# *N*-Body Algorithms for Matrices with Decay: Sparse Approximate Matrix Multiplication (SpAMM) and Inverse Factorization.

Matt Challacombe, Nicolas Bock & Terry Haut
Los Alamos National Laboratory
matt.challacombe@freeon.org

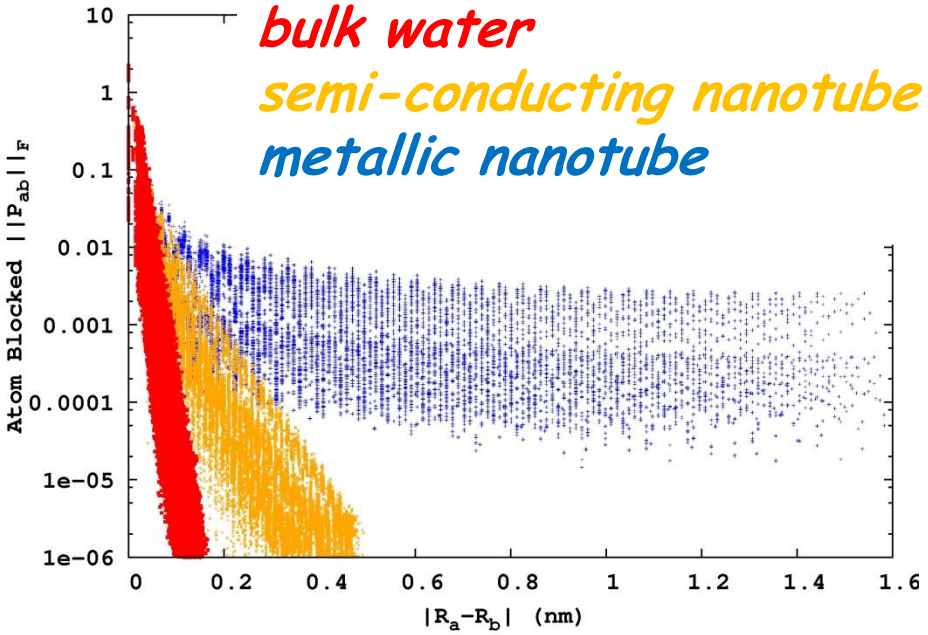
*"Linear-Complexity Dense Linear Algebra, Parallelization and Applications "*

**SIAM CS&E 2015, Salt Lake**

**LA-UR-10-07458//LA-UR 11-06091//LA-UR-14-22050//LA-UR-14-20354//**
Sponsored by: **DOE LDRD-ER grant 20110230ER** & TenBar Café

# Quantum Locality & Kohn's Nearsighted Principle

In a local, atom centered representation, quantum mechanical matrices possess decay properties. For non-metallic systems, matrix elements decay exponentially with atom-atom separation.





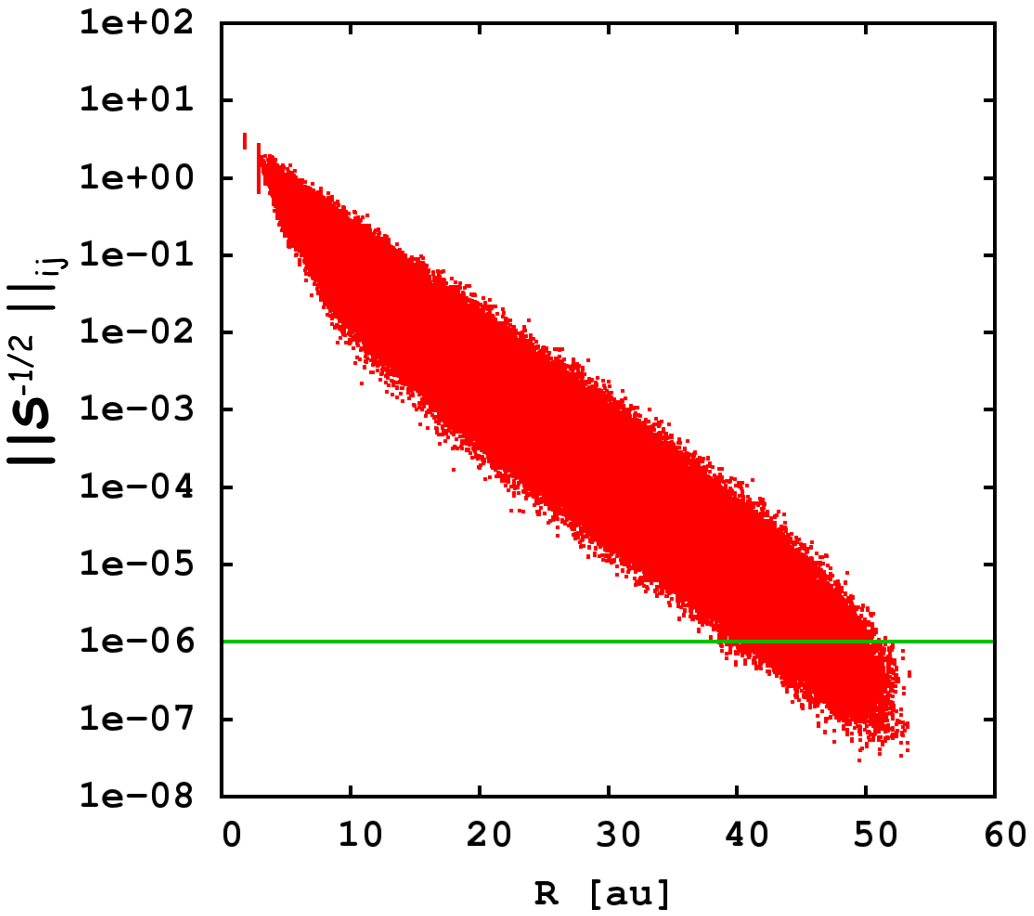
For physics poor models and uninteresting chemical problems, it is possible to achieve extreme locality leading to very sparse Hamiltonian matrices:

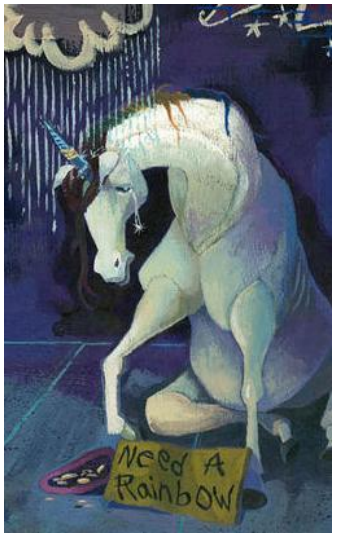
- Use conventional, sparse linear algebra to achieve an $O(n)$ cost with system size, $n$

# Dealing with a Real Basis; How Local?

***"For extended basis sets such as TZV2P, the matrices in the DM-based methods become sufficiently sparse only for large systems containing ~<u>10 000</u> water molecules" – J. Hutter group***

*J. Chem. Theory Comput., 2013, 9 (10), pp 4421–4427*



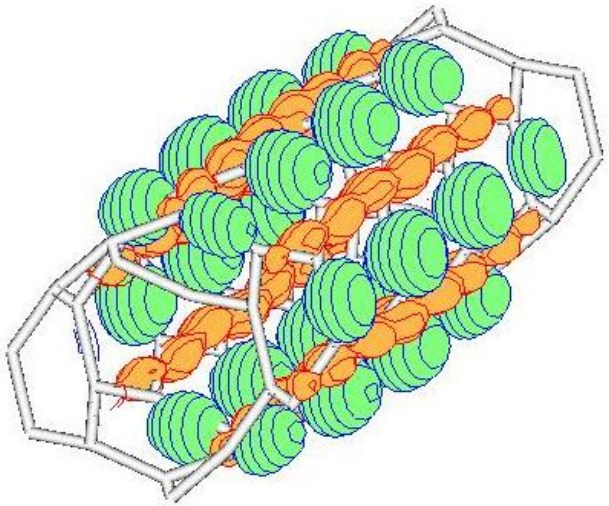




Whoa …

Even for 6-311G** 350 waters and 1d-8 error in a matrix element of $S^{-1/2}$, we don't begin to truncate until past 25Å!

# Real Wave-Functions Have Extent: Ill-Conditioning in Electronic Structure

From bond-breaking to long-range entanglement, interesting chemistries and technologically important properties emerge from long-range, strong-correlation effects.



These long range effects involve slow decay due to metric (basis) and gap ill-conditioning that preclude sparse matrix methods. Examples include gap and metric ill-conditioning in the 3,3 nanotube. We attack an example nanotube with $\boldsymbol{\kappa}(s) = 10^{11}$!

**As an alternative to sparse matrix methods, we develop the linear algebra of matrices with decay as an *n*-body problem, targeting fast methods for problems with slow exponential or algebraic decay.**

# *N*-Body Solvers in the Information, Physical & Computer Sciences

- *N*-Body solvers combine ***database operations*** (range & metric queries) with ***locality preserving heuristics***, and a wide variety of ***mathematical approximations***. Examples: the astrophysical ***Barnes-Hut tree-code***, the ***Fast Gauss Transform*** and so on.

- The ***generic n-body*** model has been extended to a vast number of fast, pairwise (kernel) summation techniques in the information sciences. Examples: see www.fast-lab.org.

- In ***functional programming***, the *n*-body problem may be developed with the formal properties of ***generacity***, involving map, fold, reduce & *etc*. Examples: the parallel map skeleton, algorithmic skeleton frameworks and so on.

# A Generalized Ecosytem of *N*-Body Solvers

## PHYSICS: **Strong Correlation**

- single determinant KSTs. Correlation on top of ***Fock exchange***, eg. B13.
- toward Mott transition, ***ill-conditioned matrix functions***

## MATH: **Functional Approximation**

- nested approximate algebras and recursive preconditioning.
- ***ill-conditioned matrix functions***

## *N*-BODY SOLVERS: **Ecosystems & Stacks**

- *N*-Body Fock exchange (NoFX)
- *N*-Body Linear Algebra (SpAMM)
- others …

## InfoSci: ***N*-Body Learning**

- fast kernel summation, fast pairwise statistical problems, …
- learning the metric: fast approaches to semi definite programming.

## CompSci: **Generic Programming**

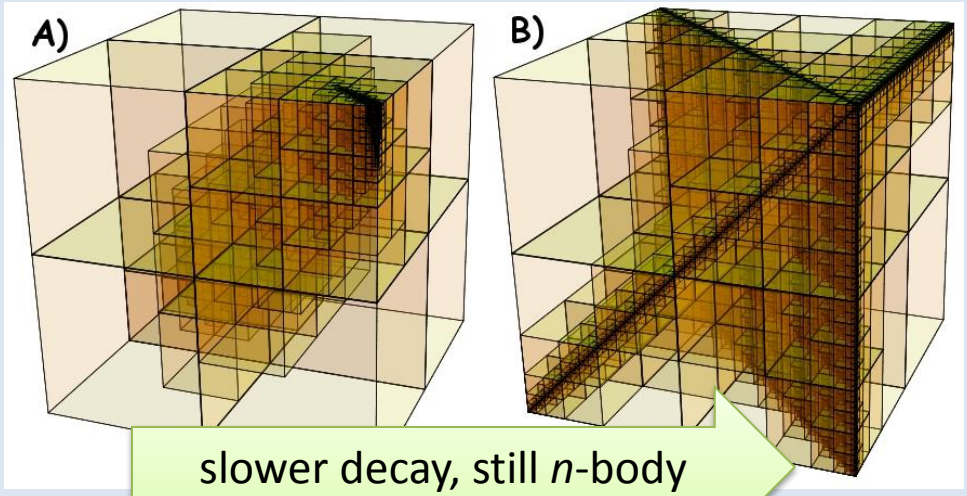
- functional programming, skeletons, recursive task parallelism, openmp 4.
- enterprise frameworks: scala/spark + neo/epiphany/phi.

# Matrix Multiplication as N-Body Solver

**Challacombe & Bock, arXiv:1011.3534**

**SpAMM** is a fast kernel for multiplication of matrices with ***decay & locality.*** Employs decorated quadtrees, rigorous sub-multiplicative norms, recursive occlusion and culling in the product space (an octree metric query on norms)**.**

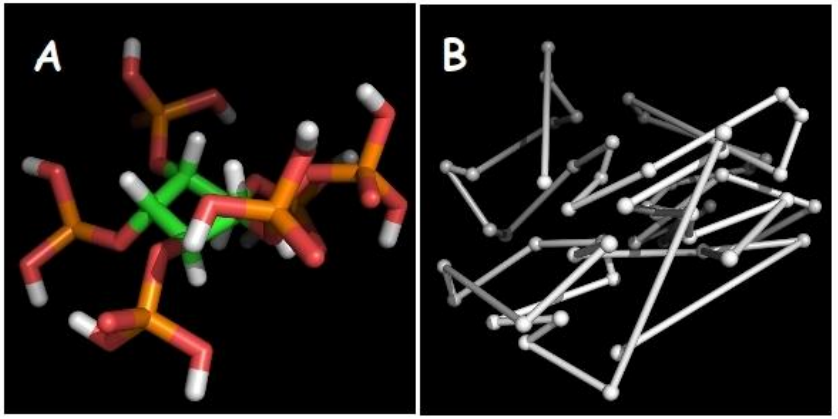
$$A^k \otimes B^k = \begin{cases} 0 & \text{if } \left\|A^k\right\| \left\|B^k\right\| < \tau, \\ A^k \cdot B^k & \text{elseif } k = k_{\max}, \\ \begin{pmatrix} A_{11}^{k+1} \otimes B_{11}^{k+1} + A_{12}^{k+1} \otimes B_{21}^{k+1} & A_{11}^{k+1} \otimes B_{12}^{k+1} + A_{12}^{k+1} \otimes B_{22}^{k+1} \\ A_{21}^{k+1} \otimes B_{11}^{k+1} + A_{22}^{k+1} \otimes B_{21}^{k+1} & A_{21}^{k+1} \otimes B_{21}^{k+1} + A_{22}^{k+1} \otimes B_{22}^{k+1} \end{pmatrix} & \text{else.} \end{cases}$$

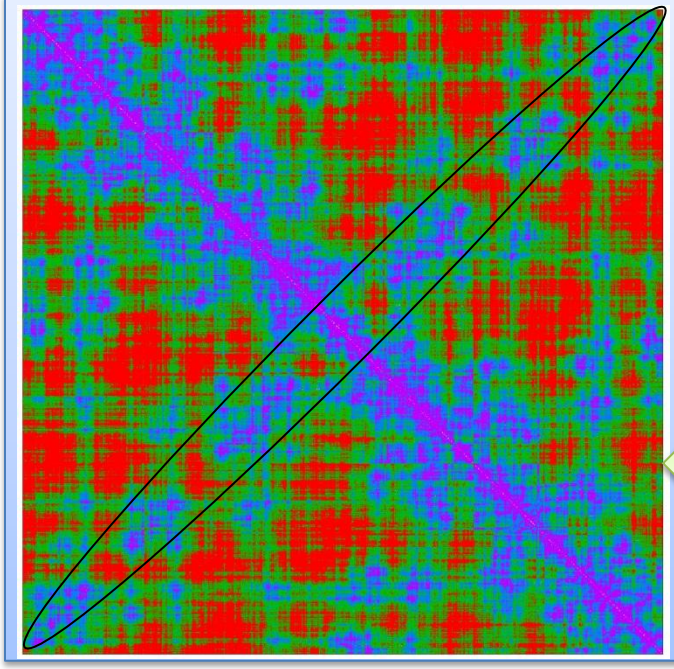



Fast occlusion and culling in the product space for matrices with:
A) exponential decay and
B) algebraic decay

# N-Body Methods Exploit Locality Heuristics

- Space-filling-curve heuristics maximize locality, clustering (A-B)
- With locality, database algorithms enable rapid distance, metric & overlap queries.
- Clustering enhances multilevel approximation, occlusion & *etc*



- Space-filling-curves (SFCs) map atoms close in space to an index where they are also close
- This SFC ordering naturally blocks & structures corresponding matrices with decay
- RB3LYP/6-31G** density matrix for $[H_2O]_{300}$. Purple is large, red is small.

# Newton Schulz for $S^{-1/2}$ (the Metric Problem)

- Problem of metric and gap ill-conditioning connected through Higham's identity (Higham '97): $\operatorname{sign}\left(\begin{bmatrix} 0 & \boldsymbol{s} \\ \boldsymbol{I} & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & \boldsymbol{s}^{1/2} \\ \boldsymbol{s}^{-1/2} & 0 \end{bmatrix}$
- A generic $NS$ with SpAMM approximation:

$$NS[\{\boldsymbol{x},\boldsymbol{z}\},\tau] := \textbf{while}\left(\,|tr\boldsymbol{x}-n|/n > \tau\,\right) \begin{bmatrix} \boldsymbol{x} \leftarrow \mathrm{map}_{NS}[\boldsymbol{x}] \\ \boldsymbol{z} \leftarrow \boldsymbol{z} \otimes_\tau \boldsymbol{x} \\ \boldsymbol{x} \leftarrow \boldsymbol{z} \otimes_\tau \boldsymbol{x} \otimes_\tau \boldsymbol{z} \end{bmatrix}$$
$$\textbf{return}\,\{\boldsymbol{x},\boldsymbol{z}\}$$

$$\boldsymbol{x} \to \boldsymbol{I} \text{ and } \boldsymbol{z} \to \boldsymbol{s}^{-1/2} \text{ with } \tau \to 0$$

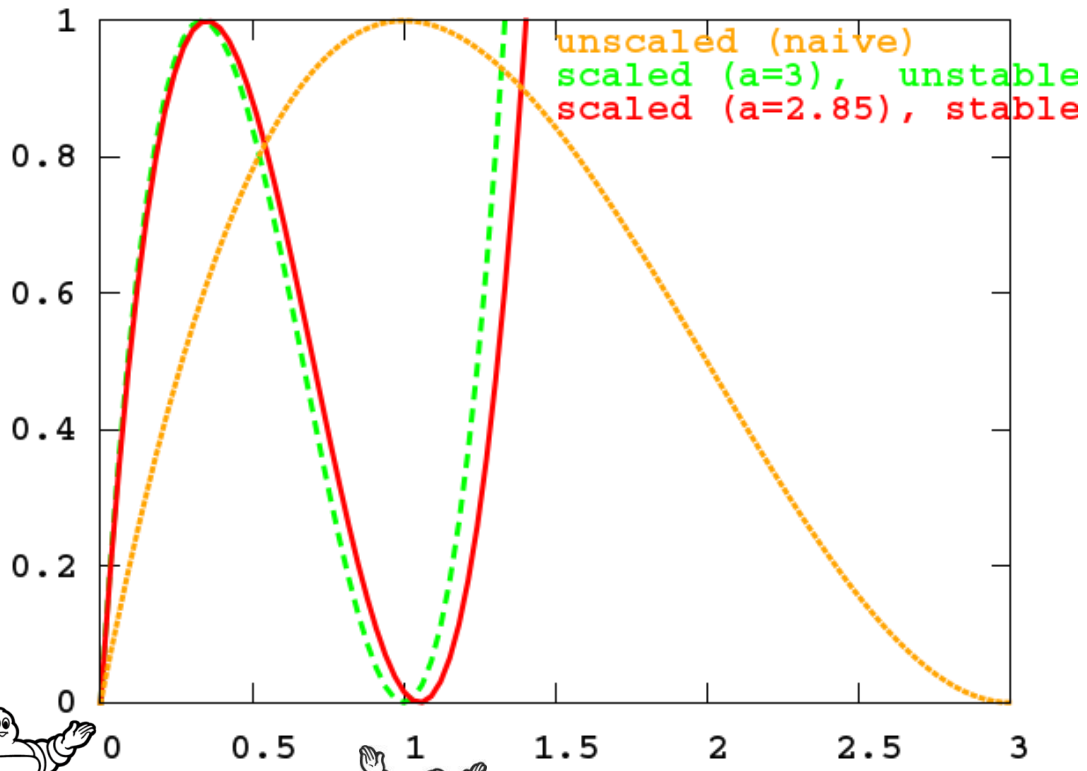
- Many variations in literature, including scaled versions, stabilized versions like "DB iteration" involving dual $\boldsymbol{y}$ & $\boldsymbol{z}$ channels, and so on. ***What is the fastest, most stable method under severe $\otimes_\tau$ approximation ( $\tau$ is large )?***

# Scaled Newton Schulz

- The naïve $NS$ map is: $\mathrm{map}_{NS}(x) \coloneqq \frac{1}{2}(3-x)$, corresponding to the logistic $x \leftarrow \frac{1}{2}(3-x) \cdot x \cdot \frac{1}{2}(3-x)$. Much recent work on the scaled $NS$ [*ie.* $\mathrm{map}_{NS}(\alpha\, x)$ ]. See for example Pan & Schreiber '91, Higham '97, Janzik et al '07, ***Chen & Chow '14***.

In the Chen & Chow scheme, scaling increases gradient at the origin, accelerating convergence of the minimum EV $x_0$

A full acceleration of $\alpha = 3$ isn't possible due to instability at $x = 1$. $\alpha \sim 2.85$ stabilizes.

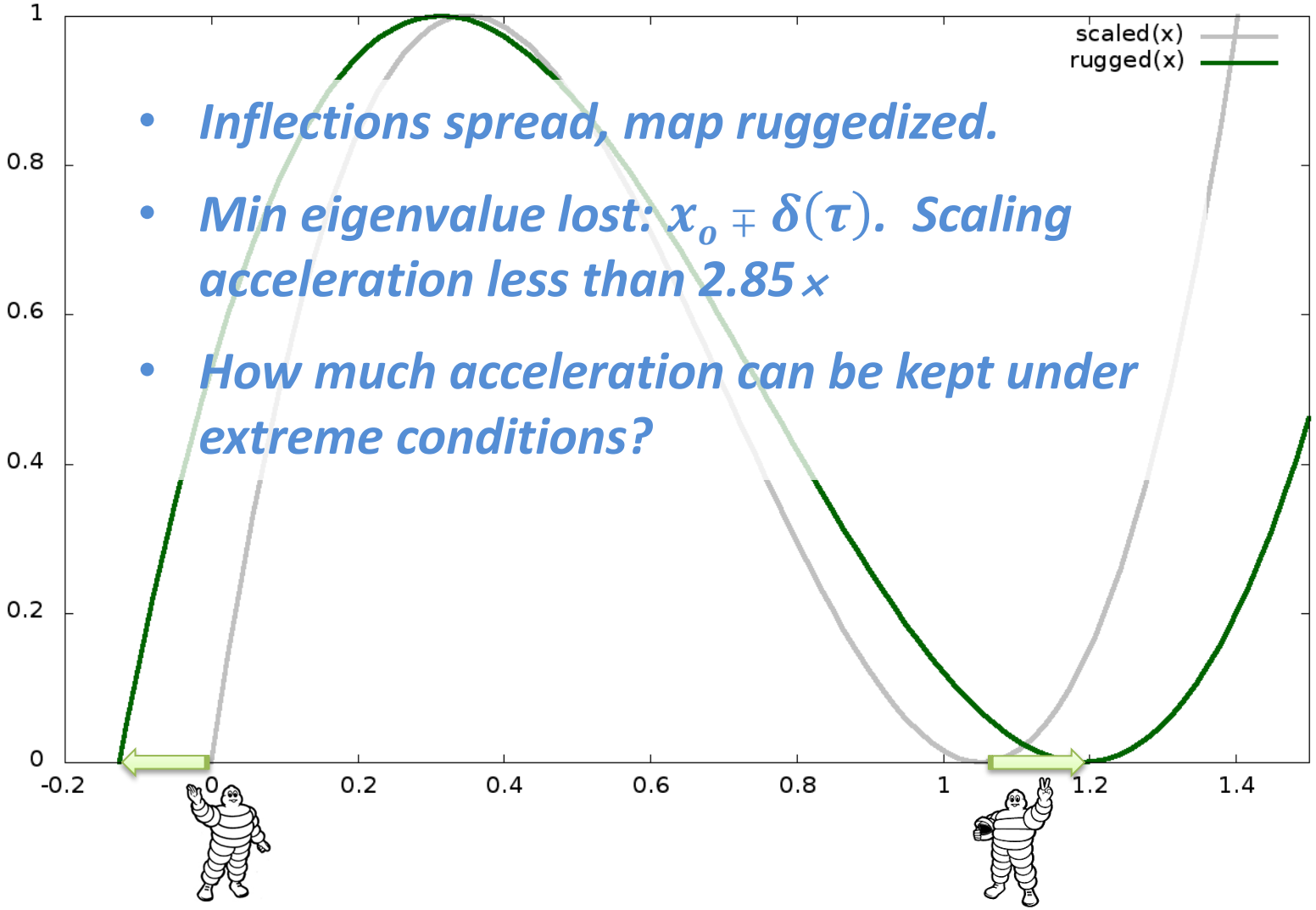



Convergence determined by gradient:
$g(x_0) = \boldsymbol{\alpha} \times \left(\frac{9}{4}\right)$.
Scaling attenuated, $\alpha \to 1$ as $x_0 \to 1$

Using $\alpha \sim 2.85$ instead of $\alpha = 3$ stabilizes map for double precision

# SpAMM Stabilized Scaled NS (I)

Ill-conditioning and SpAMM can bounce EVs out of bounds by $\mp\,\delta(\tau)$. ⇨ stabilize by spreading 0/1 inflections by $\sim\ \mp\,\delta$.




- ***Inflections spread, map ruggedized.***
- ***Min eigenvalue lost: $x_o \mp \delta(\tau)$. Scaling acceleration less than 2.85×***
- ***How much acceleration can be kept under extreme conditions?***

# SpAMM Stabilized Scaled NS (II)

- Tried many approaches to achieve stability under severe $\otimes_{\tau}$ approximation. Found use of the left transpose SpAMM, $\otimes_{LT,\tau}$ , to be most stable, especially for ill-conditioned matrices.

- Left-right stabilized SpAMM contraction:

$$x_{k+1} = \langle z^{t}{}_{k} | \, s \, | z_{k} \rangle$$

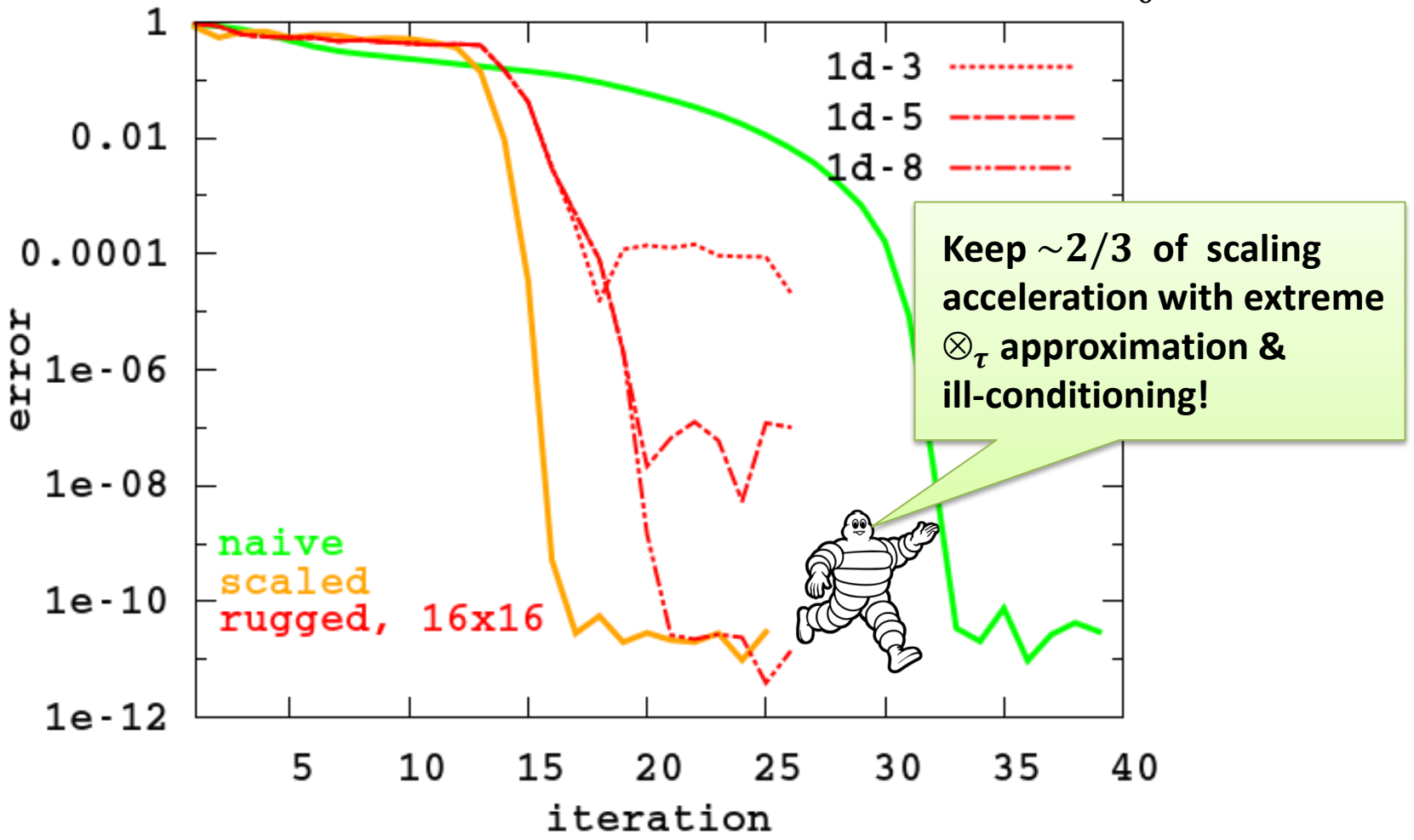
then right: $x \leftarrow a \otimes_{\tau} z$

left (T) first: $a \leftarrow z \otimes_{LT,\tau \times 10^{-2}} s$

- The full, asymmetric case with left transpose is more forgiving with respect to approximation, $\otimes_{\tau}$ , than "DB iteration" and symmetrized versions (preliminary).

# SpAMM Stabilized Scaled NS (III)

- Extreme Ill-Conditioning: $\boldsymbol{\kappa}(s) = 10^{11}$, (3,3) x 8 nanotube
- Stabilized, left transpose SpAMM, stabilized map, & scaling switched by error heuristics (don't compute $x_0$).



# Recursive Preconditioning: The SpAMM Sandwich

- Nested Newton Shulz functionals, $NS[\cdot]$ , with increasing SpAMM resolution, $\tau_m < \tau_{m-1} < \cdots < \tau_0$ .
- Use error $\varepsilon$ rather than $x_0$ (min EV) to manage transition between scaled & unscaled $NS$. Use "DB iteration" beyond 0th iteration.

nested $NS$ functionals:

$$\{\boldsymbol{x}_m, \boldsymbol{z}_m\} = NS[\, NS[\, \ldots\, NS[\, \{\boldsymbol{x}_0, \boldsymbol{z}_0\}\,,\ \tau_0\,]\ \ldots, \tau_{m-1}\,],\ \tau_m\,]$$

resolution of the identity:

$$\boldsymbol{x}_m = \langle \boldsymbol{z}^t{}_m | \langle \boldsymbol{z}^t{}_{m-1} | \cdots \langle \boldsymbol{z}^t{}_0 |\, \boldsymbol{s}\, | \boldsymbol{z}_0 \rangle \cdots | \boldsymbol{z}_{m-1} \rangle | \boldsymbol{z}_m \rangle \rightarrow \boldsymbol{I}$$

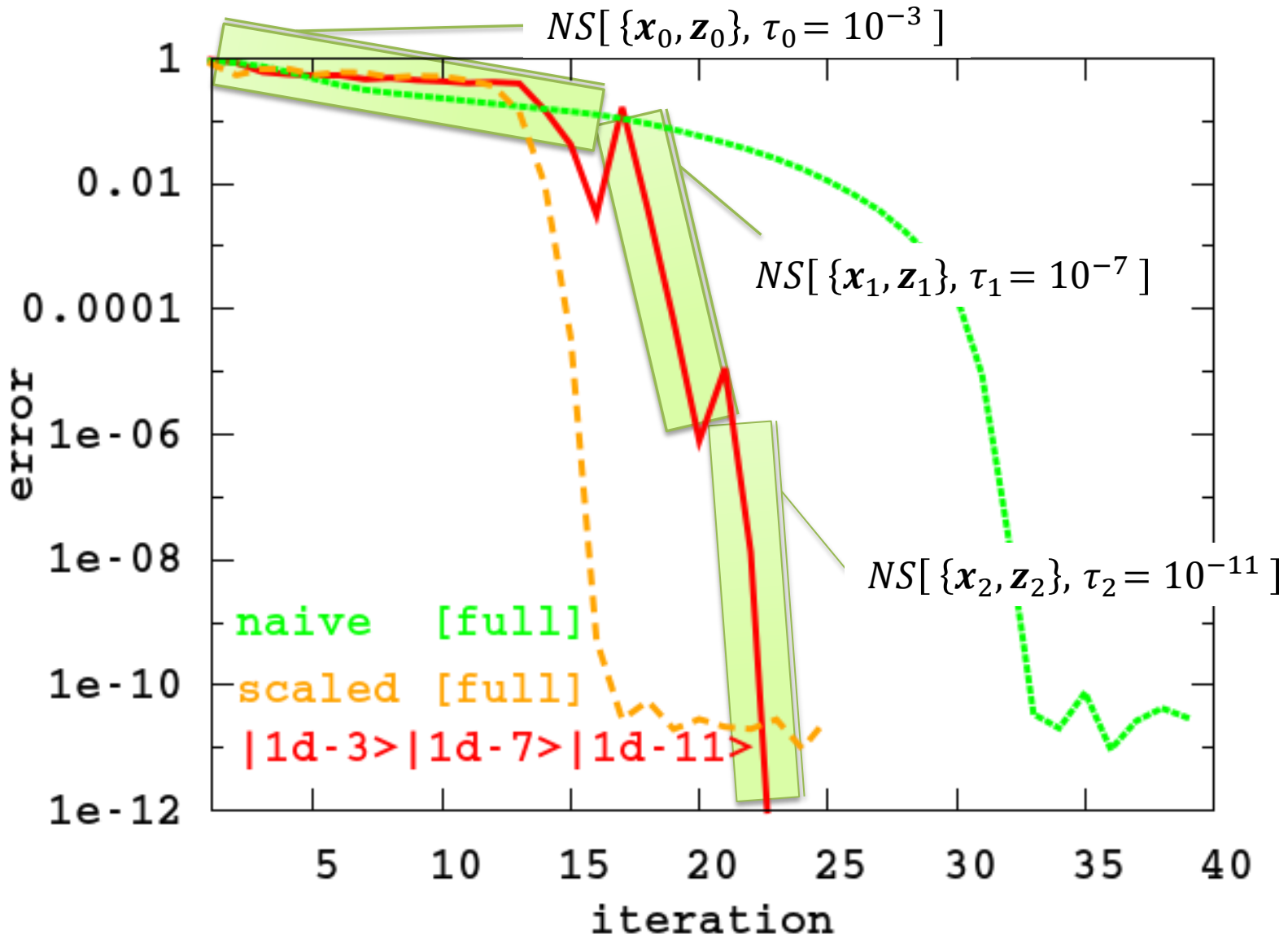
sandwich factors:

$$|\boldsymbol{z}\rangle = |\boldsymbol{z}_0\rangle \cdots |\boldsymbol{z}_{m-1}\rangle | \boldsymbol{z}_m \rangle \rightarrow |\boldsymbol{s}^{-1/2}\rangle$$

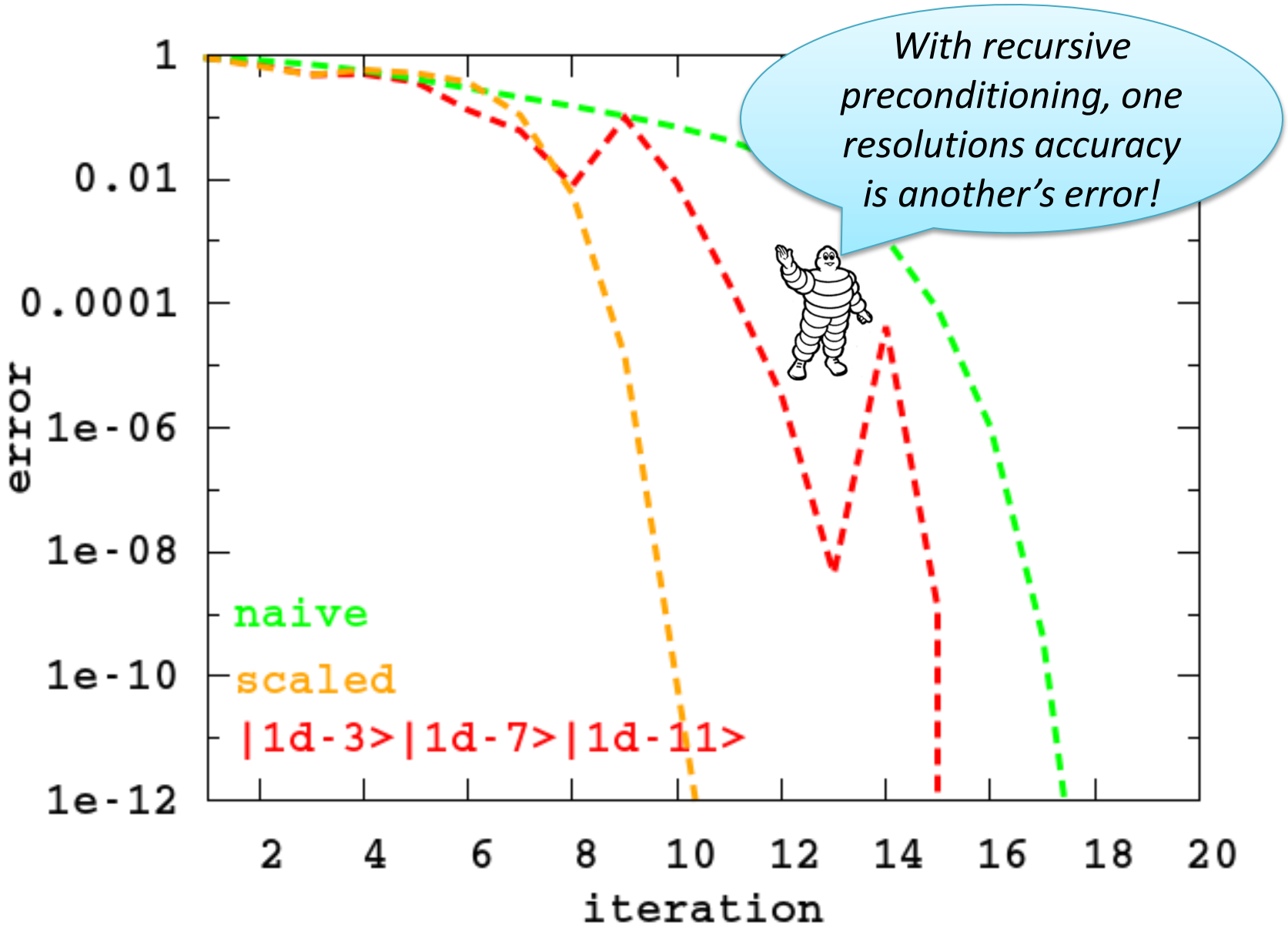
for each slice, the error:

$$\varepsilon(\tau_j) = |\, \mathrm{tr}\, \boldsymbol{x}_j - n\,|/n$$

# Ill-Conditioning: $\kappa(s) = 10^{11}$, (3,3)x8 nanotube



# Ill-Conditioning: $\kappa(s)\sim 10^5$, $[H_2O]_{70}$, TZV2P



# Inverse Factors of an Ill-conditioned S

Recall, that for large, triple-zeta + polarization basis sets:
***"sufficiently sparse only for … ~10 000 water molecules"***

For 6-311G** basis, decay in $\mathbf{S}^{-1/2}$ is exponential but very, very slow

Naïve Newton Schulz iterations [J. Chem. Phys. 126 (2007) p. 124104]. Target error 1e-8 in matrix elements.

Substantiates BCSR results of Hutter group [~$10^4$ waters]

Unoptimized SpAMM 4x4 vs max optimized MKL/DSYEV crossover @ 250 waters. ***O(N) & dense!***