

# A $N$ -Body Solver for Square Root Iteration

Matt Challacombe,\* Terry Haut,<sup>†</sup> and Nicolas Bock<sup>‡</sup>

*Theoretical Division, Los Alamos National Laboratory*

We develop the Sparse Approximate Matrix Multiply (**SpAMM**)  $n$ -body solver for first order Newton-Schulz iteration of the matrix square root and inverse square root. The solver performs an  $n$ -body occlusion-cull, yielding a bounded relative error in the matrix-matrix product and reduced complexity for problems with structured metric decay. This complexity reduction corresponds to the hierarchical resolution of algebraic structures within recursive volume of the product, a consequence of metric locality. For square root iteration, strongly localized sub-volumes are culled about plane-diagonals and along the cube-diagonal, corresponding to resolution of the identity.

The main contributions of this paper are bounds on the **SpAMM** product and demonstration of a new algebraic locality that develops in these sub-volumes with strongly contracting identity iteration. This contraction corresponds to the deflation of sub-volumes onto plane diagonals of the resolvent, and to a stronger bound on the **SpAMM** product.

Also, we carry out a first order Fréchet analyses for single and dual channel instances of the square root iteration, and look at bifurcations due to ill-conditioning and a too-aggressive **SpAMM** approximation. Then, we show that extreme **SpAMM** approximations and strongly contracting identity iteration can be achieved through iterated regularization, and demonstrate the potential for orders of magnitude acceleration with product representation of the inverse factor.

## I. INTRODUCTION

In many areas of application, long range, high value correlations lead to matrix equations with decay properties. By decay, we mean an approximate inverse relationship between matrix elements and an associated distance; this may be a simple inverse relationship between matrix elements and the Cartesian distance between corresponding support functions, or it may involve a non-Euclidean distance, *e.g.* a generalized measure between character strings.

Matrix equations with decay have history and recent development in the statistics and statistical physics literature [1–5]. Also recently, methods for mesh-free interpolation are demonstrating remarkable predictive power through delocalized correlations and corresponding ill-conditioned matrix equations with extreme slow decay [6–9], a problem equivalent to linear dependence in LCAO Gaussian basis problem in quantum chemistry [10–12]. Generally, local support functions are correlated through Lowdin’s symmetric orthogonalization based on the matrix inverse square root [13], yielding representation independent matrix equations. In electronic structure, important long-range correlations manifest in slow decay properties of the gap shifted matrix sign function, as projector of the effective Hamiltonian (Fig. I). Both of these matrix problem with decay, the matrix sign function and the matrix inverse square root, are related by

Higham’s identity:

$$\text{sign} \left( \begin{bmatrix} 0 & s \\ I & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & s^{1/2} \\ s^{-1/2} & 0 \end{bmatrix}. \quad (1)$$

The theory and computation of these matrix functions is given in Higham’s reference [14].

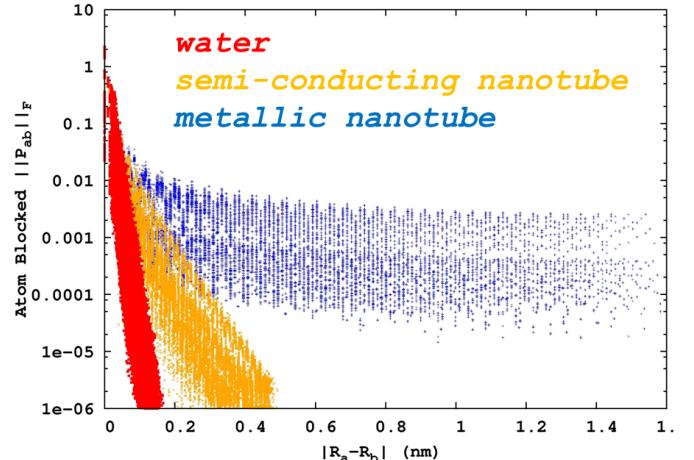


FIG. 1: Examples from electronic structure of decay for the spectral projector (gap shifted sign function) with respect to local (atomic) support. Shown is decay for systems with correlations that are short (insulating water), medium (semi-conducting 4,3 nanotube), and long (metallic 3,3 nanotube) ranged, from exponential (insulating) to algebraic (metallic).

\*Electronic address: matt.challacombe@freeon.org; URL: <http://www.freeon.org>

<sup>†</sup>Electronic address: haut@lanl.gov

<sup>‡</sup>Electronic address: nicolasbock@freeon.org; URL: <http://www.freeon.org>

A well conditioned matrix  $s$  may often correspond to matrix sign and inverse square root functions with rapid exponential decay, and be amenable to *ad hoc* matrix truncation or “sparsification”,  $\bar{s} = s + \epsilon_\tau^s$ , where  $\epsilon_\tau^s$  is the error introduced according to some criterion  $\tau$ . The criterion  $\tau$  might be a drop-tolerance,  $\epsilon_\tau^s = \{-s_{ij} * \hat{e}_i \mid |s_{ij}| <$

$\tau\}$ , a radial cutoff,  $\epsilon_\tau^s = \{-s_{ij} * \hat{\mathbf{e}}_i \mid \|\mathbf{r}_i - \mathbf{r}_j\| > \tau\}$ , or some other approach to truncation, perhaps involving a sparsity pattern chosen *a priori* for computational expedience. Then, the sparse general matrix-matrix multiply (**SpGEMM**) [15–18] may be employed, yielding fast solutions for multiplication rich iterations with fill-in modulated by truncation. Comprehensive surveys of these methods in the numerical linear algebra are given by Benzi [19, 20], and by Bowler [21] and Benzi [22] for electronic structure.

Often however, matrix truncation is ineffective for ill-conditioned problems, because of slow decay, and because of increased numerical sensitivities to poorly controlled (absolute) truncation errors, *e.g.* in the matrix-product:

$$\overline{\mathbf{a} \cdot \mathbf{b}} = \mathbf{a} \cdot \mathbf{b} + \epsilon_\tau^a \cdot \mathbf{b} + \mathbf{a} \cdot \epsilon_\tau^b + \mathcal{O}(\tau^2). \quad (2)$$

An alternative approach is to find a reduced rank approximation closed under the operations of interest. However, compression to a reduced rank may be expensive if the rank is not much smaller than the dimension. Both of these methods, truncation and rank reduction, are focused on matrix data as the target for compression. In this contribution, our target for compression is instead the matrix product itself. For problems with decay, we show that an underlying metric locality, together with a new form of algebraic locality, can lead to complexity reduction under contracting iteration and the  $n$ -body occlusion-cull. The organization of this paper follows.

## Contents

<b>I. Introduction</b>	1
<b>II. SpAMM</b>	2
A. Occlusion-cull	2
B. Bound	3
C. Related research	3
<b>III. First Order Newton-Shulz Iteration</b>	4
A. Sign iteration	4
B. Square root iteration	4
C. Mapping	5
<b>IV. Error Flows in Square Root Iteration</b>	5
A. Stability	5
B. Fréchet derivatives	5
C. Displacements	5
D. Most approximate but still stable	6
<b>V. Regularization</b>	7
A. Product representation	7
B. Effective by one order	7

C. Iterative regularization	7
<b>VI. Locality</b>	8
A. Spatial, metric and temporal locality	8
B. Algebraic locality	9
C. Complexity reduction	9
<b>VII. Conclusions</b>	9
<b>A. Implementation</b>	11
<b>B. Data</b>	11
<b>References</b>	11

## II. SpAMM

In this contribution, we further develop the **SpAMM**  $n$ -body solver for the approximation of matrix functions with decay, based on the quadtree data structure [23–31],

$$\mathbf{a}^i = \begin{bmatrix} \mathbf{a}_{00}^{i+1} & \mathbf{a}_{01}^{i+1} \\ \mathbf{a}_{10}^{i+1} & \mathbf{a}_{11}^{i+1} \end{bmatrix}. \quad (3)$$

For matrices with decay, orderings that preserve data-locality often lead to blocked-by-magnitude or other structures that concentrate data in well segregated neighborhoods; blocks, diagonals, bands & *etc.* Then, matrix elements of like size may be efficiently resolved by the quadtree data structure [23–32].

Increasing data concentration allows to avoid white space in queries and computation, which is compressive for algorithms that can exploit these properties. One common example of such data locality corresponds to an underlying order of particles, elements, support functions & *etc.* with a space filling curve [31–40]

### A. Occlusion-cull

The Sparse Approximate Matrix Multiply (**SpAMM**) carries out occlusion-culling to find only the most important sub-volumes in an approximate matrix product. **SpAMM** has evolved from a row-column oriented skipout mechanism within the BCSR and DBCSR structures [41, 42], to hierarchical approaches based on the quadtree and related to the occlusion-culling found in advanced mechanics and graphics methodologies [43], with occlusion the avoidance of unnecessary tree-work and culling the collection of significant tasks. We again amend the **SpAMM** occlusion-cull, with the recursion

$$\mathbf{a}^i \otimes_{\tau} \mathbf{b}^i = \begin{cases} \emptyset & \text{if } \|\mathbf{a}^i\| \|\mathbf{b}^i\| < \tau \|\mathbf{a}\| \|\mathbf{b}\| \\ \mathbf{a}^i \cdot \mathbf{b}^i & \text{if } (\mathbf{i} = \text{leaf}) \\ \left[ \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{00}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{10}^{i+1}, \quad \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1} \right] & \text{else} \\ \left[ \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1}, \quad \mathbf{a}_{00}^{i+1} \otimes_{\tau} \mathbf{b}_{01}^{i+1} + \mathbf{a}_{01}^{i+1} \otimes_{\tau} \mathbf{b}_{11}^{i+1} \right] & \text{else} \end{cases}, \quad (4)$$

which bounds the relative occlusion error

$$\frac{\|\Delta_{\tau}^{a \cdot b}\|}{n^2} \leq \tau \|\mathbf{a}\| \|\mathbf{b}\|, \quad (5)$$

that occurs in the approximate product

$$\widetilde{\mathbf{a} \cdot \mathbf{b}} \equiv \mathbf{a} \otimes_{\tau} \mathbf{b} = \mathbf{a} \cdot \mathbf{b} + \Delta_{\tau}^{a \cdot b}, \quad (6)$$

where  $\|\cdot\| \equiv \|\cdot\|_F$  is sub-multiplicative [44].

## B. Bound

We now prove (5).

**Proposition 1.** Let  $\tau_{A,B} = \tau \|\mathbf{A}\| \|\mathbf{B}\|$ . Then for each  $i,j$ ,

$$\left| (A \otimes_{\tau} B)_{ij} - (A \cdot B)_{ij} \right| \leq n \tau_{A,B},$$

and

$$\|A \otimes_{\tau} B - A \cdot B\|_F \leq n^2 \tau_{A,B}.$$

*Proof.* We first show the following technical result: it is possible to choose  $\alpha_{lij} \in \{0, 1\}$  such that

$$(A \otimes_{\tau} B)_{ij} = \sum_{l=1}^n A_{il} B_{lj} \alpha_{lij}, \quad (7)$$

In addition, if  $\alpha_{lij} = 0$ , then  $|A_{il}| |B_{lj}| < \tau_{A,B}$ . To show this, we use induction on the number  $k_{\max}$  of levels.

First, if  $k_{\max} = 0$ ,

$$A \otimes_{\tau} B = \begin{cases} 0 & \text{if } \|\mathbf{A}\|_F \|\mathbf{B}\|_F < \tau_{A,B}, \\ A \cdot B & \text{else.} \end{cases}$$

Therefore,  $A \otimes_{\tau} B$  is of the form (7) with either all  $\alpha_{lij} = 0$  or all  $\alpha_{lij} = 1$ . Moreover, if  $\alpha_{lij} = 0$ , then  $|A_{il}| |B_{lj}| \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F < \tau_{A,B}$ .

Now assume that the claim holds for  $k_{\max} - 1$ . We show that it holds for  $k_{\max}$ . Indeed, if  $\|\mathbf{A}\|_F \|\mathbf{B}\|_F < \tau_{A,B}$ , we have that  $A \otimes_{\tau} B = 0$ , which is of the form (7) with all  $\alpha_{lij} = 0$ . Also, if  $\alpha_{lij} = 0$ , then  $|A_{il}| |B_{lj}| < \|\mathbf{A}\|_F \|\mathbf{B}\|_F < \tau_{A,B}$ .

Now assume that  $\|\mathbf{A}\|_F \|\mathbf{B}\|_F \geq \tau_{A,B}$ . Then

$$A \otimes_{\tau} B = \begin{pmatrix} A_{11} \otimes_{\tau} B_{11} + A_{12} \otimes_{\tau} B_{21} & A_{11} \otimes_{\tau} B_{12} + A_{12} \otimes_{\tau} B_{22} \\ A_{21} \otimes_{\tau} B_{11} + A_{22} \otimes_{\tau} B_{21} & A_{21} \otimes_{\tau} B_{12} + A_{22} \otimes_{\tau} B_{22} \end{pmatrix}.$$

We need to consider four cases:  $i \leq n/2$  and  $j \leq n/2$ ,  $i > n/2$  and  $j > n/2$ ,  $i > n/2$  and  $j \leq n/2$ , and, finally,  $i > n/2$  and  $j > n/2$ . Since the analysis is similar for all four cases, we only consider  $i \leq n/2$  and  $j \leq n/2$ . We have that

$$\begin{aligned} (A \otimes_{\tau} B)_{ij} &= (A_{11} \otimes_{\tau} B_{11} + A_{12} \otimes_{\tau} B_{21})_{ij} \\ &= \sum_{l=1}^{n/2} (A_{11})_{il} (B_{11})_{lj} \alpha_{lij}^{(1)} + \\ &\quad \sum_{l=1}^{n/2} (A_{12})_{il} (B_{21})_{lj} \alpha_{lij}^{(2)} \\ &= \sum_{l=1}^n A_{il} B_{lj} \alpha_{lij}, \end{aligned}$$

where we used the induction hypothesis in the second equality.

Now suppose that  $\alpha_{lij} = 0$  for some  $l$ . Then  $\tilde{\alpha}_{lij}^{(1)} = 0$  if  $l \leq n/2$  or  $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$  if  $l > n/2$ . If, e.g.,  $\tilde{\alpha}_{l-n/2,ij}^{(2)} = 0$ , then  $|A_{il}| |B_{lj}| = |(A_{12})_{i,l-n/2}| |(B_{21})_{l-n/2,j}| < \tau_{A,B}$ , where we used the induction hypothesis in the final inequality. The analysis for  $l \leq n/2$  is similar, and the claim follows.

We can now finish the proof of Proposition 1. Indeed, by (7),

$$\begin{aligned} \left| (A \otimes_{\tau} B)_{ij} - (A \cdot B)_{ij} \right| &\leq \sum_{l=1}^n |A_{il} B_{lj}| |\alpha_{lij} - 1| \\ &= \sum_{\alpha_{lij}=0} |A_{il} B_{lj}|. \end{aligned}$$

In addition, if  $\alpha_{lij} = 0$ , then  $|A_{il} B_{lj}| < \tau_{A,B}$  and the lemma follows.  $\square$

## C. Related research

SpAMM is perhaps most closely related to the Strassen-like branch of fast matrix multiplication [45–49], and also

methods for group theoretical embedding allowing fast polynomial multiplication [50–52]. In the Strassen-like approach, disjoint tensor volumes are omitted recursively, while in the SpAMM approach to fast multiplication, the numerically least significant tensor intermediates in the naïve product space are omitted recursively, with error bounded by Eq. (5). Also, these methods for fast matrix multiplication are *stable* [46].

SpAMM is a  $n$ -body method for fast matrix multiplication, related to the generalized methods popularized by Grey [53, 54]. In our development, generalization reflects the *genericity* of recursive data access [29, 55–57], enabling new math supported by range queries, metric queries, higher dimensional queries and so on, but with common frameworks, structures, skeletons, run-times & etc. So far, we have prototyped  $n$ -body solvers for the mainstay problems in modern electronic structure theory, including the Fock exchange [58], semi-local exchange-correlation functionals [59, 60], the Hartree (Coulomb) interaction [61–64], matrix sign function [43, 65] and the matrix inverse square root (this work).

Emergent data frameworks, functional programming languages and other technologies that support generic recursion and skeletonizations [66–70] may enable exploitation of increasingly cheap (decentralized) concurrence by scientific  $n$ -body solvers, and also lower barriers to entry, enhance reliability, sustainability & etc. For centralized, distributed architectures,  $n$ -body methods offer well established protocols for turning spatial and metric locality into data and temporal locality [71–75]. Recently for example, Driscoll *et. al* showed perfect strong scaling and communication optimally for pairwise  $n$ -body methods [76]. Bridging the gap between  $n$ -body solver and fast matrix multiplication, we recently demonstrated strong scaling for fast matrix multiplication (SpAMM) [65].

This work offers a data local alternative to fast non-deterministic methods for sampling the product, which include sketching [77–81], joining [82–88], sensing [89] and probing [90]. These methods may involve probabilistic and on the fly sampling, with the potential for complexity reduction in the case of random distributions or related assumptions. SpAMM also employs on the fly weighted sampling, but with compression through locality, brought about by algebraic correlations (towards identity) and also through structured metric localities.

Finally, this work is inspired broadly by Higham's work, particularly the 2005 paper by Higham, Mackey, Mackey and Tisseur [91] on square root iteration and the group structure of matrix functions. Also, it is influenced by Chen and Chow's [92] approach to scaled NS iteration for ill-conditioned problems, and by the Helgaker group's work on NS iteration, whose notation we follow in part [11].

### III. FIRST ORDER NEWTON-SHULZ ITERATION

There are two common, first order NS iterations; the sign iteration and the square root iteration, related by the square,  $\mathbf{I}(\cdot) = \text{sign}^2(\cdot)$ . These equivalent iterations converge linearly at first, then enter a basin of stability marked by super-linear convergence.

#### A. Sign iteration

For the NS sign iteration, this basin is marked by a behavioral change in the difference  $\delta\mathbf{X}_k = \widetilde{\mathbf{X}}_k - \mathbf{X}_k = \text{sign}(\mathbf{X}_{k-1} + \delta\mathbf{X}_{k-1}) - \text{sign}(\mathbf{X}_{k-1})$ , where  $\delta\mathbf{X}_{k-1}$  is some previous error. The change in behavior is associated with the onset of idempotence and the bounded eigenvalues of  $\text{sign}'(\cdot)$ , leading to stable iteration when  $\text{sign}'(\mathbf{X}_{k-1})\delta\mathbf{X}_{k-1} < 1$ . Global perturbative bounds on this iteration have been derived by Bai and Demmel [93], while Byers, He and Mehrmann [6] developed asymptotic bounds. The automatic stability of sign iteration is a well developed theme in Ref.[14].

#### B. Square root iteration

In this work, we are concerned with resolution of the identity

$$\mathbf{I}(s) = s^{1/2} \cdot s^{-1/2}, \quad (8)$$

and its low-complexity computation with fast methods.

Starting with eigenvalues rescaled to the domain  $(0, 1]$  with the easily obtained largest eigenvalue,  $s \leftarrow s/s_{N-1}$ , and with  $\mathbf{z}_0 = \mathbf{I}$  and  $\mathbf{x}_0 = \mathbf{y}_0 = s$ , the corresponding canonical, “dual” channel square root iteration is:

$$\begin{aligned} \mathbf{y}_k &\leftarrow h_\alpha [\mathbf{y}_{k-1} \cdot \mathbf{z}_{k-1}] \cdot \mathbf{y}_{k-1} \\ \mathbf{z}_k &\leftarrow \mathbf{z}_{k-1} \cdot h_\alpha [\mathbf{y}_{k-1} \cdot \mathbf{z}_{k-1}], \end{aligned} \quad (9)$$

converging as  $\mathbf{y}_k \rightarrow s^{1/2}$ ,  $\mathbf{z}_k \rightarrow s^{-1/2}$  and  $\mathbf{x}_k \rightarrow \mathbf{I}$ , with eigenvalues aggregated towards 1 by the NS map  $h_\alpha[\mathbf{x}] = \frac{\sqrt{\alpha}}{2} (3 - \alpha \mathbf{x})$  [14, 91]. As in the case of sign iteration, this canonical iteration was shown by Higham, Mackey, Mackey and Tisseur [91] to remain strongly bounded in the super-linear regime, by idempotent Frechet derivatives about the fixed point  $(s^{1/2}, s^{-1/2})$ , in the direction  $(\delta\mathbf{y}_{k-1}, \delta\mathbf{z}_{k-1})$ :

$$\delta\mathbf{y}_k = \frac{1}{2}\delta\mathbf{y}_{k-1} - \frac{1}{2}s^{1/2} \cdot \delta\mathbf{z}_{k-1} \cdot s^{1/2} \quad (10)$$

$$\delta\mathbf{z}_k = \frac{1}{2}\delta\mathbf{z}_{k-1} - \frac{1}{2}s^{-1/2} \cdot \delta\mathbf{y}_{k-1} \cdot s^{-1/2}. \quad (11)$$

In addition to the dual channel instance, we also consider the “single” channel version of square root iteration,

$$\begin{aligned} \mathbf{z}_k &\leftarrow \mathbf{z}_{k-1} \cdot h_\alpha [\mathbf{x}_{k-1}], \\ \mathbf{x}_k &\leftarrow \mathbf{z}_k^\dagger \cdot s \cdot \mathbf{z}_k. \end{aligned} \quad (12)$$

### C. Mapping

The NS logistic map for the square root iteration is  $h_\alpha[\mathbf{x}] = \frac{\sqrt{\alpha}}{2} (3 - \alpha \mathbf{x})$ , with the initial rate of convergence controlled by  $h'_\alpha$  and the smallest eigenvalue,  $x_0$ . Various schemes for controlling the values  $\alpha$  towards convergence, notably by Pan and Schreiber [94], and more recently, Jie and Chen [92], who demonstrated  $2\times$  acceleration for very ill-conditioned problems with their new scaling approach.

In addition to scaling of the NS logistic, we introduce a stabilizing map that accounts for eigenvalues tossed out of bounds by  $\otimes_\tau$ . This stabilization is the transformation  $[0, 1] \rightarrow [0 + \varepsilon, 1 - \varepsilon]$  (shift and scale), carried out prior to application of the logistic.

The most important aspect of these scaling and stabilization maps is to turn them off towards convergence. Conventional methods often compute a lowest eigenvalue to monitor convergence [92, 94], but this may be too expensive for ill-conditioned problems. Alternatively, we monitor convergence simply with the relative trace error,  $t_k = (n - \text{tr } \tilde{\mathbf{x}}_k) n^{-1}$ . Then, empirically developed sigmoidal functions are used to damp the scaling parameter to unity,

$$\alpha(t) = 1. + 1.85 \times \left(1 + e^{-50.(t-.35)}\right)^{-1} \quad (13)$$

and the stability parameter to zero,

$$\varepsilon(t) = .1 \times \left(1 + e^{-75.(t-.30)}\right)^{-1}. \quad (14)$$

These functions are used throughout.

## IV. ERROR FLOWS IN SQUARE ROOT ITERATION

### A. Stability

Stability in the square root iteration is determined by the differential

$$\delta \mathbf{x}_k = \mathbf{x}_{\delta \hat{\mathbf{y}}_{k-1}} \times \delta \mathbf{y}_{k-1} + \mathbf{x}_{\delta \hat{\mathbf{z}}_{k-1}} \times \delta \mathbf{z}_{k-1} + \mathcal{O}(\tau^2) \quad (15)$$

which must remain bounded below one to avoid divergence. The corresponding Fréchet derivatives are

$$\mathbf{x}_{\delta \hat{\mathbf{y}}_{k-1}} = \lim_{\tau \rightarrow 0} \frac{\mathbf{x}(\mathbf{y}_{k-1} + \tau \delta \hat{\mathbf{y}}_{k-1}, \mathbf{z}_{k-1}) - \mathbf{x}_k}{\tau} \quad (16)$$

and

$$\mathbf{x}_{\delta \hat{\mathbf{z}}_{k-1}} = \lim_{\tau \rightarrow 0} \frac{\mathbf{x}(\mathbf{y}_{k-1}, \mathbf{z}_{k-1} + \tau \delta \hat{\mathbf{z}}_{k-1}) - \mathbf{x}_k}{\tau}, \quad (17)$$

along unit directions of the previous errors  $\delta \hat{\mathbf{y}}_{k-1}$  and  $\delta \hat{\mathbf{z}}_{k-1}$ , by an amount determined by the displacements

$\delta \mathbf{y}_{k-1} = \|\delta \mathbf{y}_{k-1}\|$  and  $\delta \mathbf{z}_{k-1} = \|\delta \mathbf{z}_{k-1}\|$ . In the single instance, we have simply:

$$\delta \mathbf{x}_k = \mathbf{x}_{\delta \hat{\mathbf{z}}_{k-1}} \times \delta \mathbf{z}_{k-1} + \mathcal{O}(\tau^2). \quad (18)$$

This formulation makes plain changes about the solvent, separating orientational effects of the directional derivatives, set mostly by the underlying exact linear algebra, from changes to error displacements, which involve both the action of derivatives on previous errors, as well as current SpAMM occlusion errors.

### B. Fréchet derivatives

In the dual instance, Fréchet derivatives occurring in Eq. (15) are:

$$\begin{aligned} \mathbf{x}_{\delta \hat{\mathbf{z}}_{k-1}} &= \mathbf{y}_{k-1} \cdot h'_\alpha \delta \hat{\mathbf{z}}_{k-1} \cdot \mathbf{y}_{k-1} \cdot \mathbf{z}_k \\ &\quad + \mathbf{y}_k \cdot \delta \hat{\mathbf{z}}_{k-1} \cdot h_\alpha [\mathbf{x}_{k-1}] \\ &\quad + \mathbf{y}_k \cdot \mathbf{z}_{k-1} \cdot \mathbf{y}_{k-1} \cdot h'_\alpha \delta \hat{\mathbf{z}}_{k-1}, \end{aligned} \quad (19)$$

and

$$\begin{aligned} \mathbf{x}_{\delta \hat{\mathbf{y}}_{k-1}} &= h_\alpha [\mathbf{x}_{k-1}] \cdot \delta \hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_k \\ &\quad + h'_\alpha \delta \hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_{k-1} \cdot \mathbf{y}_{k-1} \cdot \mathbf{z}_k \\ &\quad + \mathbf{y}_k \cdot \mathbf{z}_{k-1} \cdot h'_\alpha \delta \hat{\mathbf{y}}_{k-1} \cdot \mathbf{z}_{k-1}. \end{aligned} \quad (20)$$

Closer to the fixed point orbit,  $\mathbf{y}_k \cdot \mathbf{z}_{k-1} \rightarrow \mathbf{I}$ ,  $\mathbf{y}_{k-1} \cdot \mathbf{z}_k \rightarrow \mathbf{I}$ ,  $h_\alpha [\mathbf{x}_k] \rightarrow \mathbf{I}$  and  $h'_\alpha \rightarrow -\frac{1}{2}$  [91]. Then,

$$\mathbf{x}_{\delta \hat{\mathbf{y}}_{k-1}} \rightarrow \delta \hat{\mathbf{y}}_{k-1} \cdot (\mathbf{z}_k - \mathbf{z}_{k-1}) \quad (21)$$

and

$$\mathbf{x}_{\delta \hat{\mathbf{z}}_{k-1}} \rightarrow (\mathbf{y}_k - \mathbf{y}_{k-1}) \cdot \delta \hat{\mathbf{z}}_{k-1}. \quad (22)$$

Likewise, in the single channel instance:

$$\begin{aligned} \mathbf{x}_{\hat{\mathbf{z}}_{k-1}} &\rightarrow (\mathbf{z}_k - \mathbf{z}_{k-1})^\dagger \cdot \mathbf{s} \cdot \delta \hat{\mathbf{z}}_{k-1} \\ &\quad + \delta \hat{\mathbf{z}}_{k-1}^\dagger \cdot \mathbf{s} \cdot (\mathbf{z}_k - \mathbf{z}_{k-1}). \end{aligned} \quad (23)$$

About the fixed point then, error flow in the  $\mathbf{y}_k$  and the  $\mathbf{z}_k$  channels is tightly quenched, corresponding to  $\mathbf{x}_{\delta \hat{\mathbf{x}}_{k-1}} \rightarrow \mathbf{I}$  and identity iteration [91].

### C. Displacements

Countering orientational convergence, determined almost entirely by the underlying exact iterations, is the compounding displacement error, determined by SpAMM occlusion in each of three products, at each step, and also involving previous errors. Here, we look at just the displacement  $\delta \mathbf{z}_{k-1}$ , which has the largest potential

for divergence as we argue here and show numerically in the following section.

Including the SpAMM error in the  $\tilde{\mathbf{z}}_{k-1}$  update we have:

$$\begin{aligned}\tilde{\mathbf{z}}_{k-1} &= \tilde{\mathbf{z}}_{k-2} \otimes_{\tau} h_{\alpha}[\tilde{\mathbf{x}}_{k-2}] \\ &= \Delta_{\tau}^{\tilde{\mathbf{z}}_{k-2} \cdot h_{\alpha}[\tilde{\mathbf{x}}_{k-2}]} + \tilde{\mathbf{z}}_{k-2} \cdot h_{\alpha}[\tilde{\mathbf{x}}_{k-2}].\end{aligned}\quad (24)$$

Then, with  $h_{\alpha}[\tilde{\mathbf{x}}_{k-2}] = h_{\alpha}[\mathbf{x}_{k-2}] + h'_{\alpha} \delta \mathbf{x}_{k-2}$ , and taking  $\mathbf{z}_{k-1}$  from both sides, we find

$$\begin{aligned}\delta \mathbf{z}_{k-1} &= \Delta_{\tau}^{\tilde{\mathbf{z}}_{k-2} \cdot h_{\alpha}[\tilde{\mathbf{x}}_{k-2}]} \\ &\quad + \delta \mathbf{z}_{k-2} \cdot h_{\alpha}[\tilde{\mathbf{x}}_{k-2}] + \mathbf{z}_{k-2} \cdot h'_{\alpha} \delta \mathbf{x}_{k-2},\end{aligned}\quad (25)$$

which is bounded by

$$\begin{aligned}\delta \mathbf{z}_{k-1} &< \|\mathbf{z}_{k-2}\| (\tau \sigma_n \|h_{\alpha}[\tilde{\mathbf{x}}_{k-2}]\| + h'_{\alpha} \delta \mathbf{y}_{k-2} \|\mathbf{z}_{k-2}\|) \\ &\quad + \delta \mathbf{z}_{k-2} (\|h_{\alpha}[\tilde{\mathbf{x}}_{k-2}]\| + \|\mathbf{y}_{k-2}\|).\end{aligned}\quad (26)$$

In Eq. (26), the term  $h'_{\alpha} \delta \mathbf{y}_{k-2} \|\mathbf{z}_{k-2}\|^2$  is volatile, tending towards  $\delta \mathbf{y}_{k-2} \kappa(\mathbf{s})/2$ . Because of this sensitivity, and because the  $\mathbf{y}_k$  product channel maintains fidelity of the starting eigen-basis, we single out this “sensitive” product for a higher level of precision;  $\tau_s \ll \tau$ . Still, we expect different behavior from the single instance  $\tilde{\mathbf{y}}_{k-1} = \tilde{\mathbf{z}}_{k-1}^{\dagger} \otimes_{\tau_s} \mathbf{s}$ , and the dual instance  $\tilde{\mathbf{y}}_{k-1} = h_{\alpha}[\tilde{\mathbf{x}}_{k-1}] \otimes_{\tau_s} \tilde{\mathbf{y}}_{k-1}$ . This is because the spectral product is broader (resolving larger and smaller numbers) in the single instance and narrower in the dual instance.

#### D. Most approximate but still stable

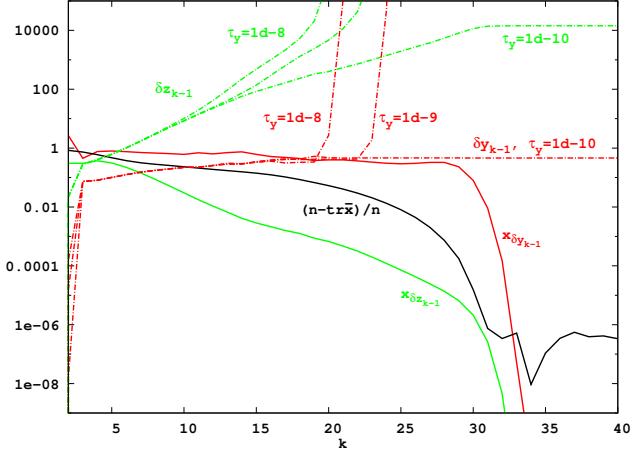


FIG. 2: Derivatives, displacements and trace error of the unscaled dual iteration. Derivatives are full lines, whilst the displacements corresponding to  $\tau_s = \{10^{-8}, 10^{-9}, 10^{-10}\}$  are the dashed lines. The trace error is shown as a full black line.

Experiments were carried out on the ill-conditioned ( $\kappa(\mathbf{s}) = 10^{10}$ ) nanotube metric-matrices described in Appendix B. We picked  $\tau = .001$  and block size  $b = 64$ .

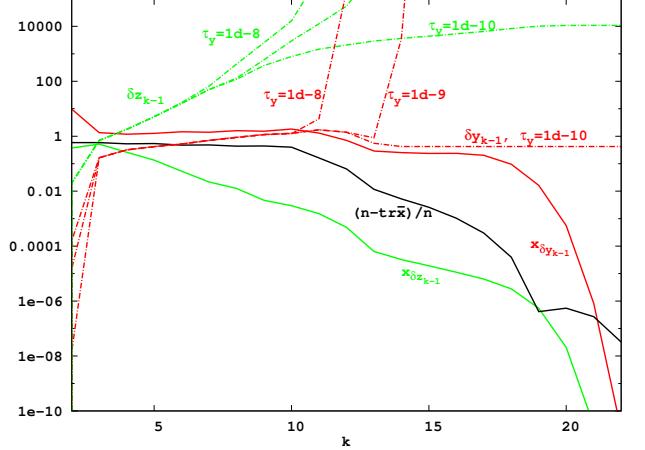


FIG. 3: Derivatives, displacements and the approximate trace of the scaled dual iteration. Derivatives are full lines, whilst the displacements corresponding to  $\tau_s = \{10^{-8}, 10^{-9}, 10^{-10}\}$  are the dashed lines. The trace error is shown as a full black line.

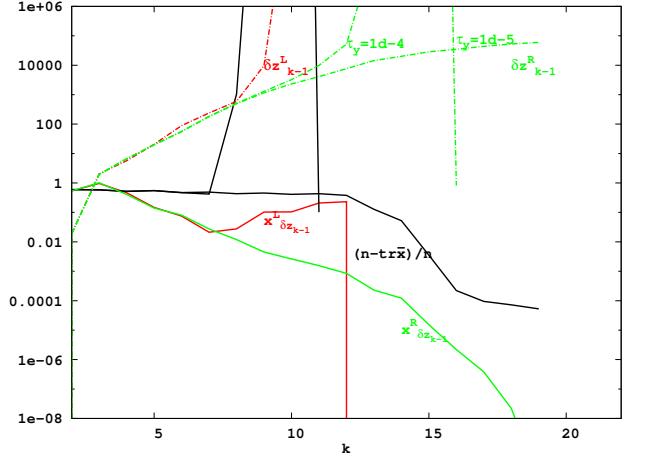


FIG. 4: Derivatives, displacements and the approximate trace of the scaled single iteration. Derivatives are full lines, whilst the displacements corresponding to  $\tau_s = \{10^{-3}, 10^{-4}, 10^{-5}\}$  are the dashed lines. The trace error is shown as a full black line.

Then, we looked at stability with respect to the tighter  $\tau_s$  threshold, the directional derivatives and the trace error.

In Fig. 2, unscaled results for the dual instance are shown. In Fig. 3, scaled results for the dual instance are given, reclaiming approximately 2/3 of the available  $2\times$  acceleration, with a 1/3 penalty due to the stabilization map described in Section III C. Then, in Fig. 4 we show results for the scaled single instance.

Not shown is complete fill-in at convergence for even the most-approximate-yet-still-stable (MAYSS) value of  $\tau_s$ . Also, with a less forgiving stability map, we find interesting left/right differences; namely, the right first

product

$$\tilde{\mathbf{x}}_k^R \leftarrow \tilde{\mathbf{z}}_k^\dagger \otimes_\tau (\mathbf{s} \otimes_{\tau_s} \tilde{\mathbf{z}}_{k-1}) , \quad (27)$$

is different from the left first product

$$\tilde{\mathbf{x}}_k^L \leftarrow (\tilde{\mathbf{z}}_k^\dagger \otimes_{\tau_s} \mathbf{s}) \otimes_\tau \tilde{\mathbf{z}}_{k-1} , \quad (28)$$

which is unsurprising. All results shown correspond to the right-first protocol.

The stability map parameters discussed Section III C are tuned away from such sensitivities, and their local variation has a negligible impact on stability. Also not shown, the intermediate volumes (before fill in) behave very differently with scaling; in all cases, the product volumes, proportional to the computational cost, are significantly increased by scaling. In the remainder of this work, we won't look at scaling again.

## V. REGULARIZATION

Even for the most approximate but stable (MAYSS) approximations, our nanotube calculations become dense in both the data and task domains, even for very large  $128 \times U.C.$  systems. For these ill-conditioned iterations, broad spectral resolutions that involve products of large and small numbers correspond to delocalized products that are not tightly bound by the SpAMM approximation. However, as we show later in Fig. 12, similarly ill-conditioned problems may achieve substantial compression with just the MAYSS approximation.

A systematic way to reduce ill-conditioning is through Tikhonov regularization [9, 95]. Regularization involves a small level shift of the eigenvalues,  $\mathbf{s}_\mu \leftarrow \mathbf{s} + \mu \mathbf{I}$ , altering the condition number of the shifted matrix to  $\kappa(\mathbf{s}_\mu) = \frac{\sqrt{s_{n-1}^2 + \mu^2}}{\sqrt{s_0^2 + \mu^2}}$  [9].

Achieving substantial acceleration with severe ill-conditioning may require a large level shift however, producing inverse factors of little practical use. One approach to recover a more accurate inverse factor is Riley's method based on Taylor's series [9];

$$\mathbf{s}^{-1/2} = \mathbf{s}_\mu^{-1/2} \cdot \left( \mathbf{I} + \frac{\mu}{2} \mathbf{s}_\mu^{-1} + \frac{3\mu^2}{8} \mathbf{s}_\mu^{-2} + \dots \right) . \quad (29)$$

For severely ill-conditioned problems and large level shifts, this expansion may converge very slowly. Also, adding powers of the full inverse may not be computationally effective.

### A. Product representation

We introduce an alternative representation of the regularized inverse factor;

$$\mathbf{s}^{-1/2} \equiv \bigotimes_{\substack{\tau=\tau_0 \\ \mu=\mu_0}} |\tau \mu ; \mathbf{s}^{-1/2}\rangle , \quad (30)$$

which is a telescoping product of preconditioned “slices” starting with a most-approximate-yet-still-effective-by-one-order (MAYEBOO) preconditioner,  $\mathbf{s}_{\tau_0 \mu_0}^{-1/2} \equiv |\tau_0 \mu_0 ; \mathbf{s}^{-1/2}\rangle^1$ . This sandwich of generic, thinly sliced SpAMM products allows to construct a nested scoping on precision via  $\tau$ , and in the effective condition number controlled by  $\mu$ .

### B. Effective by one order

We look again at the  $\kappa(\mathbf{s}) = 10^{10}$  nanotube series described in Appendix B, this time with extreme regularization,  $\mu_0 = .1$ , and at a finer granularity,  $b = 8$ . Culled  $\mathbf{y}_k$  and  $\mathbf{z}_k$  volumes (as percentage of the total work) for  $36 - 128 \times$  the (3,3) unit cell are shown for the MAYEBOO approximation in Fig. 5 for the single instance, and in Fig. 6 for the dual instance.

The behavior of these implementations is very different; in the single instance, a stable iteration could not be found at precision  $\tau_0 = .1$ . Stability could only be found at .01, and that with a poorly contained trace error and cull-volumes that continue to inflate past convergence, with a conspicuous  $\sqrt{k}$ -like dependence. On the other hand, dual iteration volumes demonstrate collapsing volumes and fast trapping of the trace error with resolution of the approximate identity.

We interpret these results as follows: The single instance suffers from an increasingly broad resolution of spectral powers,  $\tilde{\mathbf{y}}_k \rightarrow \mathbf{s}_{\tau_0 \mu_0}^{-1/2} \otimes_{\tau_0} \mathbf{s}_{\mu_0}$ , which is poorly bound by Eq. (5). In the dual instance however,  $\tilde{\mathbf{y}}_k \rightarrow \mathbf{I}_{\tau_0 \mu_0} \otimes_{\tau_0} \mathbf{s}_{\tau_0 \mu_0}^{1/2}$  and  $\tilde{\mathbf{z}}_k \rightarrow \mathbf{s}_{\tau_0 \mu_0}^{-1/2} \otimes_{\tau_0} \mathbf{I}_{\tau_0 \mu_0}$ , with Eq. (5) tightening to

$$\Delta^{\mathbf{I}_{\tau_0 \mu_0} \cdot \mathbf{s}_{\tau_0 \mu_0}^{1/2}} < \tau n \|\mathbf{s}_{\tau_0 \mu_0}^{1/2}\| \quad (31)$$

and

$$\Delta^{\mathbf{s}_{\tau_0 \mu_0}^{-1/2} \cdot \mathbf{I}_{\tau_0 \mu_0}} < \tau n \|\mathbf{s}_{\tau_0 \mu_0}^{-1/2}\| . \quad (32)$$

This contraction to the plane diagonal is compressive, leading to computational complexities that should approach quadtree copy in place.

### C. Iterative regularization

We now sketch an iterative approach to constructing the product representation, Eq. (30). In the dual instance, it is possible to obtain a first MAYEBOO approximation  $\mathbf{s}_{\tau_0=.1, \mu_0=.1}^{-1/2}$ , which improves the condition number by one order of magnitude, with a numerical resolution of one digit. Then, a next level

---

<sup>1</sup> Braket notation marks the potential for asymmetries in the intermediate representation.

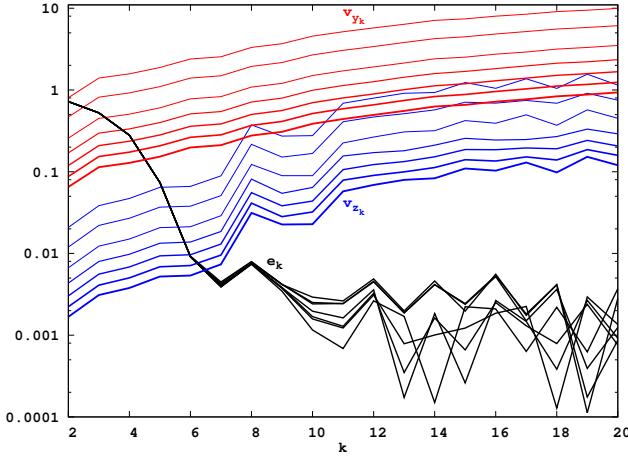


FIG. 5: Culled volumes in the thin slice, single instance approximation of  $s_{\tau_0 \mu_0}^{-1/2}$  for the (3,3) nanotube,  $\kappa(s) = 10^{10}$  matrix series described in Section B. In the “single” instance, it was not possible to achieve stability with  $\tau_0 = .1$ . In this “single” case, a thin slice corresponds to  $\mu_0 = .1, \tau_0 = 10^{-2}$  &  $\tau_s = 10^{-4}$ , and volumes are  $v_{\tilde{z}_{k-1}} = (\text{vol}_{\tilde{z}_{k-1} \otimes \tau h[\tilde{x}_{k-1}]} \times 100\%) / N^3$  and  $v_{\tilde{y}_k} = (\text{vol}_{s \otimes \tau_s \tilde{z}_k} \times 100\%) / N^3$ . Line width increases with increasing system size. Also shown is the trace error,  $e_k = (N - \text{tr } \mathbf{x}_k) / N$ .

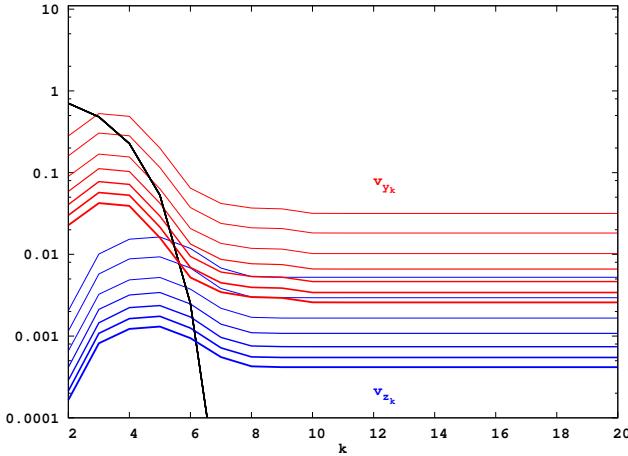


FIG. 6: Culled volumes in the thin slice, dual instance approximation of  $s_{\tau_0 \mu_0}^{-1/2}$  for the (3,3) nanotube,  $\kappa(s) = 10^{10}$  matrix series described in Section B. The thin slice corresponds to  $\mu_0 = .1, \tau_0 = .1$  &  $\tau_s = .001$  with volumes  $v_{\tilde{y}_k} = (\text{vol}_{h[\tilde{x}_{k-1}] \otimes \tau_s \tilde{y}_k} \times 100\%) / N^3$  and  $v_{\tilde{z}_k} = (\text{vol}_{\tilde{z}_{k-1} \otimes \tau h[\tilde{x}_{k-1}]} \times 100\%) / N^3$ . Line width increases with increasing system size. Also shown is the trace error,  $e_k = (N - \text{tr } \mathbf{x}_k) / N$ , which rapidly approaches  $10^{-11}$  (not shown).

slice can be found,  $s_{\tau_0 \mu_1}^{-1/2}$ , based on the residual  $(s_{\tau_0 \mu_0}^{-1/2})^\dagger \otimes_{\tau_1} (s + \mu_1 \mathbf{I}) \otimes_{\tau_1} s_{\tau_0 \mu_0}^{-1/2}$ , with e.g.  $\mu_1 = .01$  and  $\tau_1 = .01$ . The product  $s_{\tau_0 \mu_1}^{-1/2} \otimes_{\tau_1} s_{\tau_0 \mu_0}^{-1/2}$  then improves the condition number by two orders of magnitude, still with a numerical resolution of one digit. In preliminary work,

it appears necessary to compute the residual at a higher level of precision, e.g. using  $\otimes_{\tau_1}$  instead of  $\otimes_{\tau_0}$  and with  $\tau_1 > \tau_0$ .

In this way, it may be possible to obtain product representation of the inverse square root at a SpAMM resolution that is potentially far more permissive than otherwise possible,

$$s_{\tau_0}^{-1/2} = s_{\tau_0 \mu_n}^{-1/2} \otimes_{\tau_1} s_{\tau_0 \mu_{n-1}}^{-1/2} \otimes_{\tau_1} \dots s_{\tau_0 \mu_0}^{-1/2}, \quad (33)$$

assuming  $.1 \geq \mu_0 > \mu_1 > \dots$ . Likewise, it may also be possible to obtain the full inverse factor with increasing numerical resolution as

$$s^{-1/2} = s_{\tau_m}^{-1/2} \otimes_{\tau_{m+1}} s_{\tau_{m-1}}^{-1/2} \otimes_{\tau_m} \dots s_{\tau_0}^{-1/2}, \quad (34)$$

and  $.1 \geq \tau_0 > \tau_1 > \dots$

With each step a well conditioned generic slice, it may be possible to find a more effective logistic map optimized for the corresponding eigenvalue distributions. Also, there are other paths that can be taken in the scoping of regularization and precision that may be far more efficient.

## VI. LOCALITY

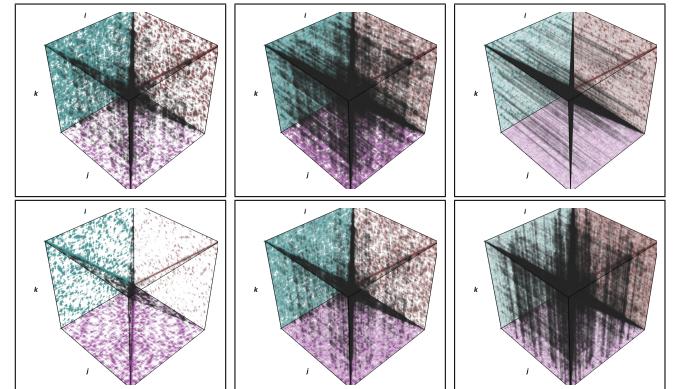


FIG. 7: The  $ijk$  task and data space for construction of the unregularized preconditioner  $|\tau_0 = .001, \mu_0 = .0; s^{-1/2}\rangle$ , with dual instance square root iteration, and for the 6-311G\*\* metric of 100 periodic water molecules at STP. At top its  $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$  for  $k = 0, 4, \& 15$ , while on the bottom we have  $\mathbf{x}_k = \mathbf{y}_k \otimes_{\tau} \mathbf{z}_k$  for  $k = 0, 4, \& 15$ . Maroon is  $a$ , purple is  $b$ , green is  $c$ , and black is the volume  $\text{vol}_{a \otimes_r b}$  in the product  $c = a \otimes_r b$ .

### A. Spatial, metric and temporal locality

Astrophysical  $n$ -body algorithms employ range queries over spatial databases to hierarchically discover and compute approximations that commit only small errors. Often, these spatial databases are ordered with a space filling curve (SFC) [23–32], which maps points that are close

in space to an index where they are also close. The block-by-magnitude structures empowering the SpAMM approximation are metric localities corresponding to metric decay, in quantum chemical examples corresponding to an underlying SFC ordering of Cartesian coordinates.

Warren and Salmon showed how to parlay spatial locality into temporal locality, remapping and repartitioning the space filling curve to rebalance distributed  $n$ -body tasks, based on accumulated histories (persistence data). In a similar way, we showed how persistence can be used to achieve strong parallel scaling for SpAMM with commonly available runtimes [65]. Persistence data, providing temporal locality, may also be useful in mathematical approximation.

In Figure 7, we show  $\otimes_\tau$  volumes for square root iteration, corresponding to the metric of a small, periodic water box with the large 6-311G\*\* basis (Appendix B). For the 3-d periodic case, diminishing Cartesian separations lead to long-skinny delocalizations (pillae) and much denser matrices, relative to e.g. a one-dimensional nano-tube.

These delocalizations correspond to weakness in Eq. (5), and to the tighter thresholds required to maintain a single iteration in the MAYSS approximation. This effect is even more pronounced in the single instance (not shown), where delocalizations are exaggerated due to spectral resolutions that are broader. Eventually, Cartesian separation will thin the density of these delocalizations, with complexity reduction due to metric separation and decay only. However, this may be painfully slow.

## B. Algebraic locality

In Figures 8 and 9, we show a new kind of locality that is uniquely exploited by  $n$ -body approximation of the square root iteration. This locality increases compressively towards convergence, as the contractive identity iteration develops. We call this compression *lensing*, involving collapse of the culled volume about plane diagonals of the resolvent. Lensing corresponds to strengthening Eq.(5), and to strongly contracting directional derivatives, Eqs. (21)-(22). This is an important, mitigating computational effect for the sensitive  $\mathbf{y}_k$  channel, with  $\tau_s \sim .01 \times \tau$ .

In addition, non-Euclidean measures are relevant for achieving metric locality in the SpAMM algebra, including information measures, space filling curve generalizations, as well as graph reorderings that band matrix elements about the diagonal, a common approach in structural mechanics. In Figure 12 we show development of a first, unregularized (MAYSS) preconditioner for such an example; the structural matrix  $\mathbf{s} = \text{bcsstk14}$  is a  $\kappa(\mathbf{s}) = 10^{10}$  matrix corresponding to the roof of the Omni Coliseum in Atlanta. These results show remarkable gossamer sheeting and flattening along plane diagonals, at top for development of  $\mathbf{y}_k$ , as well as hollow accumulation of  $\text{vol}_{\mathbf{y}_k \otimes_\tau \mathbf{z}_k}$  at bottom. Surprisingly, this

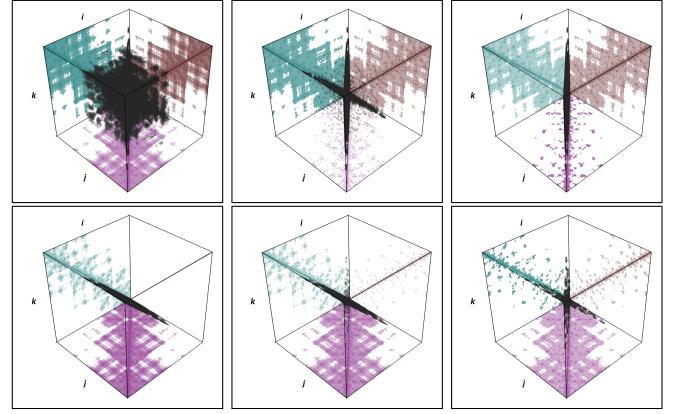


FIG. 8: The  $ijk$  task and data space for construction of the MAYEBOO preconditioner  $|\tau_0 = .1, \mu_0 = .1; \mathbf{s}^{-1/2}\rangle$ , with dual instance square root iteration, and for an  $8 \times U.C. (3,3)$   $\kappa(\mathbf{s}) = 10^{11}$  nanotube.  $\mathbf{y}_k$  appears wider than  $\mathbf{z}_k$  because it is computed at a higher precision,  $\tau_s = .001$ , and because the first multiply involves  $\mathbf{s}^2$ . At top its  $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$  for  $k = 0, 4, \& 16$ , while on the bottom we have  $\mathbf{x}_k = \mathbf{y}_k \otimes_\tau \mathbf{z}_k$  for  $k = 0, 2, \& 16$ . Maroon is  $\mathbf{a}$ , purple is  $\mathbf{b}$ , green is  $\mathbf{c}$ , and black is the volume  $\text{vol}_{\mathbf{a} \otimes_\tau \mathbf{b}}$  in the product  $\mathbf{c} = \mathbf{a} \otimes_\tau \mathbf{b}$ .

example demonstrates lensing for a tight MAYSS approximation, while the equally ill-conditioned & lower dimensional  $\kappa(\mathbf{s}) = 10^{10}$  nanotube MAYSS approximation remains full (dense) through U.C.  $\times 128$ .

## C. Complexity reduction

Finally, we show complexity reduction at convergence of the MAYEBOO approximation relative to the MAYSS approximation, in Fig. 10 for periodic water boxes, and in Fig. 11 for the ill-conditioned nano-tube. The two-orders difference between  $\mathbf{y}_k$  and  $\mathbf{z}_k$  volumes corresponds precisely to  $\tau_s \sim \tau \times .01$ , with  $\mathbf{x}_k$  in between. Except for the slower trend in Fig. (10)'s  $\mathbf{x}_k$  volume, we see the potential for continued strong acceleration with increasing system size. Understanding these subtleties is the subject of future work.

## VII. CONCLUSIONS

In this work, we developed the  $n$ -body solver SpAMM for square root iteration. Main contributions include a modified Cauchy-Schwarz criterion, Eq. 4, and proof that the cooresponding product relative error is bound by Eq. (5). Also, we demonstrated a new kind of algebraic locality that develops with strongly contractive identity iteration.

This work is gauged against other methods for fast matrix multiplication discussed in Section II C. Against SpGEMM, the  $n$ -body approach offers a bounded control over relative errors in the product and the ability to resolve complex algebraic structures, about plane-diagonals of the  $ijk$ -cube, along tall-skinny pillae and for

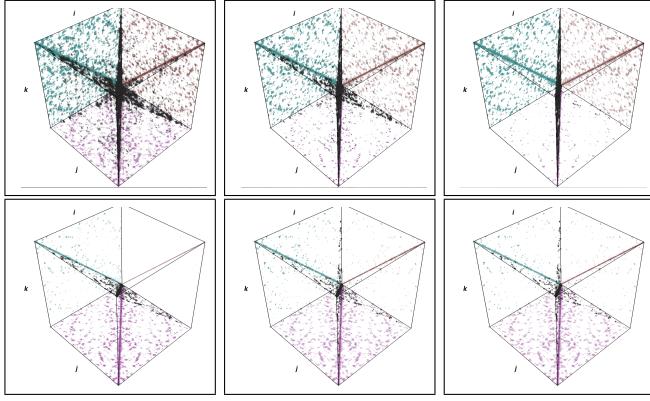


FIG. 9: The  $ijk$  task and data space for construction of the MAYEBOO preconditioner  $|\tau_0 = .1, \mu_0 = .1; s^{-1/2}\rangle$ , with dual instance square root iteration and for 6-311G\*\* metric of 100 periodic water molecules at STP. At top its  $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$  for  $k = 0, 4, \& 15$ , while on the bottom we have  $\mathbf{x}_k = \mathbf{y}_k \otimes_{\tau} \mathbf{z}_k$  for  $k = 0, 4, \& 15$ . Maroon is  $\mathbf{a}$ , purple is  $\mathbf{b}$ , green is  $\mathbf{c}$ , and black is the volume  $\text{vol}_{a \otimes_{\tau} b}$  in the product  $\mathbf{c} = \mathbf{a} \otimes_{\tau} \mathbf{b}$ .

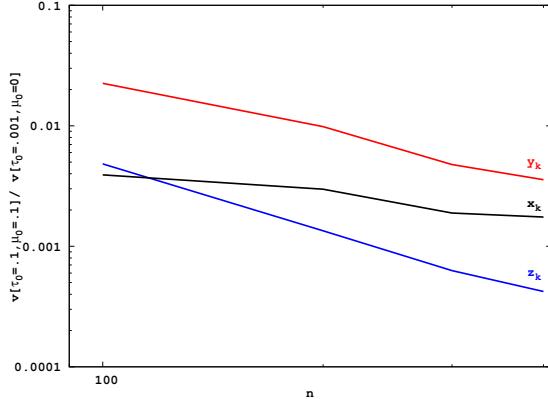


FIG. 10: Complexity reduction in metric square root iteration for periodic 6-311G\*\* water. Shown is the ratio of lensed product volumes for the regularized most-approximate-yet-effective-by-one-order (MAYEBOO) approximation and the unregularized most-approximate-yet-still-stable (MAYSS) approximation.

volumetric contractions to lower dimensional objects via lensing. The  $n$ -body method uniquely and synergistically exploits two distinct forms of locality, metric locality cooresponding to a Cartesian or non-Euclidean decay principle, and algebraic locality cooresponding to contractive identity iteration. Also, strong parallel scaling for the  $\mathcal{O}(n)$  electronic structure problem has been demonstrated with the SpAMM kernel, a feature that remains elusive for methods based on the SpGEMM [96].

Beyond the fast matrix multiply,  $n$ -body frameworks may enable additional, layered functionalities and economizations in complex solver ecosystems, with facile interoperability and mathematical agility, through generic recursion and skelitization, and with common runtimes

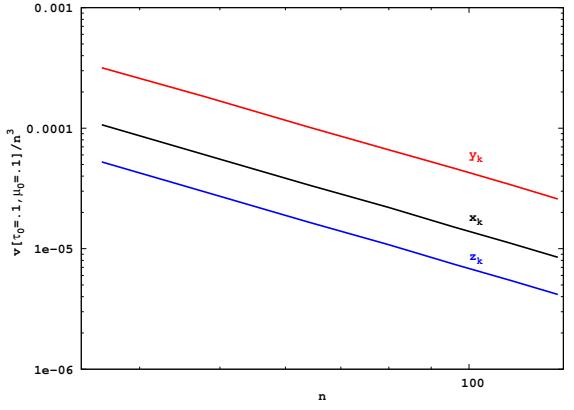


FIG. 11: Complexity reduction in metric square root iteration for periodic 6-311G\*\* water. Shown is the ratio of lensed product volumes for the regularized most-approximate-yet-effective-by-one-order (MAYEBOO) approximation and the unregularized most-approximate-yet-still-stable (MAYSS) approximation.

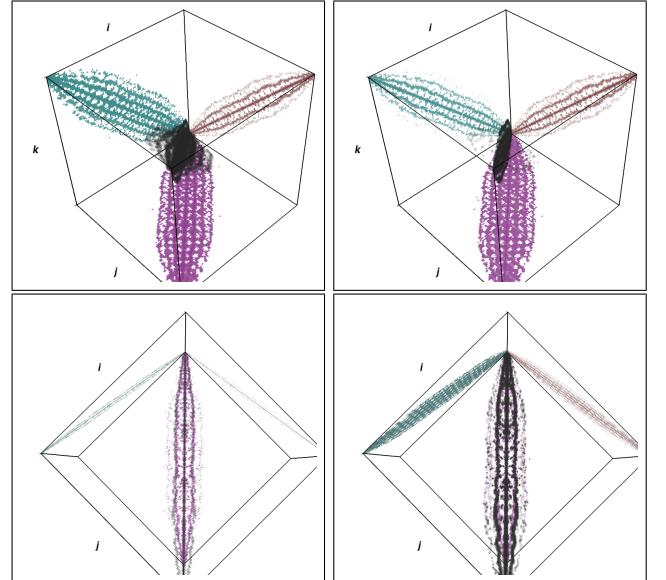


FIG. 12: The  $ijk$  task and data space for construction of the unregularized preconditioner  $|\tau_0 = .001, \mu_0 = .0; s^{-1/2}\rangle$ , with the dual instance of square root iteration and for 6-311G\*\* metric of 100 periodic water molecules at STP. At top its  $\mathbf{y}_k = h_\alpha[\mathbf{x}_{k-1}] \otimes_{\tau_s} \mathbf{y}_{k-1}$  for  $k = 0, 4, \& 15$ , while on the bottom we have  $\mathbf{x}_k = \mathbf{y}_k \otimes_{\tau} \mathbf{z}_k$  for  $k = 0, 4, \& 15$ . Maroon is  $\mathbf{a}$ , purple is  $\mathbf{b}$ , green is  $\mathbf{c}$ , and black is the volume  $\text{vol}_{a \otimes_{\tau} b}$  in the product  $\mathbf{c} = \mathbf{a} \otimes_{\tau} \mathbf{b}$ .

able to exploit temporal and data localities. For example, we recently generalized SpAMM recursion to the problem of Fock-exchange, with a recursive triple (hextree) metric querry on the Almlöf-Alhrichs direct SCF criteria [58]. Also, equivalence with the matrix sign function via Higham's identity, Eq. (1), and close structural relationships with the polar decomposition [91] may enable to extend functionality of the  $n$ -body iterations developed

here.

Despite these compelling qualities,  $n$ -body square-root iteration must be gauged by its ability to compute a high quality inverse factor. Here, we have only looked at complexity and stability of the most approximate preconditioners for a few systems and for a few numerical experiments. However, the results are new and encouraging, showing the potential for many orders of magnitude re-

duction in complexity.

## Appendix A: Implementation

## Appendix B: Data

- 
- [1] O. Penrose and J. Lebowitz, *Commun. Math. Phys.* **184**, 1 (1974).
- [2] J. Voit, *The Statistical Mechanics of Financial Markets, Theoretical and Mathematical Physics* (Springer Berlin Heidelberg, 2006).
- [3] L. Anselin, *Int. Reg. Sci. Rev.* **26**, 153 (2003).
- [4] J. Hardin, S. R. Garcia, and D. Golan, *Ann. Appl. Stat.* **7**, 1733 (2013).
- [5] I. Krishtal, T. Strohmer, and T. Wertz, *Found. Comp. Math.* **15**, 931 (2015).
- [6] R. Schaback, *Adv. Comput. Math.* **3**, 251 (1995).
- [7] R. Schaback and H. Wendland, *Acta Numerica* **15**, 543 (2006).
- [8] B. Fornberg and N. Flyer, *Acta Numerica* (2015), Solving PDEs with Radial Basis Functions.
- [9] S. A. Sarra, *Engineering Analysis with Boundary Elements* **44**, 76 (2014).
- [10] U. Röthlisberger, M. Sprik, and J. Hutter, in *Bridging Time Scales: Molecular Simulations for the Next Decade*, edited by P. Nielaba, M. Mareschal, and G. Ciccotti (Springer Berlin Heidelberg, 2002), vol. 605 of *Lecture Notes in Physics*, pp. 413–442.
- [11] B. Jansik, S. Host, P. Jorgensen, J. Olsen, and T. Helgaker, *J. Chem. Phys.* **126**, 124104 (2007).
- [12] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-structure Theory* (Wiley, 2008).
- [13] P. O. Lowdin, *Advances in Physics* **5**, 1 (1956).
- [14] N. J. Higham, *Functions of Matrices* (Society for Industrial & Applied Mathematics, 2008).
- [15] F. G. Gustavson, *ACM Transactions on Mathematical Software (TOMS)* **4**, 250 (1978).
- [16] S. Toledo, *IBM J. Res. Dev.* **41**, 711 (1997).
- [17] M. Challacombe, *Comput. Phys. Commun.* **128**, 93 (2000).
- [18] D. R. Bowler and T. M. and M. J. Gillan, *Comp. Phys. Comm.* **137**, 255 (2000).
- [19] M. Benzi and M. Tuma, *Appl. Numer. Math.* **30**, 305 (1999).
- [20] M. Benzi, *J. Comput. Phys.* **182**, 418 (2002).
- [21] D. R. Bowler and T. Miyazaki, *Reports Prog. Phys.* **75**, 36503 (2012).
- [22] M. Benzi, P. Boito, and N. Razouk, *SIAM Rev.* **55**, 3 (2013).
- [23] D. S. Wise, *SIGSAM Bull.* **18**, 24 (1984).
- [24] S. K. Abdali and D. S. Wise, in *Symbolic and Algebraic Computation*, edited by P. Gianni (Springer Berlin / Heidelberg, 1989), vol. 358 of *Lecture Notes in Computer Science*, pp. 96–108.
- [25] H. Samet, *The design and analysis of spatial data structures* (Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990).
- [26] D. Wise, *J. Parallel Distrib. Comput.* **9**, 282 (1990).
- [27] D. S. Wise, in *Euro-Par 2000 Parallel Processing*, edited by A. Bode, T. Ludwig, W. Karl, and R. Wismüller (Springer Berlin / Heidelberg, 2000), vol. 1900 of *Lecture Notes in Computer Science*, pp. 774–783, ISBN 978-3-540-67956-1.
- [28] K. P. Lorton and D. S. Wise, in *Proceedings of the 2006 workshop on Memory performance: Dealing with Applications, systems and architectures* (ACM, New York, NY, USA, 2006), MEDEA ’06, pp. 5–12.
- [29] H. Samet, *Foundations of Multidimensional and Metric Data Structures* (Morgan Kaufmann, 2006).
- [30] M. D. Adams and D. S. Wise, in *Proceedings of the 2006 workshop on Memory system performance and correctness* (ACM, New York, NY, USA, 2006), MSPC ’06, pp. 41–50.
- [31] M. Bader, *Space-Filling Curves: An Introduction with Applications in Scientific Computing*, vol. 9 of *Texts in Computational Science and Engineering* (Springer Berlin Heidelberg, 2013).
- [32] N. Bock and M. Challacombe, *SIAM Journal on Scientific Computing* **35**, C72 (2013).
- [33] S. Aluru and F. E. Sevilgen, in *Proceedings of the 4th IEEE Conference on High Performance Computing* (1997), pp. 230–235.
- [34] M. Bader and C. Zenger, *Lecture Notes in Computer Science* **3911**, 1042 (2006).
- [35] P. M. Campbell, K. D. Devine, J. E. Flaherty, L. G. Gervasio, and J. D. Teresco, Tech. Rep., Williams College Department of Computer Science (2003).
- [36] K. D. Devine, E. G. Boman, R. T. Heaphy, B. A. Hendrickson, J. D. Teresco, J. Faik, J. E. Flaherty, and L. G. Gervasio, *Applied Numerical Mathematics* **52**, 133 (2005), ADAPT ’03: Conference on Adaptive Methods for Partial Differential Equations and Large-Scale Computation.
- [37] I. Lashuk, A. Chandramowlishwaran, H. Langston, T.-A. Nguyen, R. Sampath, A. Shringarpure, R. Vuduc, L. Ying, D. Zorin, and G. Biros, in *SC ’09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (ACM, New York, NY, USA, 2009), pp. 1–12.
- [38] F. Gunther, M. Mehl, M. Pogl, and C. Zenger, *SIAM J. Sci. Comput.* **28**, 1634 (2006).
- [39] J. J. Camata and A. L. G. a. Coutinho, *2010 22nd Int. Symp. Comput. Archit. High Perform. Comput.* pp. 151–158 (2010).
- [40] I. Bartholdi, J.J. and P. Goldsman, *Visualization and Computer Graphics, IEEE Transactions on* **10**, 484 (2004).
- [41] M. Challacombe, *The Journal of chemical physics* **110**,

- 2332 (1999).
- [42] M. Challacombe, Comp. Phys. Comm. **128**, 93 (2000).
- [43] M. Challacombe and N. Bock, arXiv:cs.DS **1011.3534** (2010).
- [44] Kahan, *A Tutorial Overview of Vector and Matrix Norms* (2013), URL [www.cs.berkeley.edu/~wkahan/MathH110/NormOvrv.pdf](http://www.cs.berkeley.edu/~wkahan/MathH110/NormOvrv.pdf).
- [45] V. Strassen, Numer. Math. **13**, 354 (1969).
- [46] J. Demmel, I. Dumitriu, and O. Holtz, Numerische Mathematik **108**, 59 (2007).
- [47] G. Ballard, J. Demmel, O. Holtz, and O. Schwartz, Communications of the ACM **57**, 107 (2014).
- [48] F. Le Gall, in *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation* (ACM, New York, NY, USA, 2014), ISSAC '14, pp. 296–303.
- [49] Y. F. Andris Ambainis and F. L. Gall, in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing* (ACM, New York, NY, USA, 2015), STOC '15, pp. 585–593.
- [50] H. Cohn and C. Umans, in *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on* (IEEE, 2003), pp. 438–449.
- [51] H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans, in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on* (IEEE, 2005), pp. 379–388.
- [52] C. Umans, in *Proceedings of the 2006 International Symposium on Symbolic and Algebraic Computation* (ACM, New York, NY, USA, 2006), ISSAC '06, pp. 5–5, ISBN 1-59593-276-3.
- [53] A. Gray and A. Moore, in *Advances in Neural Information Processing Systems 13* (MIT Press, 2000), pp. 521–527.
- [54] A. G. Gray, Ph.D. thesis (2003).
- [55] G. Kuper, L. Libkin, and J. Paredaens, eds., *Genericity in Spatial Databases* (2000).
- [56] F. Geerts, S. Haesevoets, and B. Kuijpers, in *Database Programming Languages*, edited by G. Ghelli and G. Grahe (Springer Berlin Heidelberg, 2002), vol. 2397 of *Lecture Notes in Computer Science*, pp. 198–212.
- [57] P. Gottschling, D. Wise, and A. Joshi, Int. J. Parallel, ... (2009).
- [58] M. Challacombe and N. Bock, arXiv preprint arXiv:1401.6961 (2014).
- [59] M. Challacombe, J. Chem. Phys. **113**, 10037 (2000).
- [60] C. K. Gan and M. Challacombe, J. Comp. Phys. **118**, 9128 (2003).
- [61] M. Challacombe, E. Schwegler, and J. Almlöf, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1996), vol. 1 of *Computational Chemistry: Reviews of Current Trends*, pp. 53–107.
- [62] M. Challacombe, E. Schwegler, and J. Almlöf, J. Chem. Phys. **104**, 4685 (1996).
- [63] M. Challacombe and E. Schwegler, J. Chem. Phys. **106**, 5526 (1997).
- [64] M. Challacombe, C. White, and M. Head-Gordon, J. Chem. Phys. **107**, 10131 (1997).
- [65] N. Bock, M. Challacombe, and L. V. Kalé (2014), URL <http://arxiv.org/abs/1403.7458>.
- [66] A. Sarje and S. Aluru, in *Proc. Int. Conf. Parallel Process.* (2010), pp. 343–352.
- [67] H. González-Vélez and M. Leyton, Softw. Pract. Exper. **40**, 1135 (2010).
- [68] P. Totoo and H.-W. Loidl, Concurrency and Computation: Practice and Experience **26**, 987 (2014).
- [69] C. Rodrigues, T. Jablin, A. Dakkak, and W.-M. Hwu, in *ACM SIGPLAN Notices* (ACM, 2014), vol. 49, pp. 247–258.
- [70] W. B. March, B. Xiao, and G. Biros, SIAM J. Sci. Comput. **37**, A1089 (2015).
- [71] M. S. Warren and J. K. Salmon, in *Proceedings of the 1993 ACM/IEEE conference on Supercomputing* (ACM, New York, NY, USA, 1993), Supercomputing '93, pp. 12–21.
- [72] M. S. Warren and J. K. Salmon, *A parallel, portable and versatile treecode* (SIAM, Philadelphia, 1995), chap. 1.
- [73] M. S. Warren and J. K. Salmon, Computer Physics Communications **87**, 266 (1995).
- [74] M. S. Warren and J. K. Salmon, *Simulating the motion of 322,000,000 self-gravitating particles* (1997), ACM Gordon Bell Prize.
- [75] M. S. Warren, in *Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2013), p. 72.
- [76] M. Driscoll, E. Georganas, P. Koanantakool, E. Solomonik, and K. Yelick, in *Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on* (IEEE, 2013), pp. 1075–1084.
- [77] T. Sarlos, In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)* (2006).
- [78] P. Drineas, R. Kannan, and M. W. Mahoney, SIAM Journal on Computing **36**, 132 (2006).
- [79] M. W. Mahoney, pp. 1–53 (2012), arXiv:1502.03032v1.
- [80] R. Pagh, ACM Trans. Comput. Theory **5**, 1 (2013).
- [81] D. P. Woodruff, **10**, 1 (2015), arXiv:1411.4357v3.
- [82] P. Mishra and M. H. Eich, ACM Comput. Surv. **24**, 63 (1992).
- [83] E. G. Hoel and H. Samet, in *Parallel Processing, 1994. ICPP 1994. International Conference on* (1994), vol. 3, pp. 227–234.
- [84] E. H. Jacox and H. Samet, ACM Trans. Database Syst. **28**, 230 (2003).
- [85] S. Chen, A. Ailamaki, P. B. Gibbons, and T. C. Mowry, ACM Trans. Database Syst. **32** (2007).
- [86] R. R. Amossen and R. Pagh, in *Proceedings of the 12th International Conference on Database Theory* (ACM, 2009), pp. 121–126.
- [87] M. D. Lieberman, J. Sankaranarayanan, and H. Samet, in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering* (IEEE Computer Society, Washington, DC, USA, 2008), pp. 1111–1120, ISBN 978-1-4244-1836-7.
- [88] C. Kim, T. Kaldewey, V. W. Lee, E. Sedlar, A. D. Nguyen, N. Satish, J. Chhugani, A. Di Blas, and P. Dubey, Proc. VLDB Endow. **2**, 1378 (2009).
- [89] M. A. Iwen and C. V. Spencer, Information Processing Letters **109**, 468 (2009).
- [90] J. Chiu and L. Demanet, SIAM Journal on Numerical Analysis **50**, 171 (2012).
- [91] N. J. . Higham, N. M. D . Steven Mackey, and F. . coise Tisseur, SIAM J. Matrix Anal. Appl. **26**, 849 (2005).
- [92] J. Chen and E. Chow, Preprint ANL/MCS-P5059-0114, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439 (2014).
- [93] Z. Bai and J. Demmel, SIAM J. Matrix Anal. Appl. **19**, 205 (1998).

- [94] V. Pan and R. Schreiber, SIAM J. Sci. Stat. Comput. **12**, 466 (1991).
- [95] A. Neumaier, SIAM review **40**, 636 (1998).
- [96] D. R. Bowler, T. Miyazaki, L. A. Truflandier, and M. J. Gillan, arXiv:cont-mat.mtrl-sci **1402.6828** (2014).