Algorithms for domain wall fermions

Peter Boyle
Brookhaven National Laboratory
RBC-UKQCD collaboration

* Work carried out under SciDAC-5 HEP LAB-22-2580

Work with Shuhei Yamamoto, Patrick Oare, Chulwoo Jung and Christoph Lehner

Outline

Warning: this talk is an exercise in false advertising and doesn't reflect the abstract

- Originally relatively content free, speculative and visualization heavy
  - discussion of gauge and fermion roles in topology change and sampling

- Became more multigrid heavy with concrete progress after I submitted the abstract

Part 1: New chiral fermion multigrid solver with potential use in HMC
Part 2: topology change in HMC & the index theorem

- **True advertising:** see other talks by
  - Shuhei Yamamoto - FTHMC and accelerating sampling of long distance physics
  - Patrick Oare — Krylov-Schur non-hermitian eigensolver: complex eigenvalues of Dw instead of Hw
  - Chulwoo Jung — Harmonic Krylov-Schur & tracking transition of complex paired modes of Dw to real chirality
  - Christoph Lehner - RBC-UKQCD ensemble generation along with muon g-2

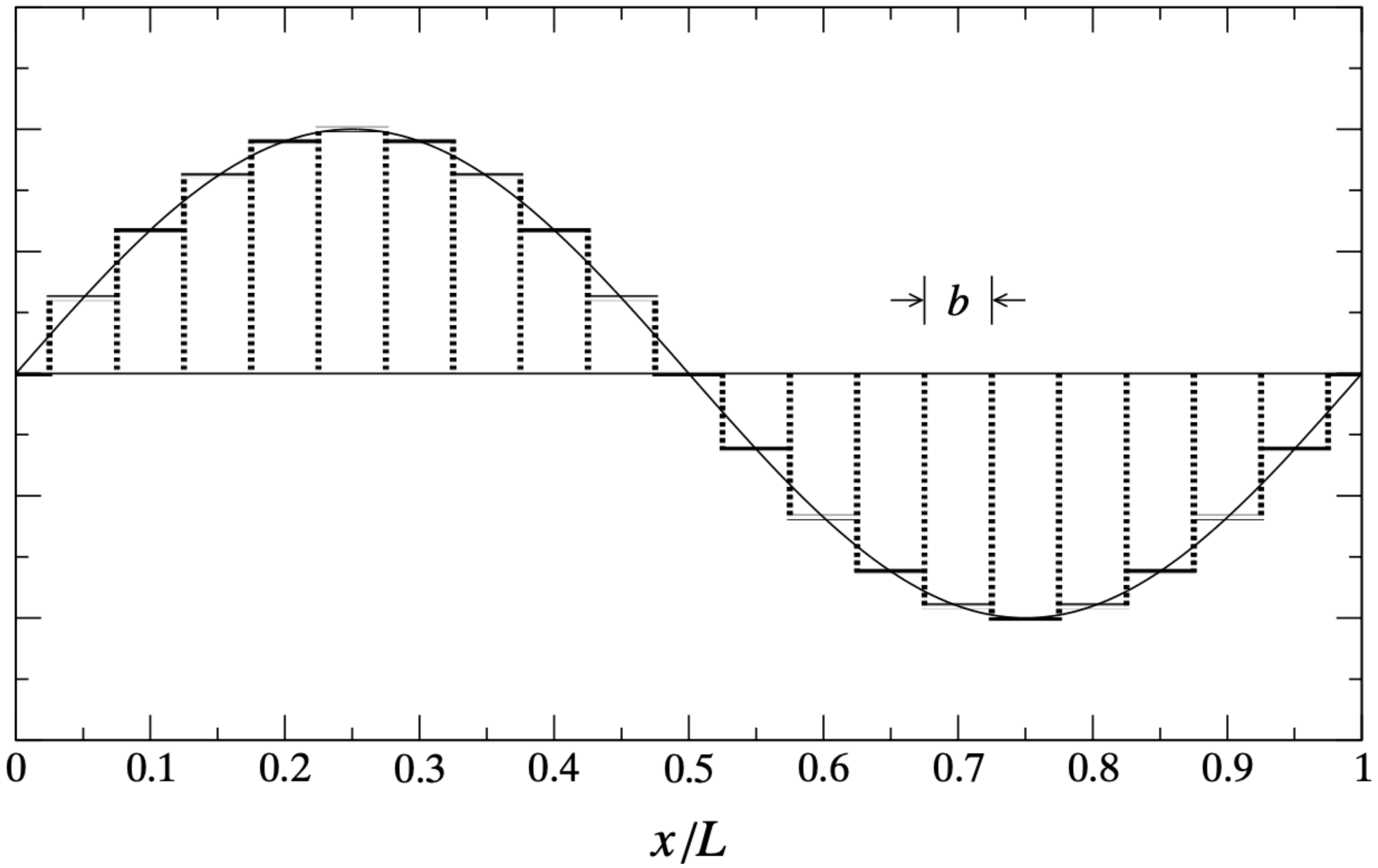
Part 1: New chiral fermion multigrid solver with potential use in HMC

Boyle, Oare

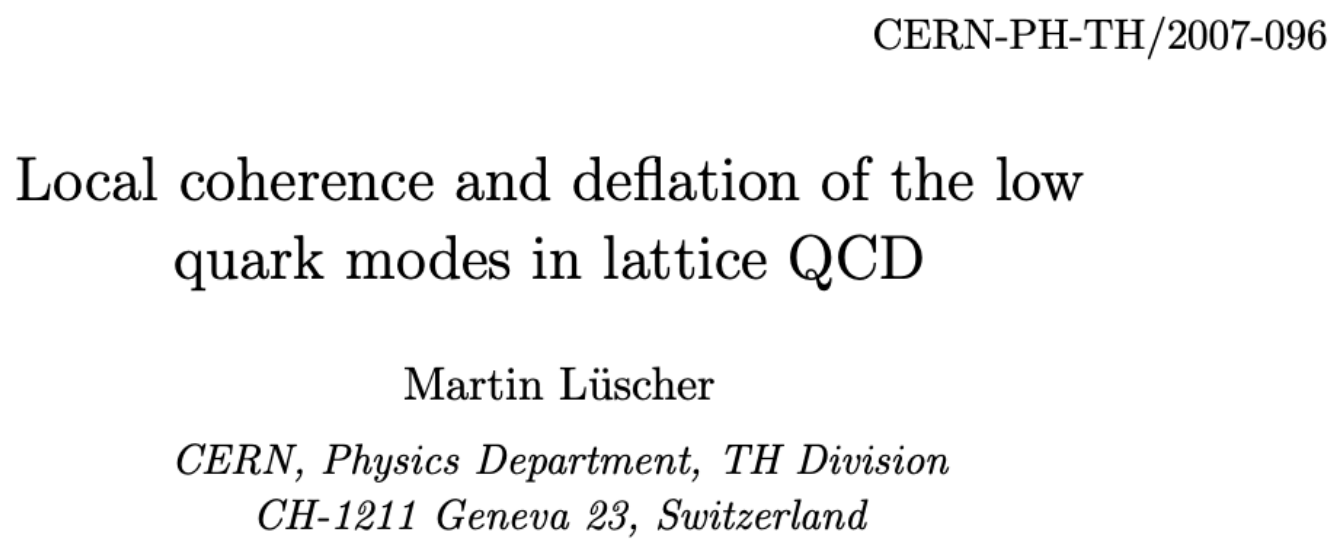
Part 1: Multigrid algorithms. (PB, P. Oare)

Multigrid Dirac solvers:

- *Learn near null space of Dirac matrix* (~60 vectors): non-trivial in gauge theory
- Break vectors into local wavelet basis chunks
- Calculate a representation of Dirac operator within this critical subspace
- Use as an approximate near-null space multigrid preconditioner



(Figure ``borrowed'' from Martin Lüscher's paper)



CERN-PH-TH/2007-096

Local coherence and deflation of the low quark modes in lattice QCD

Martin Lüscher

CERN, Physics Department, TH Division
CH-1211 Geneva 23, Switzerland

- Recent progress on multiple right-hand side multigrid for DWF in Grid: arXiv:2409.03904
  - Use (introduce?) preconditioned BlockCGrQ
    - Made appropriate for multigrid by using **stationary** coarse grid solver and smoother
  - Can use GPU matrix multiplication hardware to gain extraordinary performance

Preconditioned BlockCGrQ

1. $z_0 = B - \mathcal{H}X_0$
2. $Z_0 = M^{-1} z_0$
3. $C_0^\dagger C_0 = Z_0^\dagger \cdot z_0 = \langle Z_0 | Z_0 \rangle_M$
4. $q_0 = z_0 C_0^{-1} \quad ; \quad Q_0 = Z_0 C_0^{-1}$
5. $D_0 = Q_0$
6. for iteration $k \in 0, 1, 2 \ldots$
7. $\quad z_k = \mathcal{H} D_k$
8. $\quad Z_k = M^{-1} z_k$
9. $\quad N_k = [D_k^\dagger \cdot z_k]^{-1} = \langle D_k | Z_k \rangle_M^{-1}$
10. $\quad X_{k+1} = X_k + D_k N_k$
11. $\quad t_k = q_k - z_k N_k \quad ; \quad T_k = Q_k - Z_k N_k$
12. $\quad S_k^\dagger S_{k+1} = T_k^\dagger \cdot t_k = \langle T | T \rangle_M$
13. $\quad q_{k+1} = t_k S_k^{-1} \quad ; \quad Q_{k+1} = T_k S_k^{-1}$
14. $\quad D_{k+1} = Q_{k+1} + D_k S_k^\dagger$
15. $\quad C_{k+1} = S_k C_k$

2409.03904v1 [hep-lat] 5 Sep 2024

Multiple right hand side multigrid for domain wall fermions with a multigrid preconditioned block conjugate gradient algorithm.

Peter Boyle

[a] Physics Department, Brookhaven National Laboratory, Upton, 11777, NY, USA

Abstract

We introduce a class of efficient multiple right-hand side multigrid algorithm for domain wall fermions. The simultaneous solution for a modest number of right hand sides concurrently allows for a significant reduction in the time spent solving the coarse grid operator in a multigrid preconditioner. We use a preconditioned block conjuate gradient with a multigrid preconditioner, giving additional algorithmic benefit from the multiple right hand sides. There is also a very significant additional to computation rate benefit to multiple right hand sides. This both increases the arithmetic intensity in the coarse space and increases the amount of work being performed in each subroutine call, leading to excellent performance on modern GPU architectures. Further, the software implementation makes use of vendor linear algebra routines (batched GEMM) that can make use of high throughput tensor hardware on recent Nvidia, AMD and Intel GPUs. The cost of the coarse space is made sub-dominant in this algorithm, and benchmarks from the Frontier supercomputer system show up to a factor of twenty speed up over the standard red-black preconditioned conjugate gradient algorithm on a large system with physical quark masses.

Keywords: Multigrid, Linear Solver, Lattice QCD, Algorithms, BlockCG

- > 20x performance gain over red-black CG
  - Eliminates V^2 scaling of deflation approach: 16 times bigger volume requires 256 more cost and storage !
  - (Coarse space is exact deflated with cost quadratic in coarse volume - not a problem in my lifetime)

- Coarsens the 4-hop operator $\mathscr{H} = (M_{ee} - M_{eo}M_{oo}^{-1}M_{oe})^\dagger(M_{ee} - M_{eo}M_{oo}^{-1}M_{oe})$
  - "Only" 81 point coarse stencil if blocking is 4^4 or more
    - $\hat{0}, (\pm\hat{x}), (\pm\hat{x} \pm \hat{y}), (\pm\hat{x} \pm \hat{y} \pm \hat{z}), (\pm\hat{x} \pm \hat{y} \pm \hat{z} \pm \hat{t})$ + rotations
    - No hop by 2 or more in any direction due to minimum block constraint

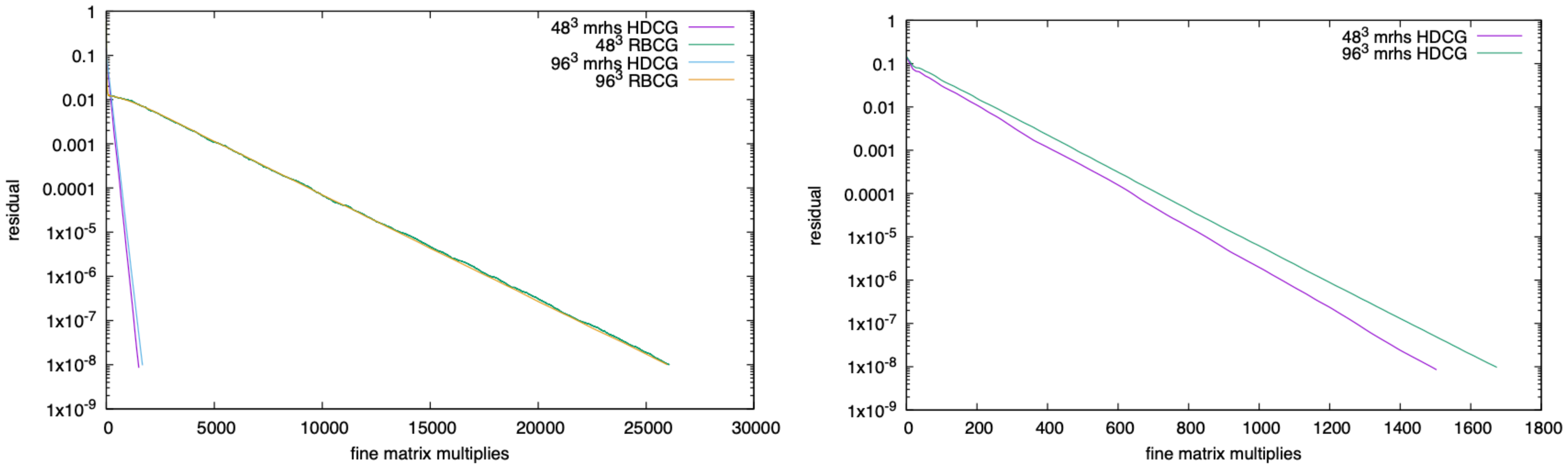


Figure 7: (Left) Convergence of mrhs-HDCG and red-black preconditioned Conjugate Gradient on sample $48^3 \times 96$ and $96^3 \times 192$ configurations. (Right) Zoomed comparison between the two volumes on the HDCG convergence history.

MultiRHS coarse operator implemented using Batched BLAS calls

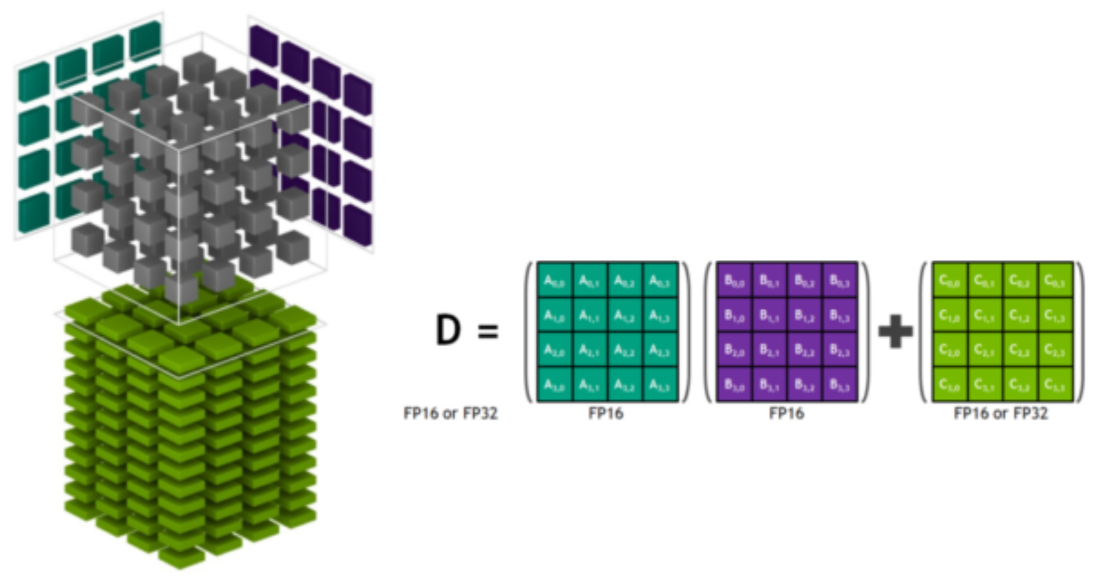
Coarse operations implemented using “Batched GEMM” APIs in HIP, CUDA and OneAPI (and Eigen for CPU)

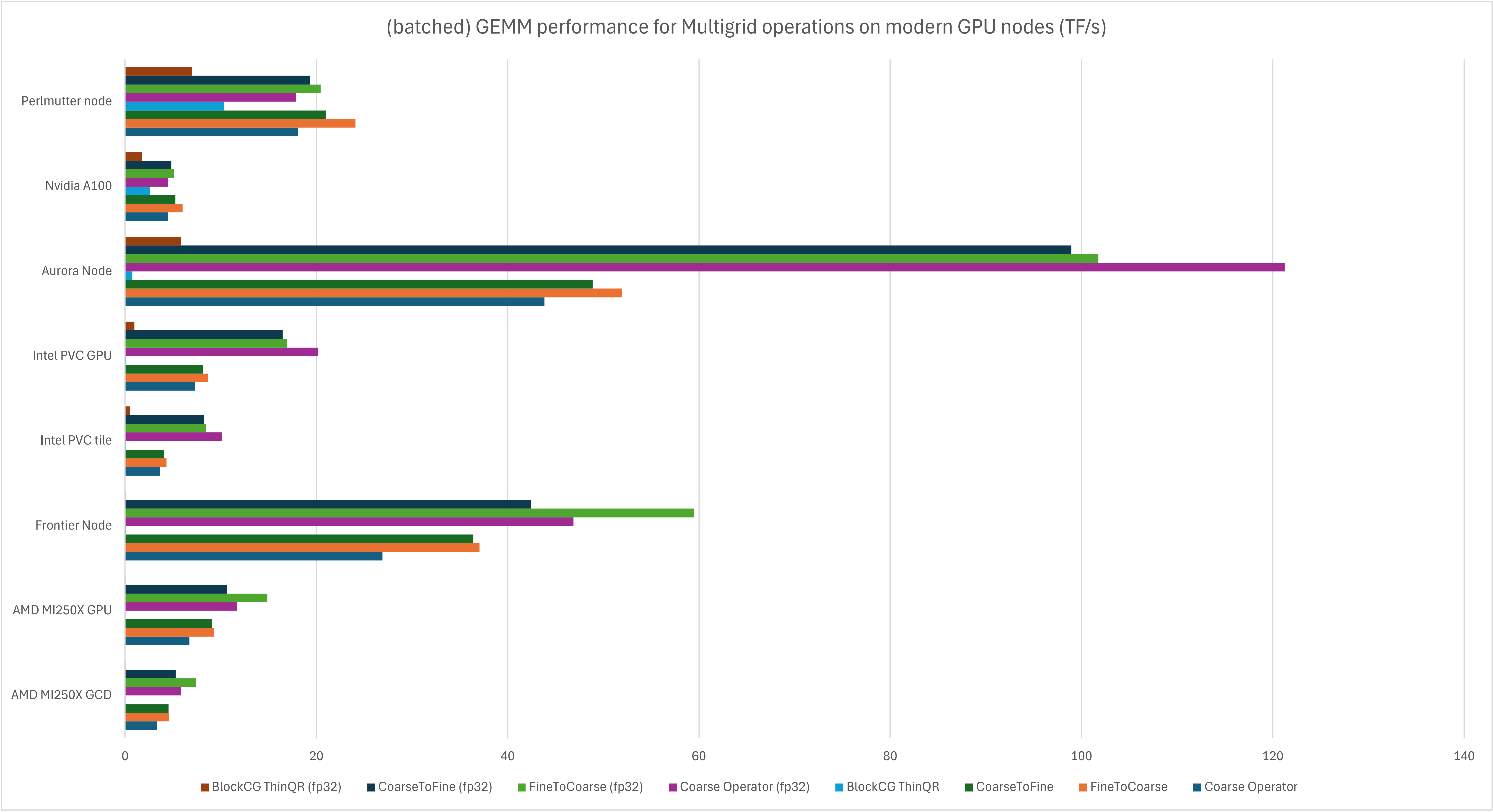
Using machine learning tensor hardware in GPUs to speed up multi-right-hand side solvers in QCD

Grid / Grid / algorithms / **blas** /

paboyle Batched blas, but not working yet on OneAPI

Name
..
BatchedBlas.cc
BatchedBlas.h





Setup cost of HDCG

Why am I not happy with mrhs-HDCG ?

Setup cost

- 81 point stencil + 60 vectors : 4860 matrix multiplies to calculate coarse operator
- Lowest eigenvectors work best as setup: 20x gain
- Various filtered noise approaches work with 10-15x gain

MultiRHS

- HDCG (2014) was efficient on CPUs, e.g. BG/Q
- Address Amdahl on coarse space on GPU’s by going to multiRHS and using BLAS

Both affect use in HMC: single solve on a NEW gauge field
Seek an algorithm based on more compact fine operator

This work: Directly coarsen 2-hop fine operator

$$\mathcal{F}(m_{\mathrm{PV}}, m_l) = D^\dagger_{\mathrm{mobius}}(m_{\mathrm{PV}}) D_{\mathrm{mobius}}(m_l)$$

- $\mathcal{F}$ considered by arXiv:2004.07732 (Brower, Clark, Howarth, Weinberg)
  - 2D U(1), direct coarsen D_W, make a 5D Mobius of the coarse Wilson in preconditioner
- Attempt in 4D QCD arXiv:2103.05034 (Boyle, Yamaguchi) unsuccessful due to 5D coarse cost
- This work: block entire 5th dimension and coarsen 2-hop operator, 33 point stencil
  - $\hat{0}, (\pm\hat{x}), (\pm\hat{x} \pm \hat{y})$ + rotations
- 1980 matrix multiplies to re-evaluate coarse operator

Taken from arXiv:2004.07732

- Naively evade the “half plane condition”
- GCR converges in 33,000 iterations
- Comparable to RBCG 26500 iterations

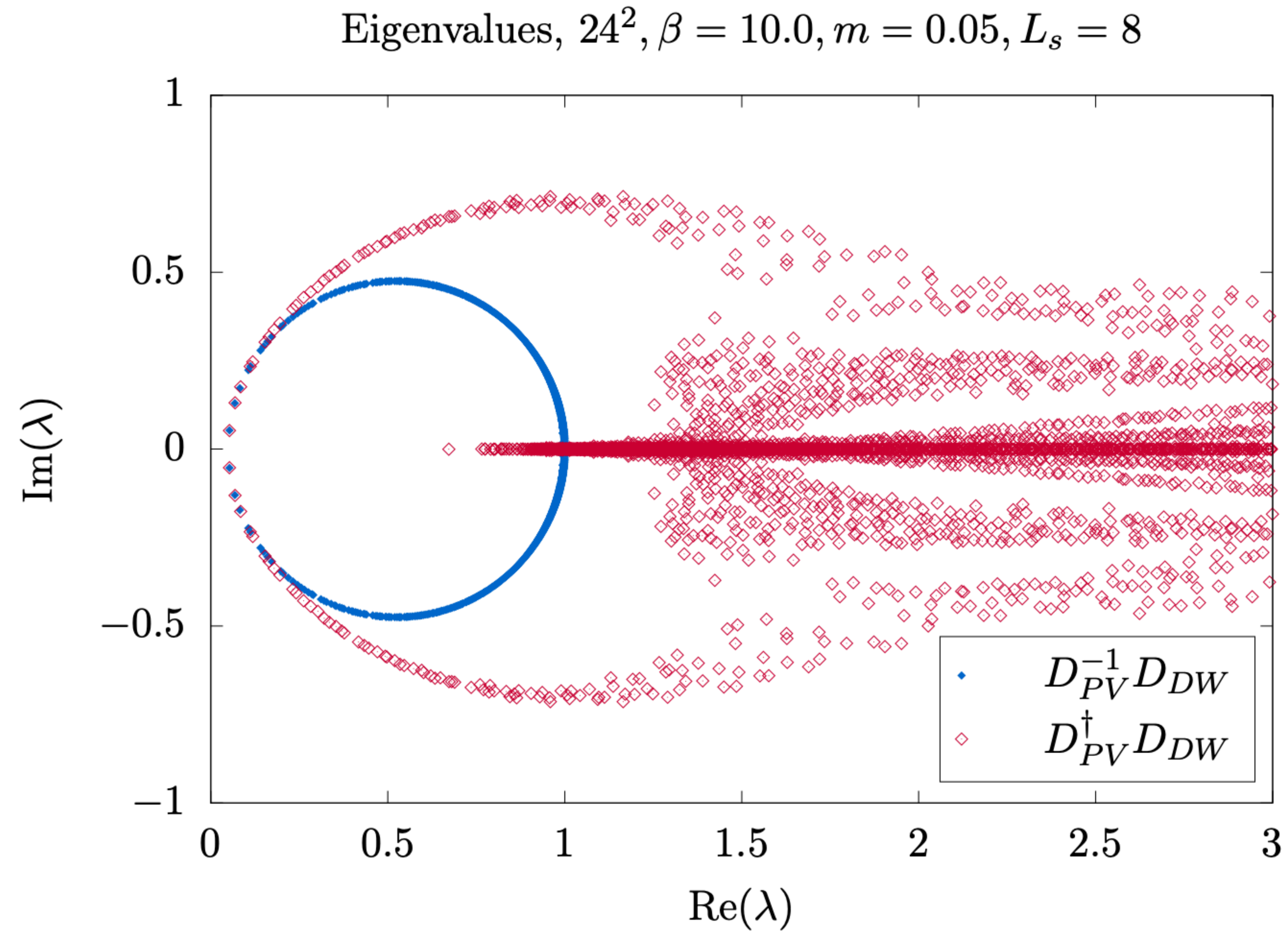


Figure 2.5: The spectrum of our target multigrid operator, $D_{PV}^{\dagger} D_{DW}$, compared with the effective overlap spectrum, $D_{PV}^{-1} D_{DW}$. For clarity of presentation we truncate the x-axis; the spectrum of $D_{PV}^{\dagger} D_{DW}$ extends out to Re(λ)$\approx$ 25.

Failure mechanisms of Abnormality

- Problem : $\mathcal{F}(m_{\rm PV}, m_l) = D^\dagger_{\rm mobius}(m_{\rm PV}) D_{\rm mobius}(m_l)$ is non-normal
- Eigenvectors are NOT mutually orthogonal

Field of values:

contained within spectrum if hermitian
overlap between EV's matters if non-hermitian
can spill into negative half plane, even if spectrum is in right half!

$$A v_i = \lambda_i v_i$$

$$x = c_i v_i \qquad \sum_i |c_i|^2 = 1$$

$$x^\dagger A x = c_i^* c_j \lambda_j v_i^\dagger v_j = \left\{ \begin{pmatrix} |c_i|^2 \lambda_i & ; & \text{herm} \\ c_i^* c_j \lambda_j v_i^\dagger v_j & ; & \text{nonherm} \end{pmatrix} \right.$$

GCR:

$$r' = r - \alpha A r$$

minimise residual via $\alpha = \dfrac{\langle r | A | r \rangle}{|Ar|^2}$

$$\Rightarrow \frac{|r'|^2}{|r|^2} = 1 - \frac{|\langle r | A | r \rangle|^2}{|r|^2 |Ar|^2}$$

Stagnation is possible if FoV contains 0 !

PVdagM is abnormal

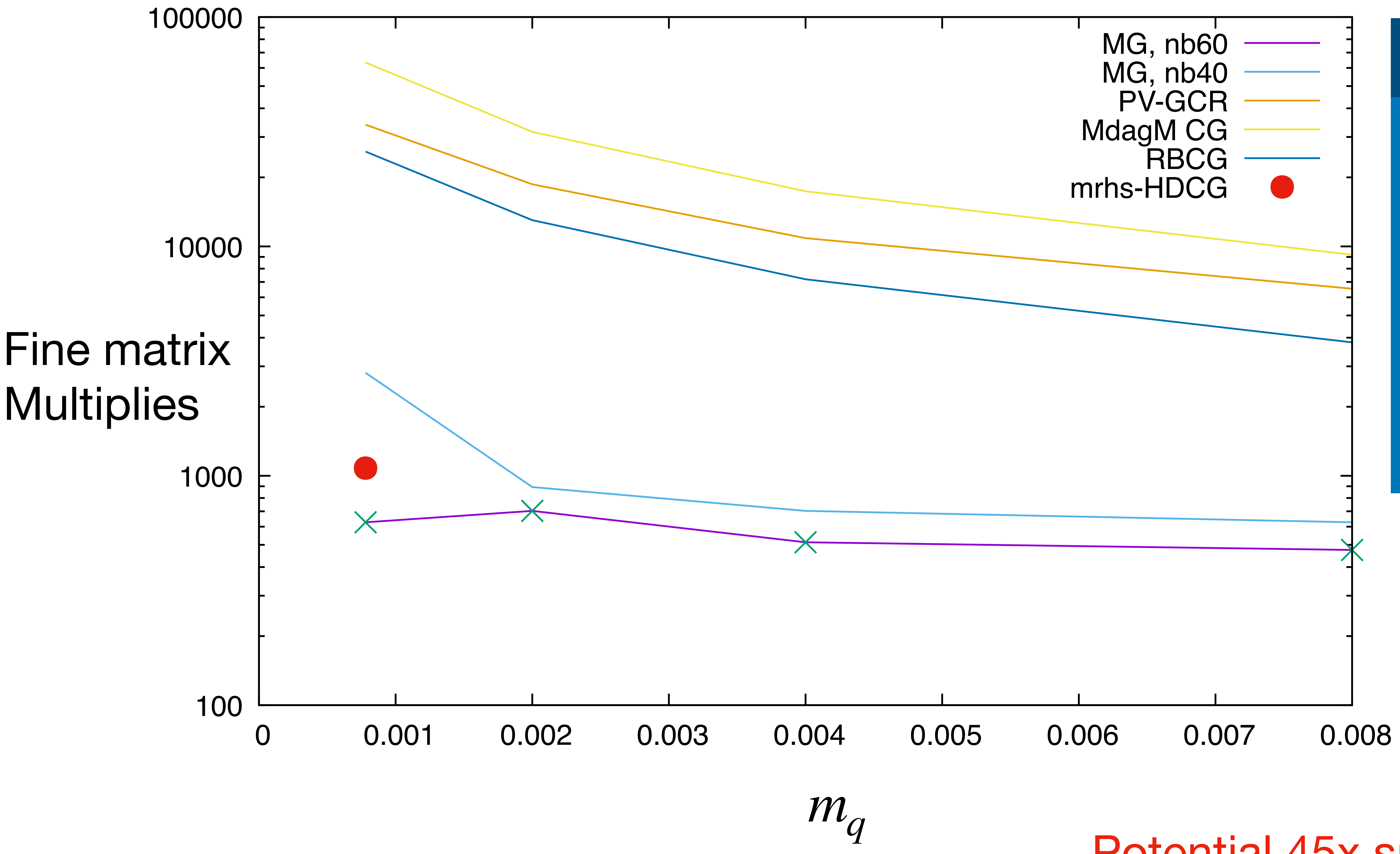
- Claude Fable & Opus clarified pathology & half plane condition correctly
- Eisenstat, Elman, Schultz GCR paper
  - Constraint on the Hermitian projection $M = (A^\dagger + A)/2$
  - Convergence bound requires positivity of real eigenvalues of M
    - λ_min(M) is exactly the leftmost real extent of the field of values
  - Convergence rate depends also on singular values (~eigenvalues of $A^\dagger A$)

THEOREM 3.3. *If* $\{r_i\}$ *is the sequence of residuals generated by* GCR, *then*

$$\|r_i\|_2 \leq \min_{q_i \in P_i} \|q_i(A)\|_2 \|r_0\|_2 \leq \left[1 - \frac{\lambda_{\min}(M)^2}{\lambda_{\max}(A^T A)}\right]^{i/2} \|r_0\|_2. \tag{3.3}$$

- Oft repeated “CGNR squares the condition number, so is dumb” is superficial
  - complex spectrum has phases so $\kappa = \dfrac{\lambda_{\max}}{\lambda_{\min}}$ has phases - CG bound doesn’t apply

Smaller block size 2^4: better coarse space on $48^3 \times 96$ mobius physical ensemble



Algo	Fine MatMul
HDCG	1100
PVdagM-MG-GCR	578
RBCG	26500
CGNR	66000
PVdagM GCR	33000

Potential 45x speed up(!)

Problem: coarse space is far too expensive

- 1100 iterations to converge *coarse* to 3e-2

  90% of runtime is in coarse solver !

  - Rapidly tried many options with Claude code
    - Squaring coarse operator and deflating (worse! too ill conditioned) ❌
    - SVD init guess deflation ❌
    - SVD deflation as a preconditioner ❌
    - 3 level, 4 level, 5 level recursive ❌
    - MultiRHS coarse solve ✅

- **Significant finding:**
  - fine null vectors make good, free coarse null vectors, by simple direct promotion
  - significant implications for HMC & reduced setup cost

Why not SVD deflation?

For staggered (mass + antihermitian) SVD deflation is used (BU/Nvidia)

$$D_s = V(m + i\Lambda)V^\dagger$$

- Positive field of values
- U=V up to a phase ⇒ stays deflated up to rounding!

General abnormal case: SVD of fine matrix $\mathscr{F} = U\Sigma V^\dagger$ has $U \neq V$

Init guess for k-singular vectors: $x_k^{\text{guess}} = V_k \Sigma_k^{-1} U_k^\dagger b$

$$r_0 = b - \mathscr{F} x_k^{\text{guess}} = (1 - U_k U_k^\dagger) b \ \Rightarrow r_0 \perp U_k$$

But deflation does not survive next step:

$$\mathscr{F} r_0 = U\Sigma V^\dagger r_0 \Rightarrow u_j^\dagger \mathscr{F} r_0 = \sigma_j v_j^\dagger r_0$$

The modes we orthogonalised against leak back in when the left and right singular spaces do not align
Demonstrated mutual local coherence of left/right singular vectors and right eigenspace saves us

multiRHS PVdagM MG-GCR (3 level)

36 nodes of Frontier, $48^3 \times 96$, $m_l = 0.00078$

Block by 2^4 and $2^2 \times 3^2$

	Time per RHS
RBCG	363s
PVdagM	136s
Mrs PVdagM	32s

12x gain over RBCG per RHS

Coarse space again accelerated by multiRHS

Good result, but similar performance to mrhs-HDCG

We lack equivalent of Lanczos deflation of coarse operator

Measuring our Abnormality

Measured the lowest eigenvalue of $(\mathscr{F} + \mathscr{F}^\dagger)$ in fine, coarse and coarse coarse space

- The ESS bound is broken for $\mathscr{F}(m_{\text{adj}}, m_l)$ for all $m_{\text{adj}} \geq 0.03$
- Lowest eigenvalue was -0.06 in case of $m_{\text{adj}} = 1$
  - Introduce a shift in the coarse and fine level smoothers

Evading the consequences of our Abnormality

For a normal matrix eigenvectors and singular vectors are the same up to a phase

- Null space filter projected on near null right eigenvector space

Performed coarse operator SVD (Lanczos on $D_c^\dagger D_c$)

- Right AND left singular vectors only 6% complete in this eigenvector basis

But…

- Local coherence completeness was 98% !
  - ***=> extra super coarse level, handle as dense non-iterative : plays role of deflation***

Likely resolution - breakthrough hot off press (i.e. yesterday!)

- Perform two levels of coarsening

	Global	Local	Blocking
Fine	48.48.48.96.24	16.8.12.24	-
Coarse	24.24.24.48.1	8.4.6.12	2.2.2.2.24
CoarseCoarse	3.6.8.8	1.1.2.2	8.4.3.6

- Max out the second coarsening
- Minimize coarse smoother count to 2
- Solve coarse coarse to high precision
  - **578 fine grid PVdagM applications**
  - **153 coarse grid applications per solve, 51 coarse outer iterations**

Coarse coarse is small enough 70k x 70k it can be inverted in seconds
applied as a dense matrix in ms timescale.
(Estimates, but Claude Opus and I agree!) **Plays even better with multiRHS !**

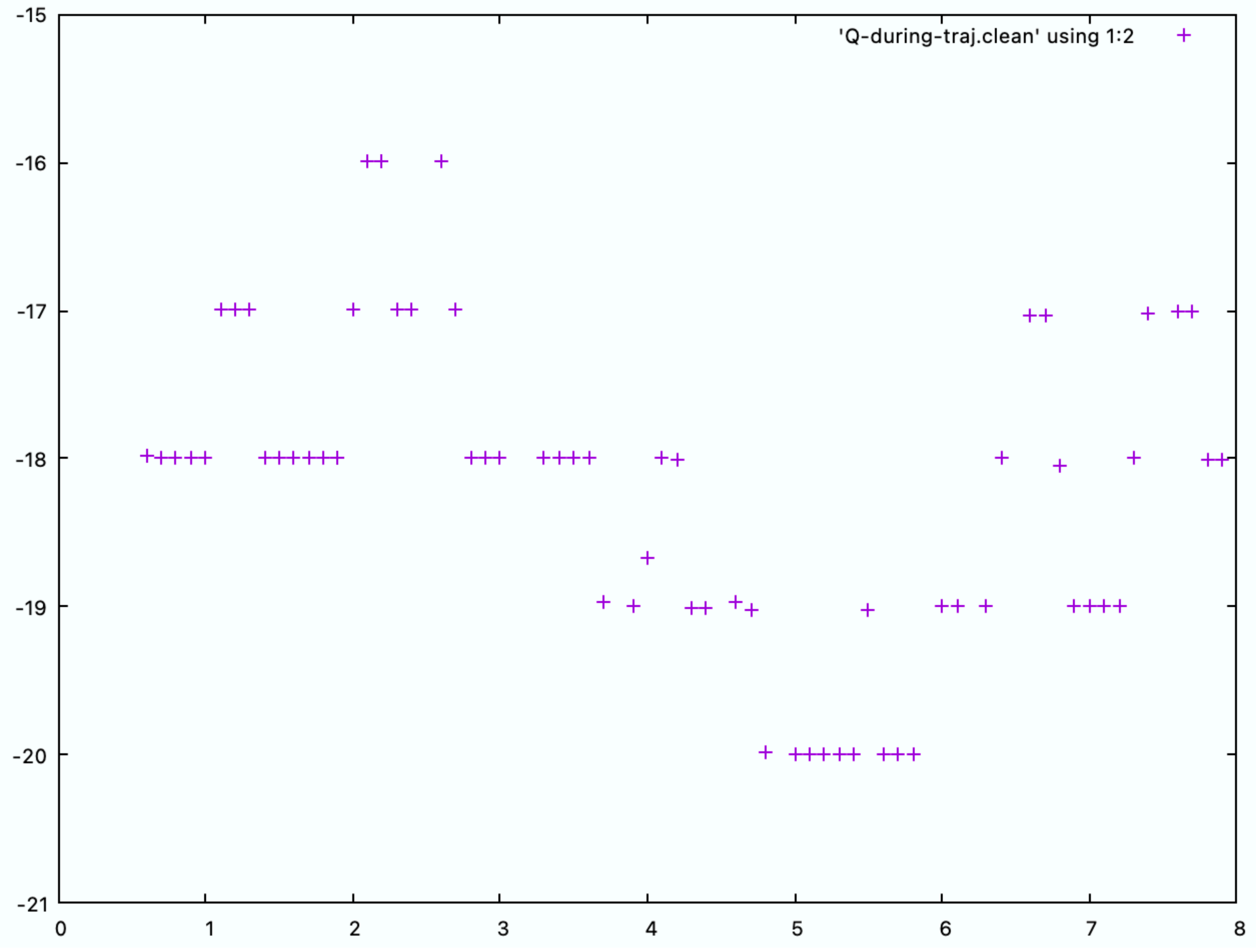
Natural, exact, non-iterative top level may play same role as deflation in herm case

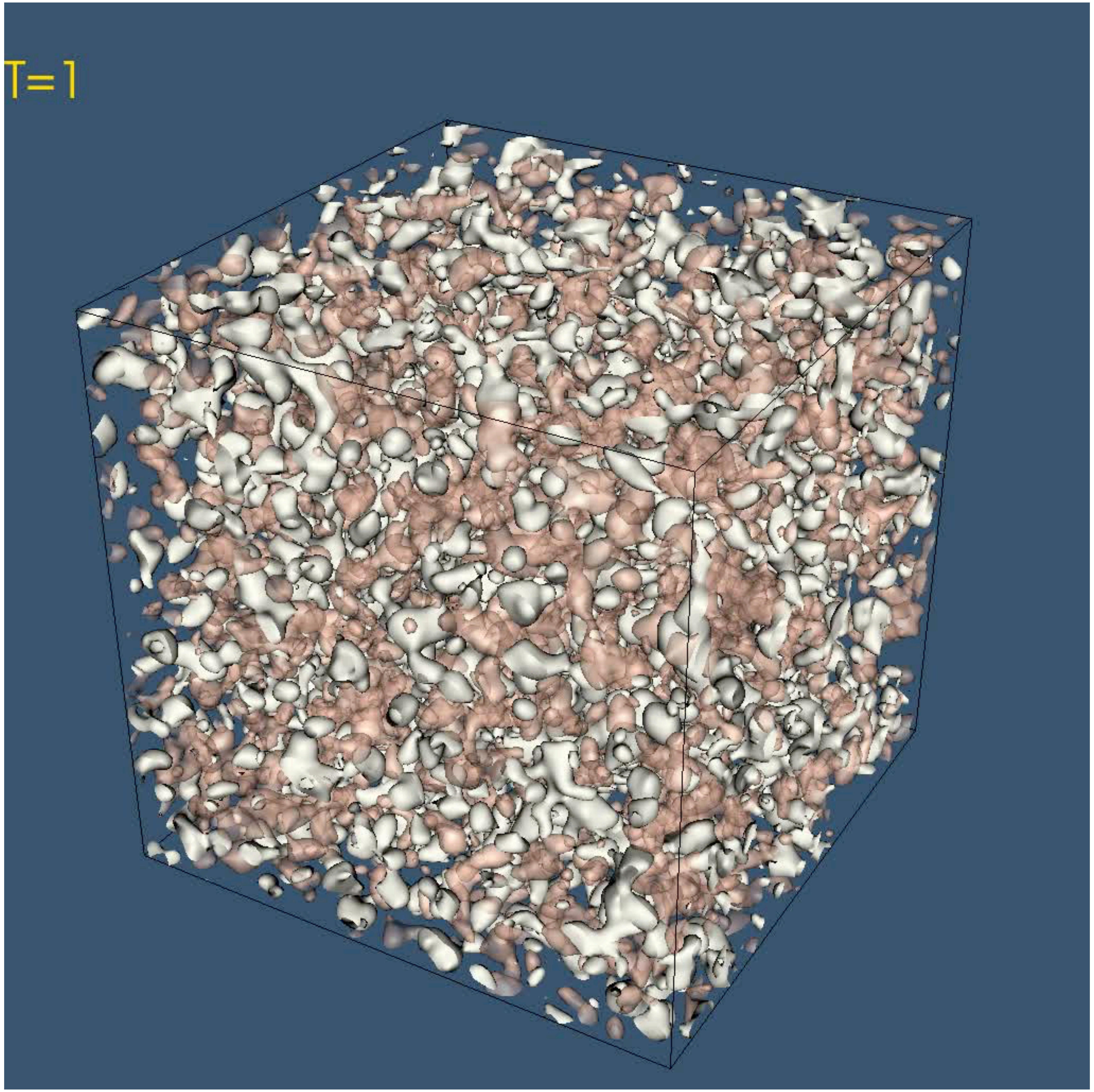
Part 2: topology change in HMC & the index theorem

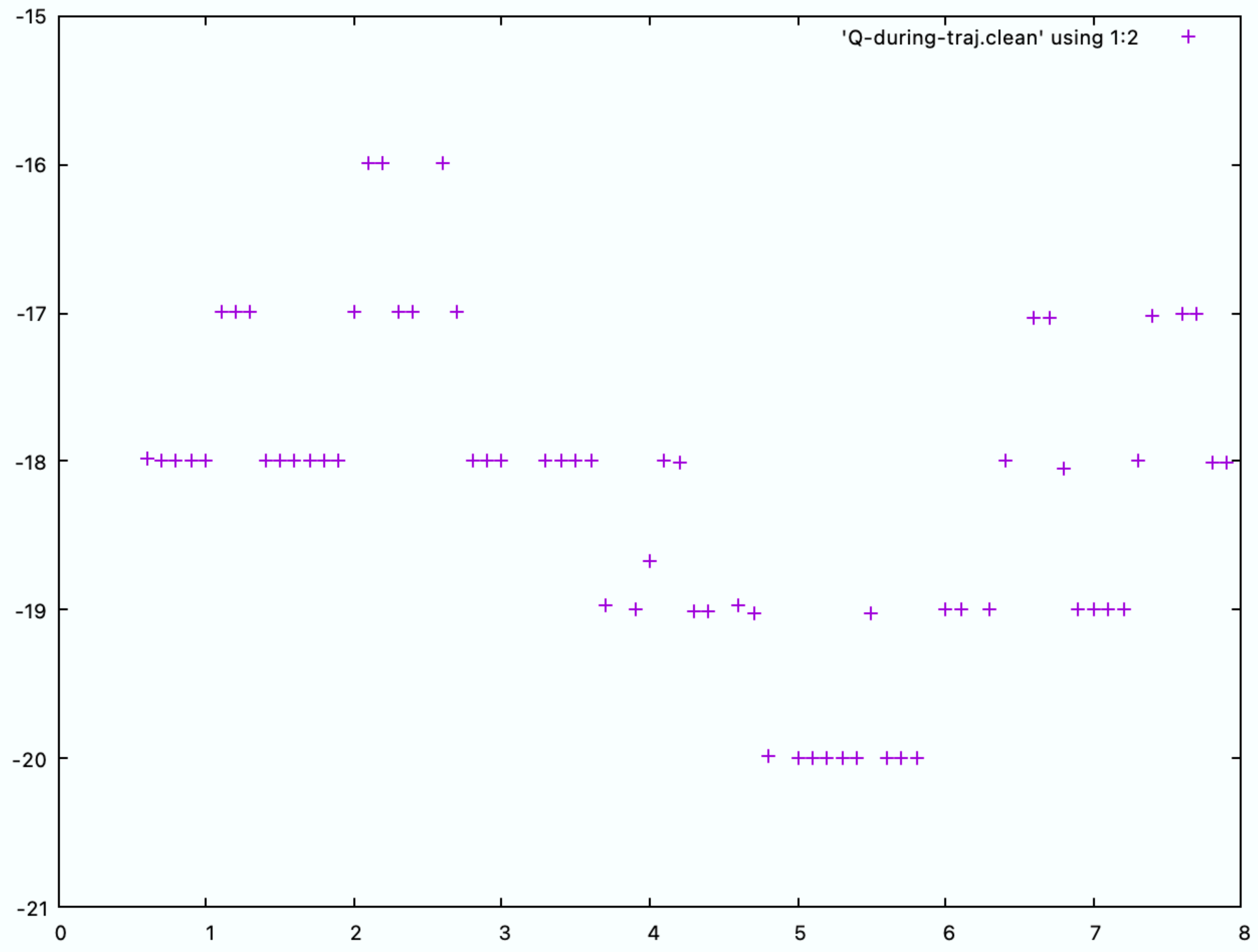
PB, Yamamoto, Lehner, Jung

Interested in accelerating topology and the sample of our $128^3 \times 288$ and other lattices

- Evolution in GPT by Christoph Lehner
- Grid has new visualization features
- Topology **not** frozen, but suspiciously slow sampling of full distribution







total Wilson flowed topological charge, measured every timestep along a trajectory
There is more activity than is consistent with staying around -18 => regression to the mean is occurring

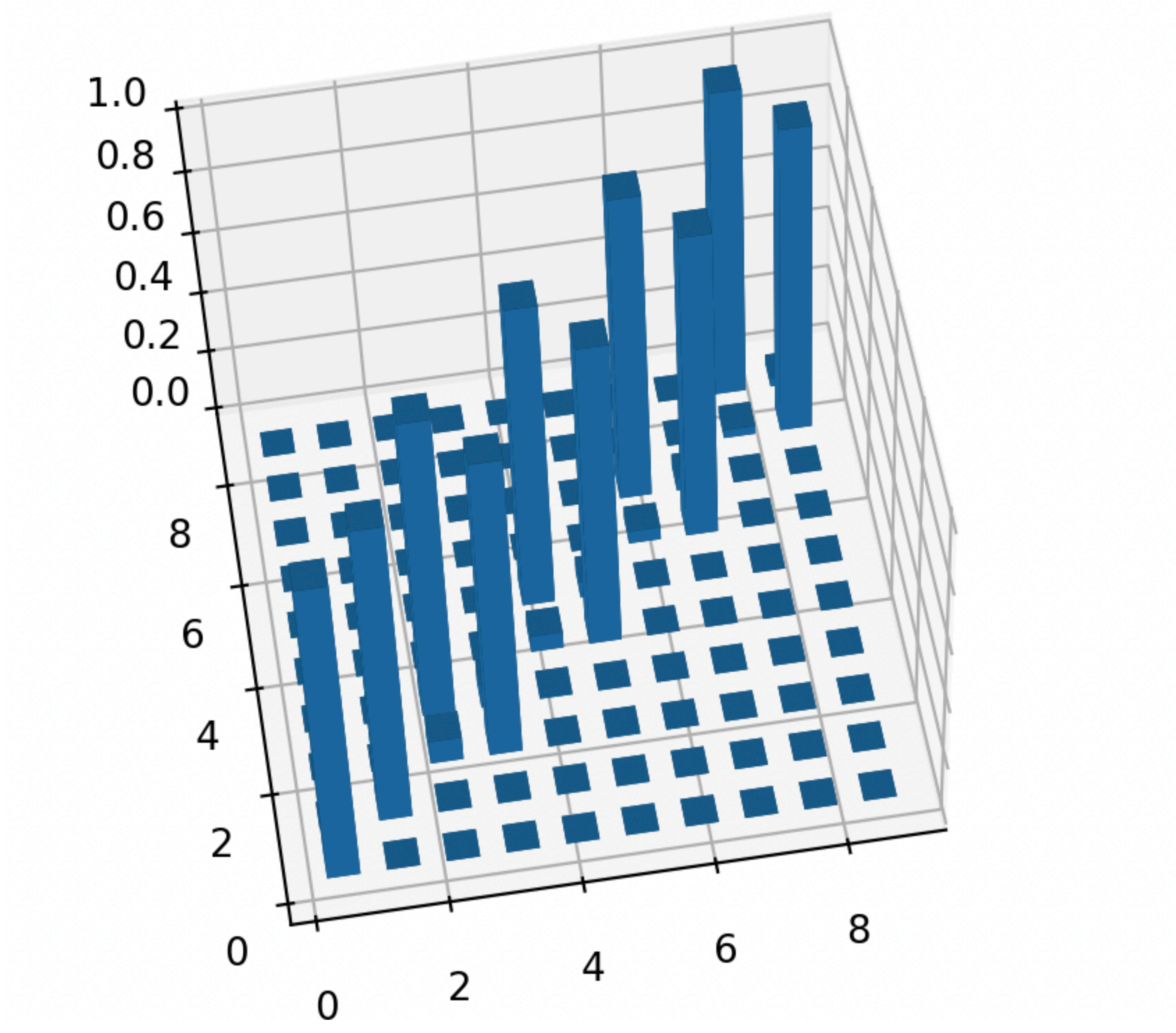
Current conjecture:

a) (Wilson) flow projects from continuous topology into discrete space:
   Non-linear transform; there must be bifurcation points
b) Small changes in initial configuration can lead to big changes in flowed t-charge
c) Possible explanation for WHY FT-HMC is working:
   - Consistent with Luscher's argument about the connectedness of the space of likely configs.
   - Transform from an integration variable in which the space is much closer to connected:
     - Step between valleys by going up, across then down!
d) Can we change/learn a field transform to encourage growth of large, physical instantons?

Fermionic topology has been well studied with DWF

Fermionic definition: produce eigenvectors of $H = \Gamma_5 D$ (Lanczos algorithm, large Ls, zero quark mass)
Zero modes coincide and have definite chirality

Topology change involves paired complex eigenvalues of Dw migrating to real axis and unpairing
Overlap sign function projects a zero mode of Hw(m) either down to zero or up into lattice cut-off depending on mass



$\langle \psi_i^H | \Gamma_5 | \psi_j^H \rangle$

arXiv:hep-lat/0105006

CU-TP-1014, BNL-HET-01/10, RBRC-185

Chirality Correlation within Dirac Eigenvectors from Domain Wall Fermions

T. Blum[a], N. Christ[b], C. Cristian[b], C. Dawson[c], X. Liao[b], G. Liu[b], R. Mawhinney[b], L. Wu[b], Y. Zhestkov[b]

[a]*RIKEN-BNL Research Center, Brookhaven National Laboratory, Upton, NY 11973*
[b]*Physics Department, Columbia University, New York, NY 10027*
[c]*Physics Department, Brookhaven National Laboratory, Upton, NY 11973*

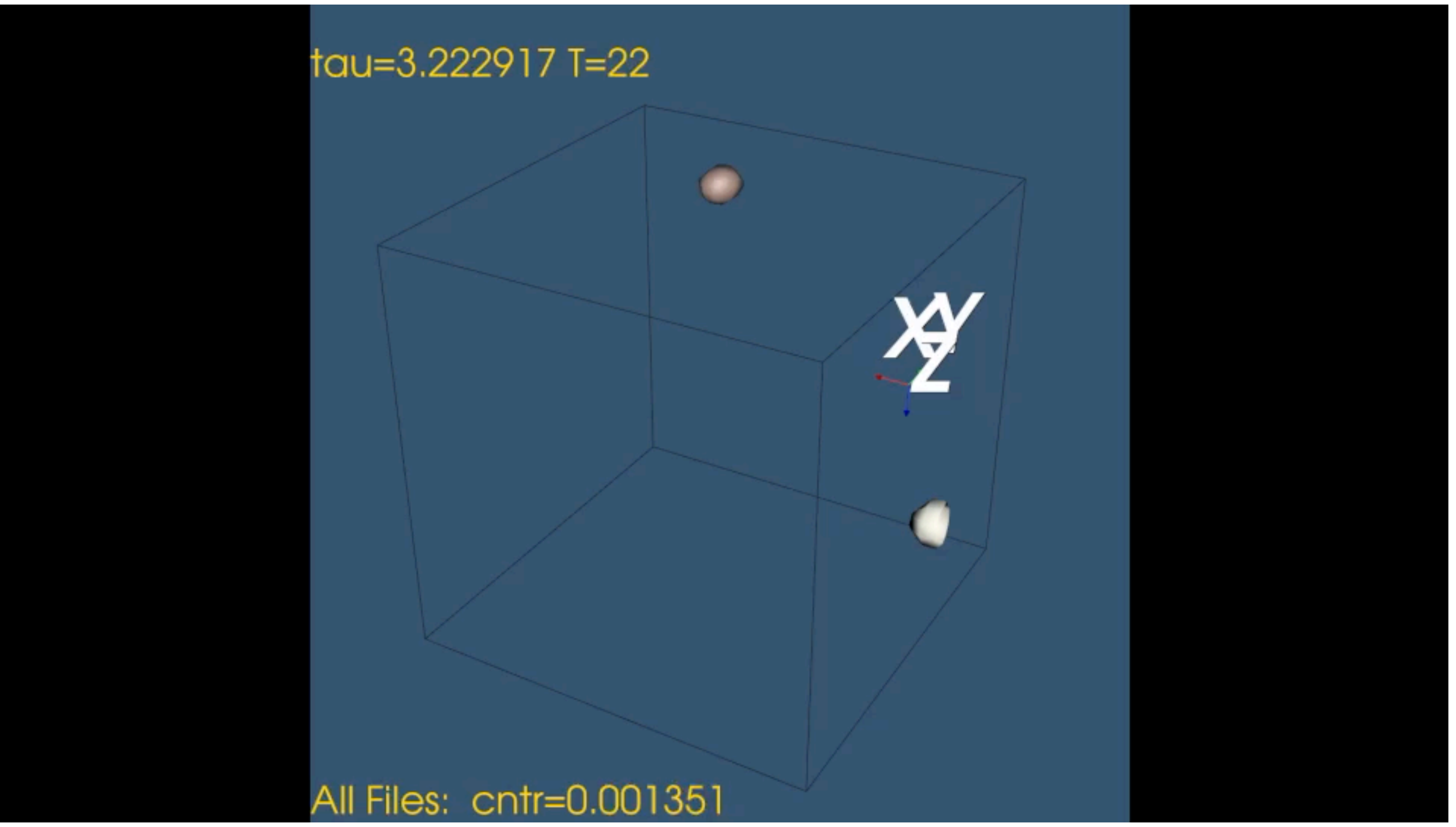
(May 6, 2001)

“Lego plot”

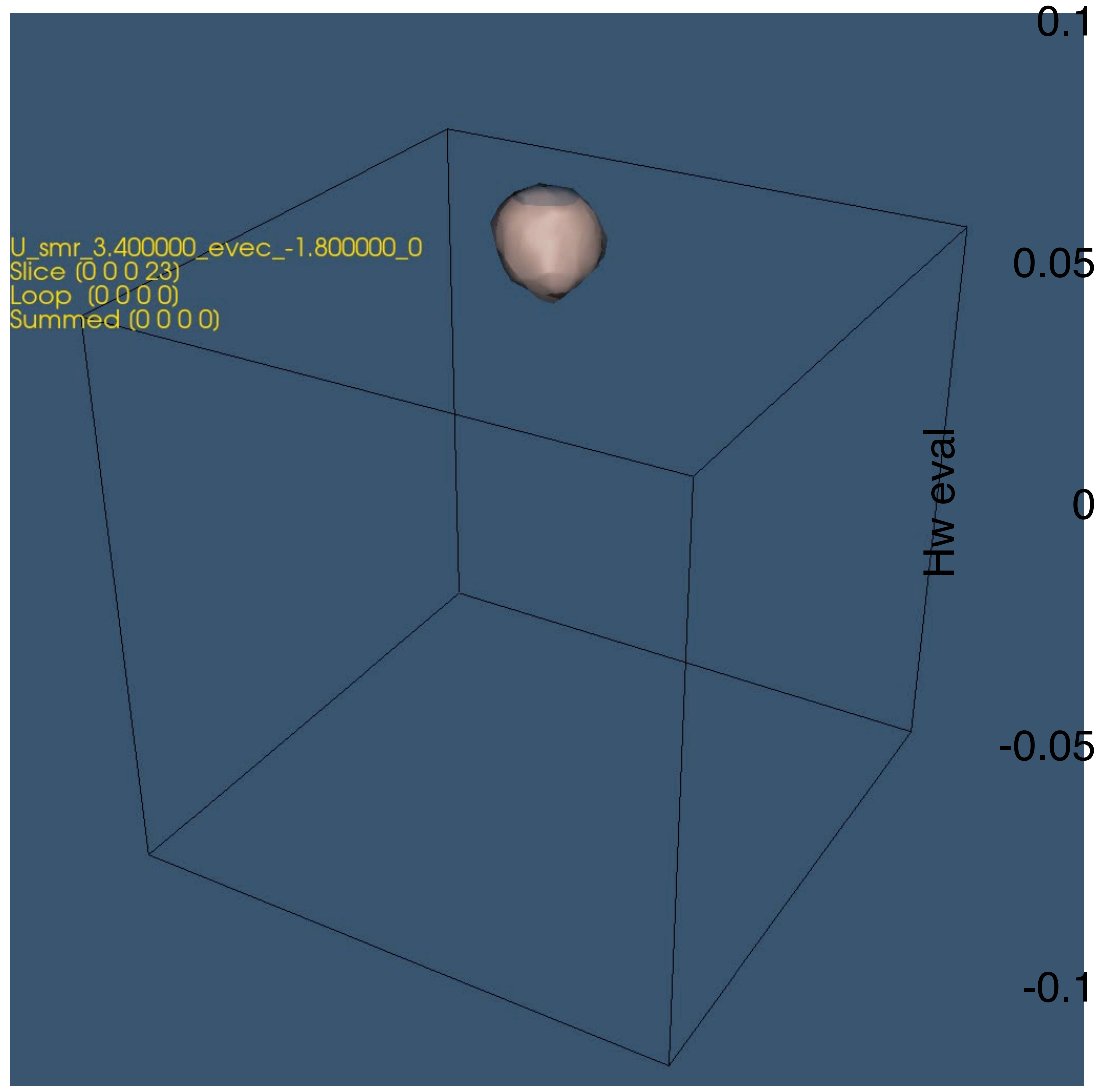
Counts unpaired modes for net topology

- Detailed / microscopic investigation of topology change in 32^4 HMC simulation
  - Wilson flowed topological charge is changing multiple times on each trajectory
    But long term exploration is slower than this would suggest
  - Topology change is often accompanied by local reversion

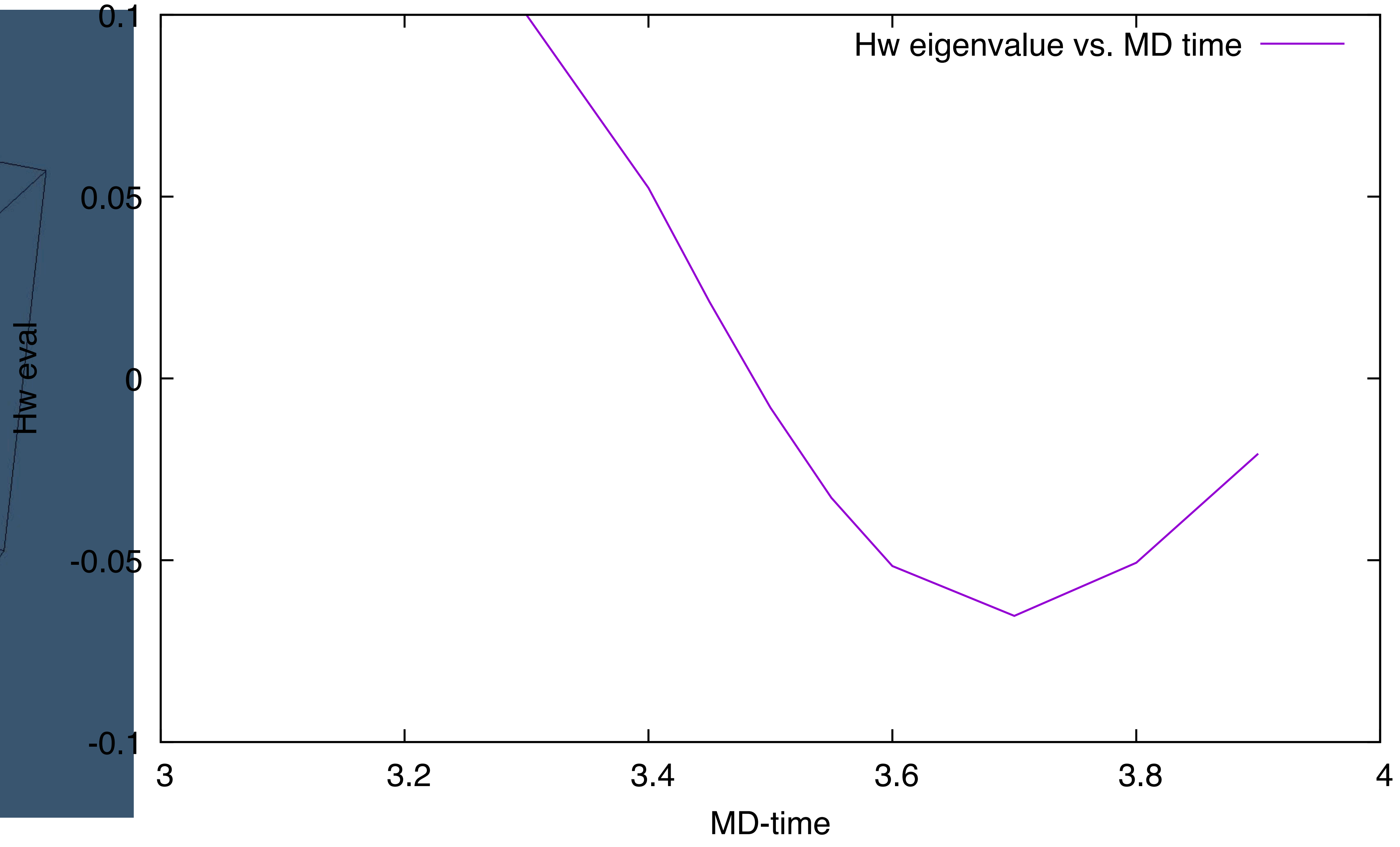
$$D_{ov} = \frac{1}{2}\left(1 + \gamma_5 \frac{H_W}{\sqrt{H_W^2}}\right) = \frac{1}{2}(1 + \gamma_5 \mathrm{sgn} H_W)$$



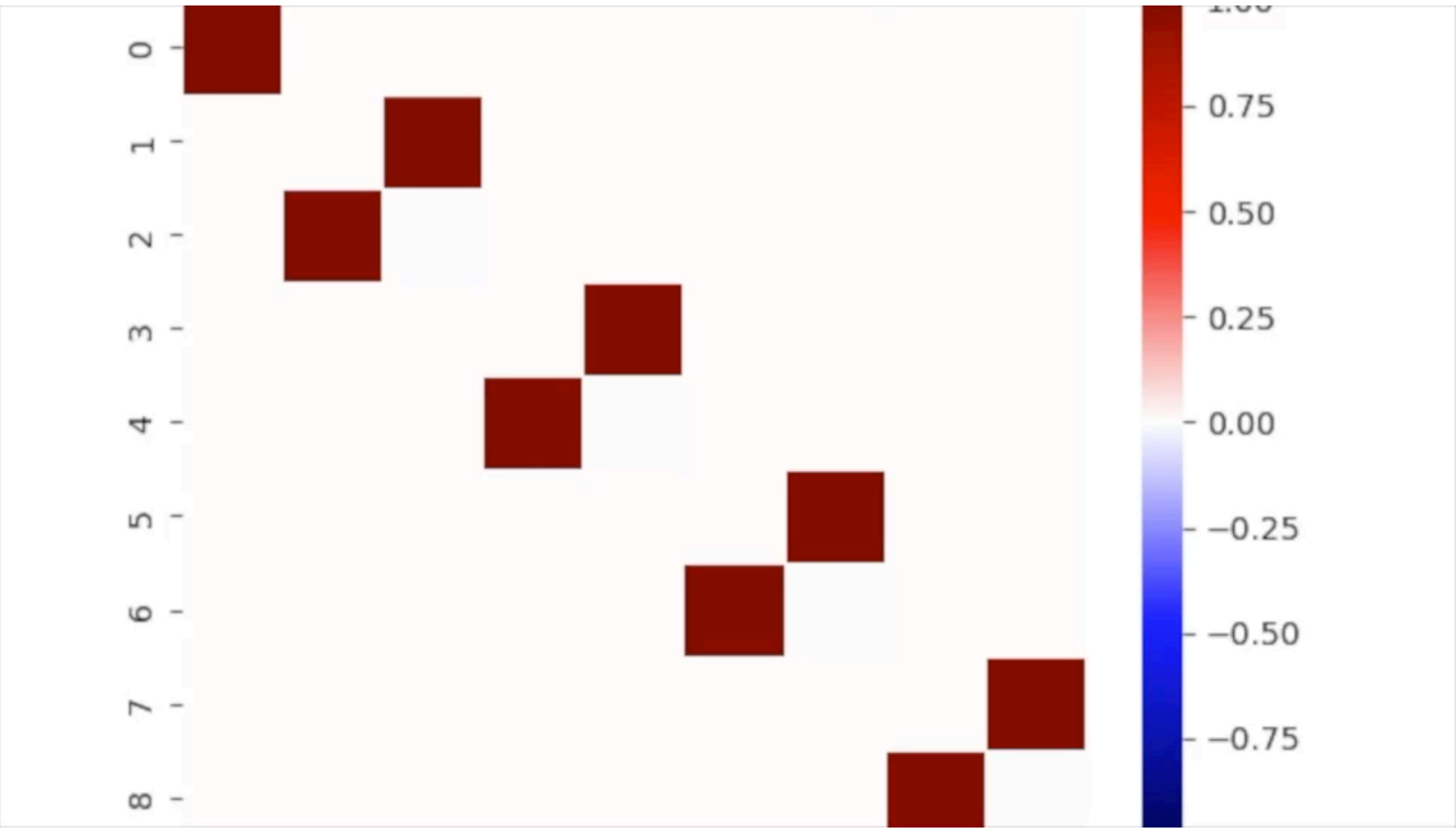
32^4 test : topological charge Density



Lowest Hw eigenvector

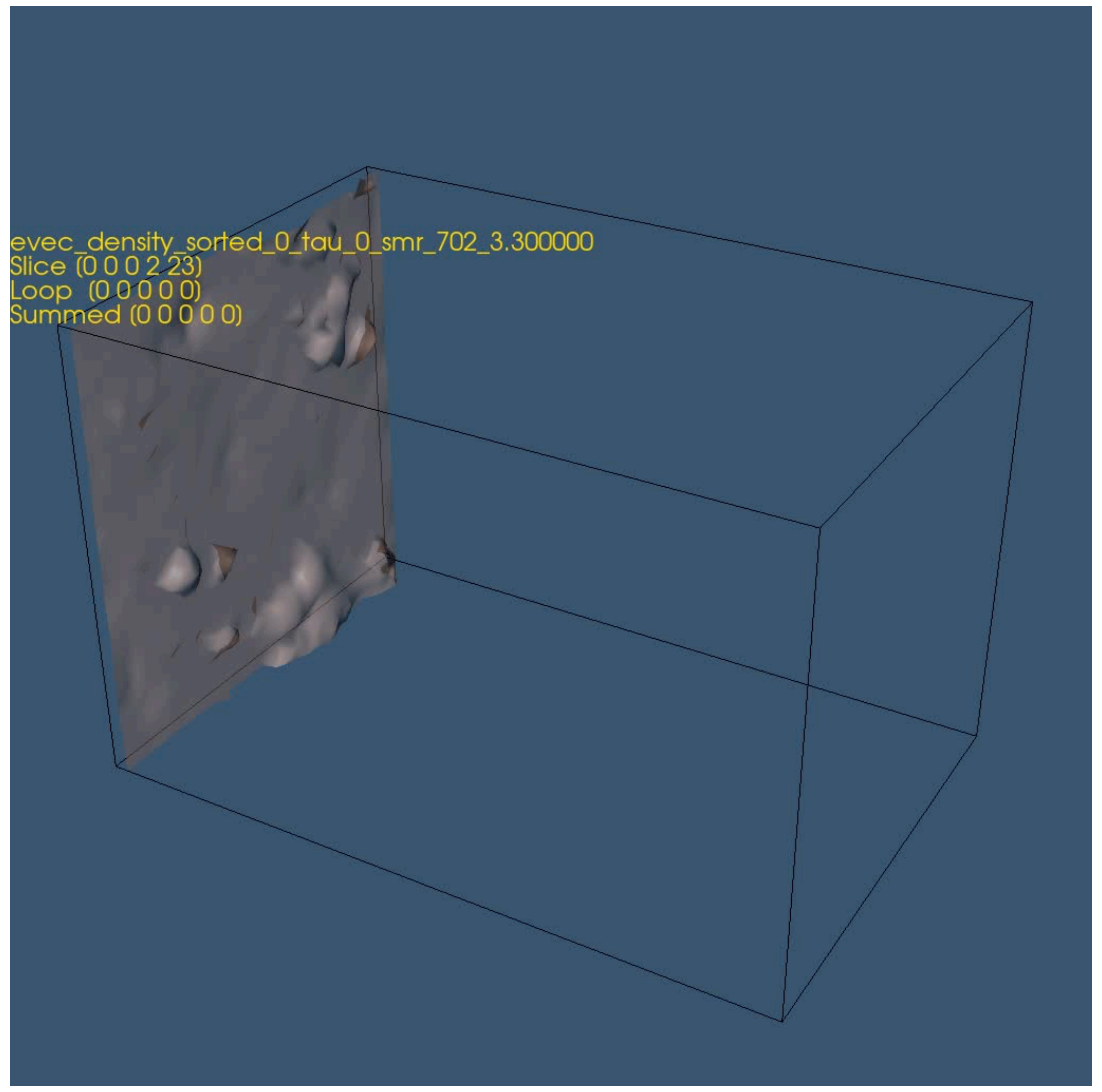


$\langle \psi_i | \Gamma_5 | \psi_j \rangle$

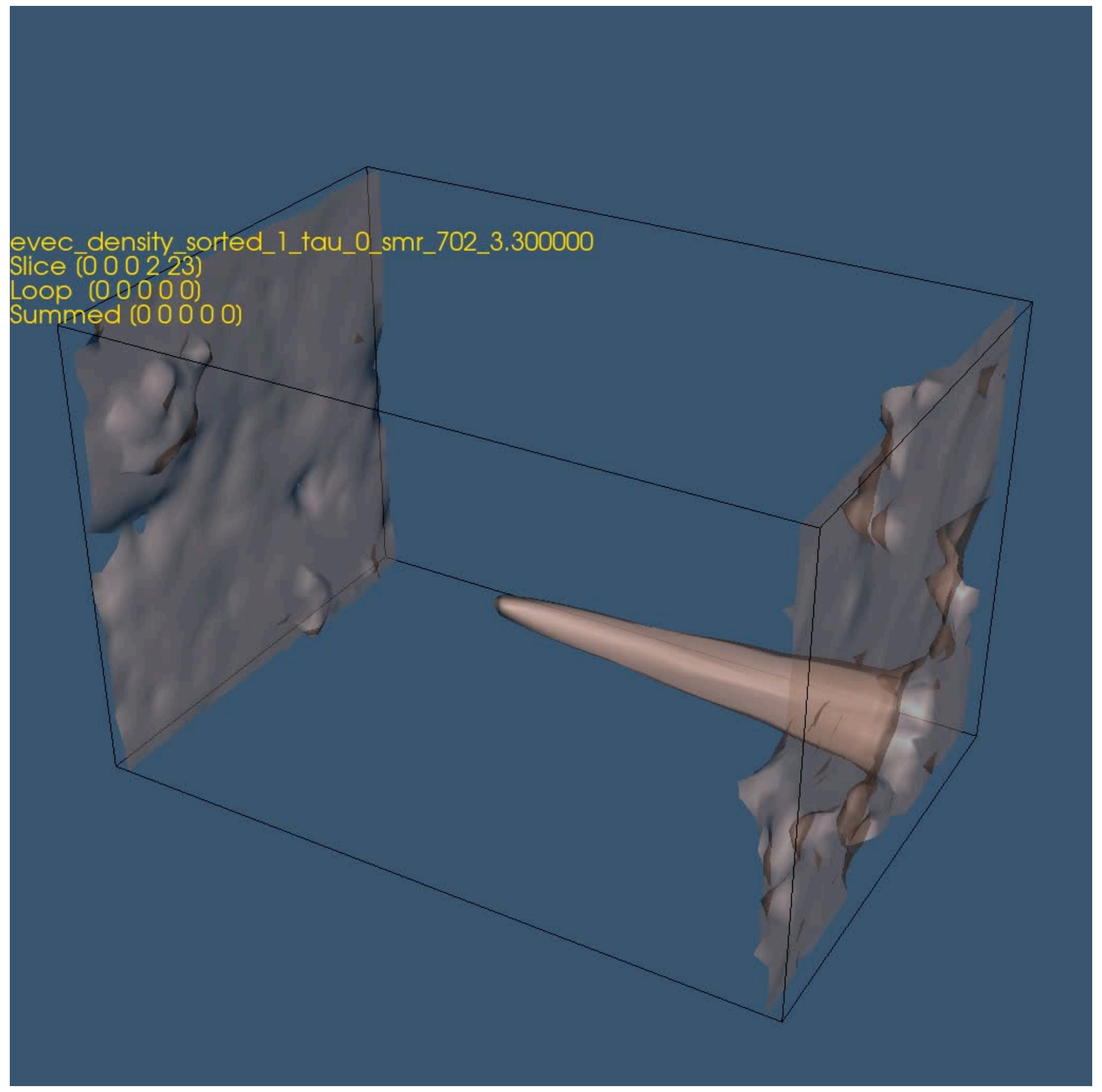


Continuum chiral properties:
eigenmodes are
EITHER
paired and mixed chirality
OR
Chiral, unpaired (+topological)

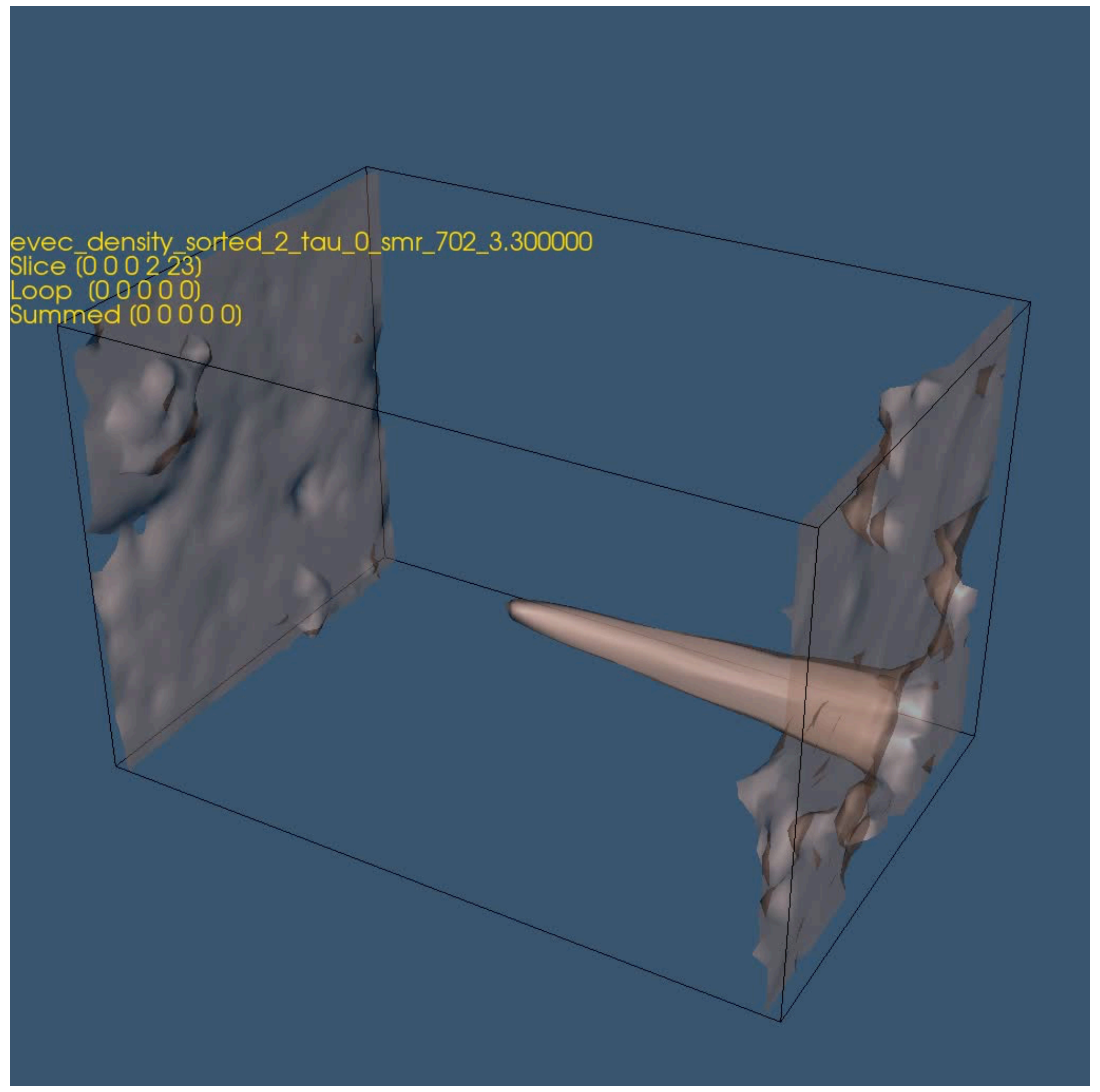
ψ_0



ψ_1



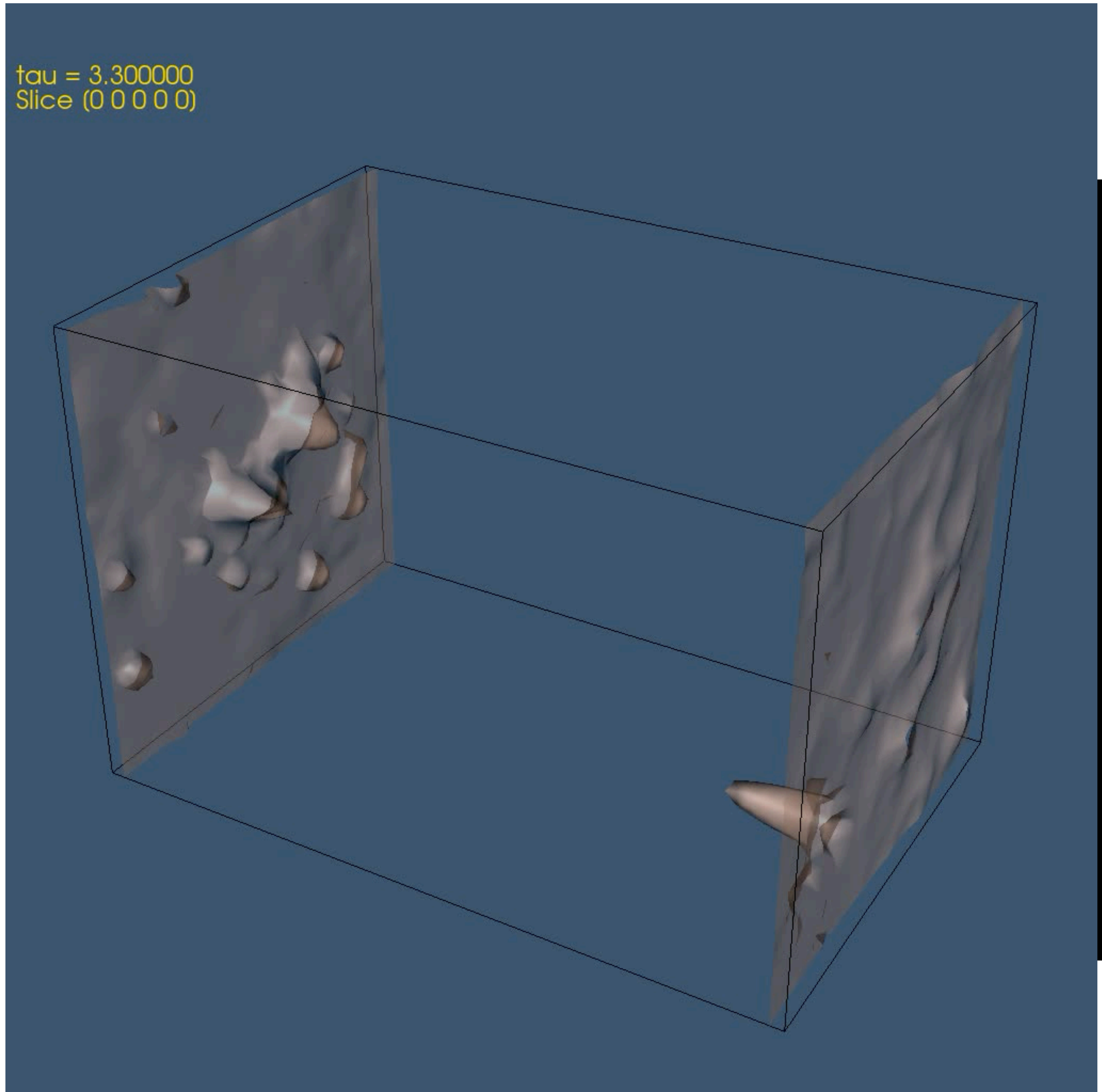
ψ_2

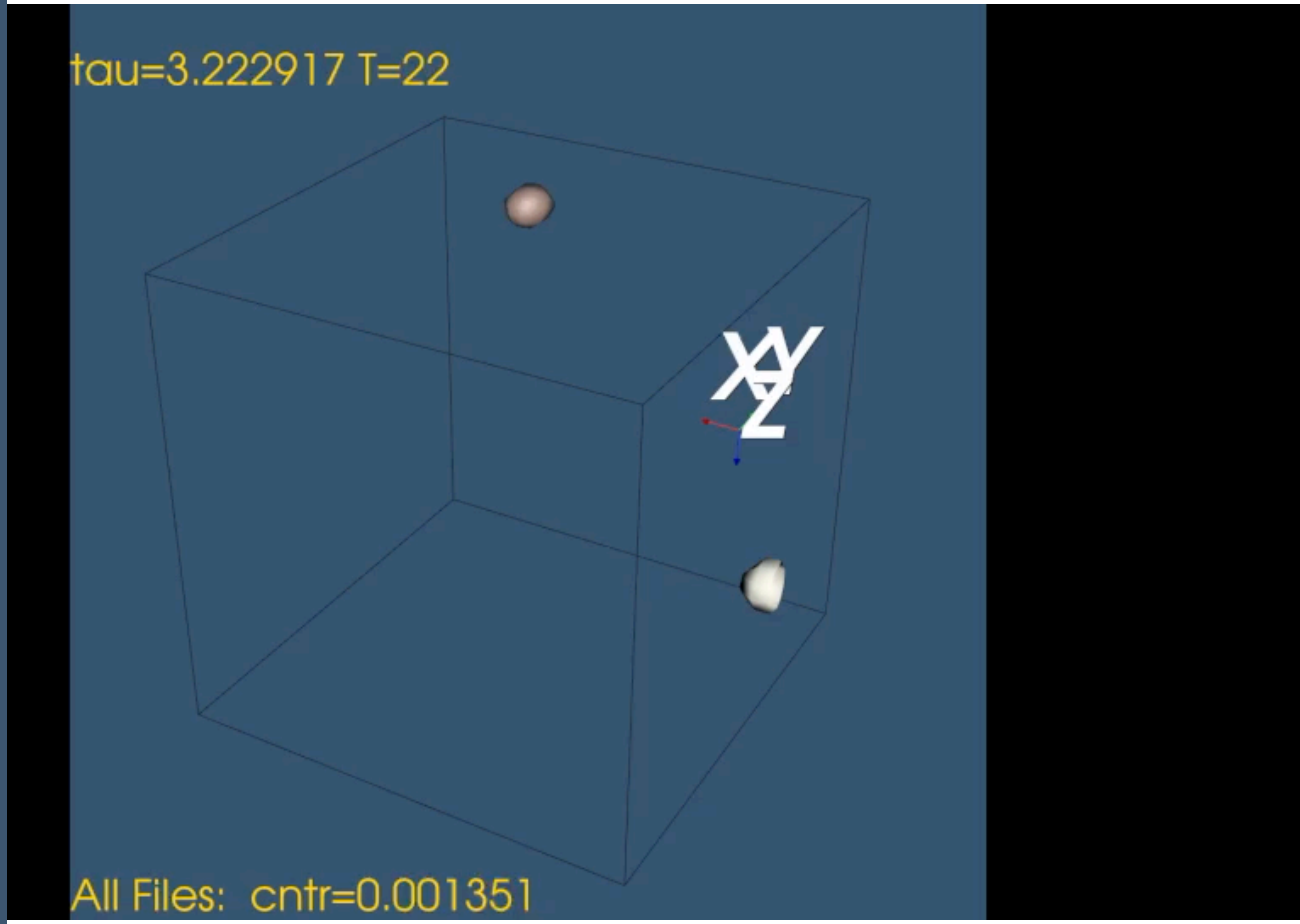


[Claude]

It's a remarkably clean demonstration of the mechanism. The visualisation makes it tangible — you can literally watch the norm drain from one wall and the mode asymptotically approach a chirality eigenstate, then reverse. The fact that it's mode #1 doing this, while mode #0 sits stationary on its wall throughout, is a direct geometric realisation of the index theorem on the lattice.

Natural language control of vis!





Conclusions

Multigrid progress

- New $PV^{\dagger}M$ approach to chiral fermion multigrid is promising.
  - MultiRHS *already* on a par with mrhs-HDCG.
  - Planned exact inverse on super-coarse : algorithmically proven, but not used yet!
- Optimistic for both valence mrhs and HMC single rhs usage.

Topology change:

- Intriguing local mode reversion of Hw during HMC almost tunneling events
  - Wilson flow projects distribution of un-smoothed Q into discrete space => bifurcation.
- Suspect this is related to how FTHMC is working and can be improved.
- Have always wanted to capture chiral mode rearrangement visually!