \lg^2 p)$	$O(kr^2 \lg^3 p + r^2 \frac{N}{p})$
	TSQR	$O(k^2 r^3 \lg^3 p + kr^2 \frac{N}{p} \lg^2 N)$	$O(\frac{\lg^2 p}{k})$	$O(kr^2 \lg^3 p + r^2 \frac{N}{p})$
	CholeskyQR (SVD)	$O(kr^2 \lg^3 p + kr^2 \frac{N}{p} \lg^2 N)$	$O(\frac{\lg^2 p}{k})$	$O(kr^2 \lg^3 p + r^2 \frac{N}{p})$
Single level	Householder QR	$O(r^2 \frac{N}{p} \lg^2 N)$	$O(r \lg^2 p)$	$O(r^2 \lg^3 p + r^2 \frac{N}{p})$
	TSQR	$O(r^3 \lg^3 p + r^2 \frac{N}{p} \lg^2 N)$	$O(\lg^2 p)$	$O(r^2 \lg^3 p + r^2 \frac{N}{p})$
	CholeskyQR (SVD)	$O(r^2 \lg^3 p + r^2 \frac{N}{p} \lg^2 N)$	$O(\lg^2 p)$	$O(r^2 \lg^3 p + r^2 \frac{N}{p})$

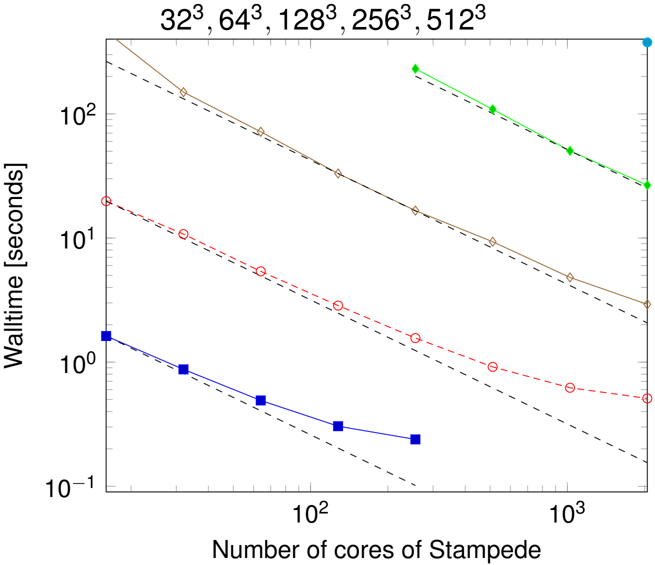
2D composition with weak admissibility ($r = 8$)



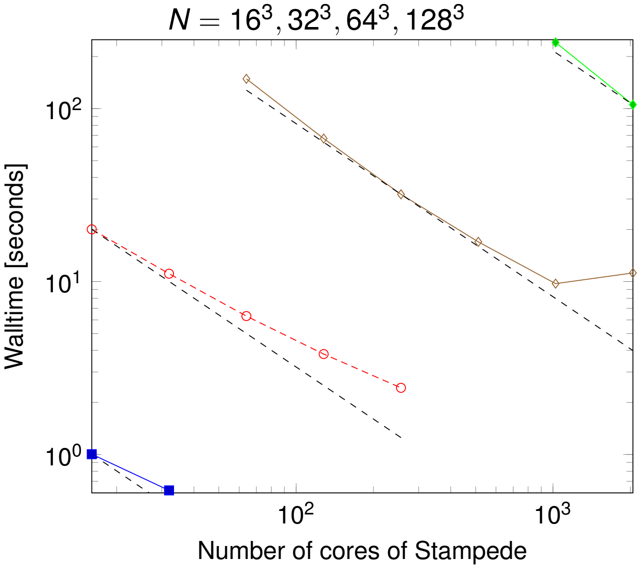
2D composition with standard admissibility ($r = 8$)



3D composition with weak admissibility ($r = 8$)



3D composition with edge admissibility ($r = 8$)



Outline

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Future directions

- Efficient parallel conversion of a weakly-admissible $\mathcal{H}^2$ ULV factorization to $\mathcal{H}$-matrix form [Lin/Lu/Ying-2009,Martinsson-2011]
- Probing for the minimum viable $\mathcal{H}^2$ rank
- Large-scale dense inversion tests
- Extend the above to structured multifrontal method with standard admissibility
- Improving data locality for $\mathcal{H}$-matrix application
- Support for more general topologies
- Hierarchical Interpolative Factorizations [Ho/Ying-2013,cf. Gillman et al.] instead of HSS ULV?

Acknowledgments and Availability

KAUST

Computational resources

TACC

My host

David Keyes and the ECRC

Availability

Prototype implementations available at

`bitbucket.org/poulson/dmhm`

Questions?

Memory usage

Update Method	Sequential Memory	Parallel Memory
Full	$O(rN\lg^2 N)$	$O(r^2\lg^3 p + r^2\frac{N}{p}\lg^3 N)$
k levels	$O(krN\lg N)$	$O(k^2r^2\lg p + k^2r^2\frac{N}{p}\lg N)$
Single level	$O(rN\lg N)$	$O(r^2\lg p + r^2\frac{N}{p}\lg N)$