NVIDIA

STATE OF THE ART MULTI-GRID ALGORITHMS IN QUDA

Evan Weinberg, July 26 2021

AGENDA

Overview

Brief summary of what's in QUDA, recent accomplishments, asks for the community

Wilson-Clover Multigrid

It's still awesome

HISQ Multigrid through MILC

What's there and how it's always improving

Domain Wall Multigrid

In theory

Community Asks and Future Work

Where we're going in the next year and how you can help

OVERVIEW

QUDA MULTIGRID TL;DR

If you pay attention to any one slide…

- Wilson-clover (+/- a twist) has been there and is amazingly performant
  - MG-accelerated HMC is available in Chroma
  - Mathias Wagner, “Strong scaling RHMC on NVIDIA GPUs”, Software Development Parallel, July 28 @ 1:15 pm EST
- HISQ MG “beta” is available through MILC
  - https://github.com/lattice/quda/wiki/HISQ-MG-for-Measurements
  - With fresh performance optimizations and 30% memory overhead reductions
  - *I want to help you use it!*
- MG setup has been accelerated with tensor cores
  - Jiqun Tu, “Use Tensor Cores to Accelerate Math Intensive Kernels in QUDA”, poster
- Domain wall/mobius MG is still a work in progress
  - Please find me more hours in the day

WHAT IS QUDA?

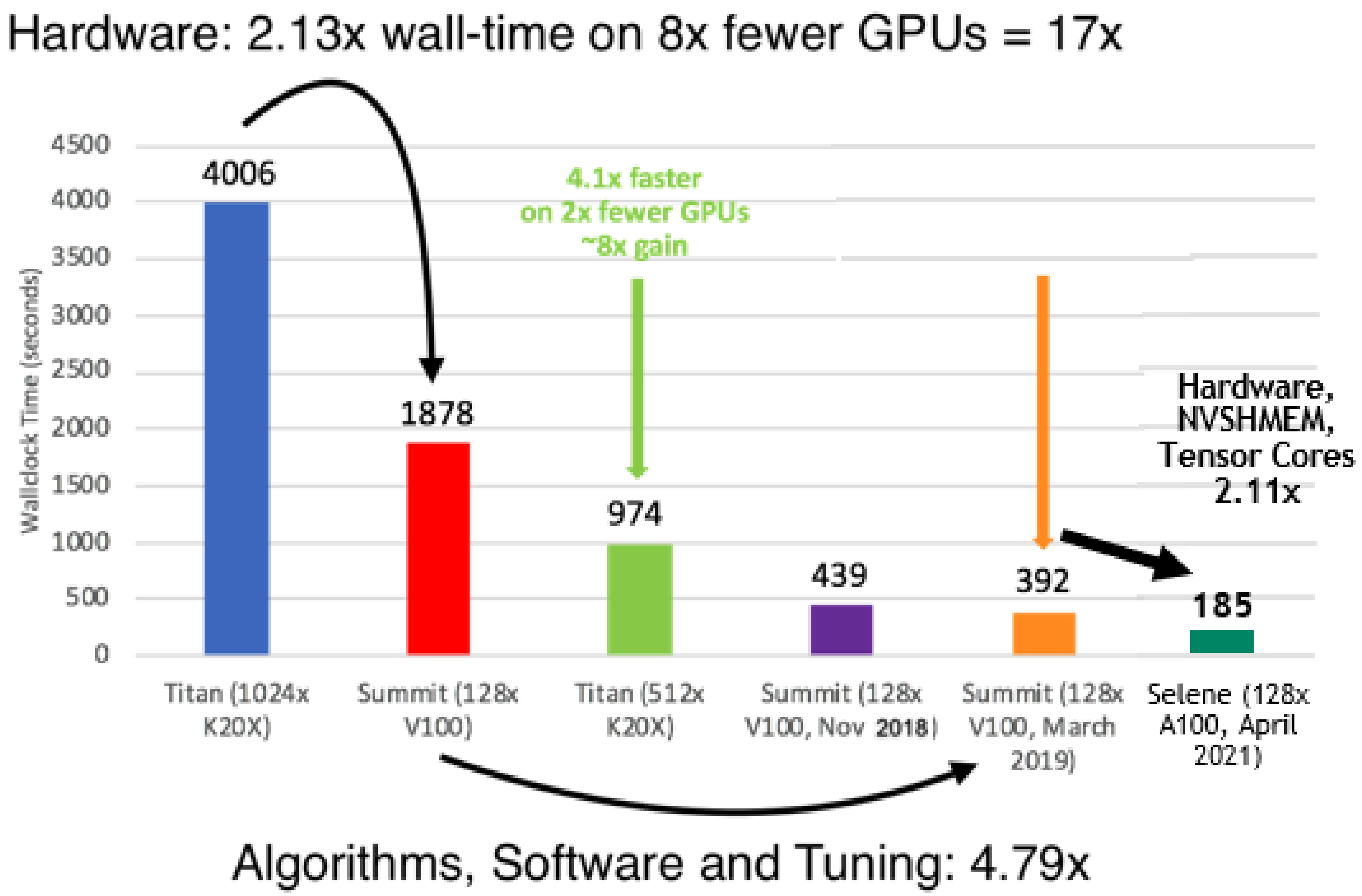
If you aren't tired of hearing yet

- See Kate Clark, "Preparing for QUDA 2.0," July 28 @ [time] to learn more
- But, in brief…
- "QCD on CUDA": https://github.com/lattice/quda
    - Open source, BSD license
    - Formal support for more architectures coming
- Effort started at Boston University in 2008, now in wide use as the GPU backend for BQCD, Chroma, CPS, MILC, TIFR, …
- Provides multi-GPU solvers for all major fermion discretizations at maximum performance

WILSON-CLOVER MULTIGRID

IN SUMMARY

What Multigrid has done for Chroma HMC



Chroma w/ QDP-JIT and QUDA, ECP FOM data,

V=64^3x128 sites, m_π ~172 MeV, (QDP-JIT by F. Winter, Jefferson Lab)

Original figure credit Balint Joo ~2 years ago, new numbers from M. Wagner

HISQ MULTIGRID IN MILC

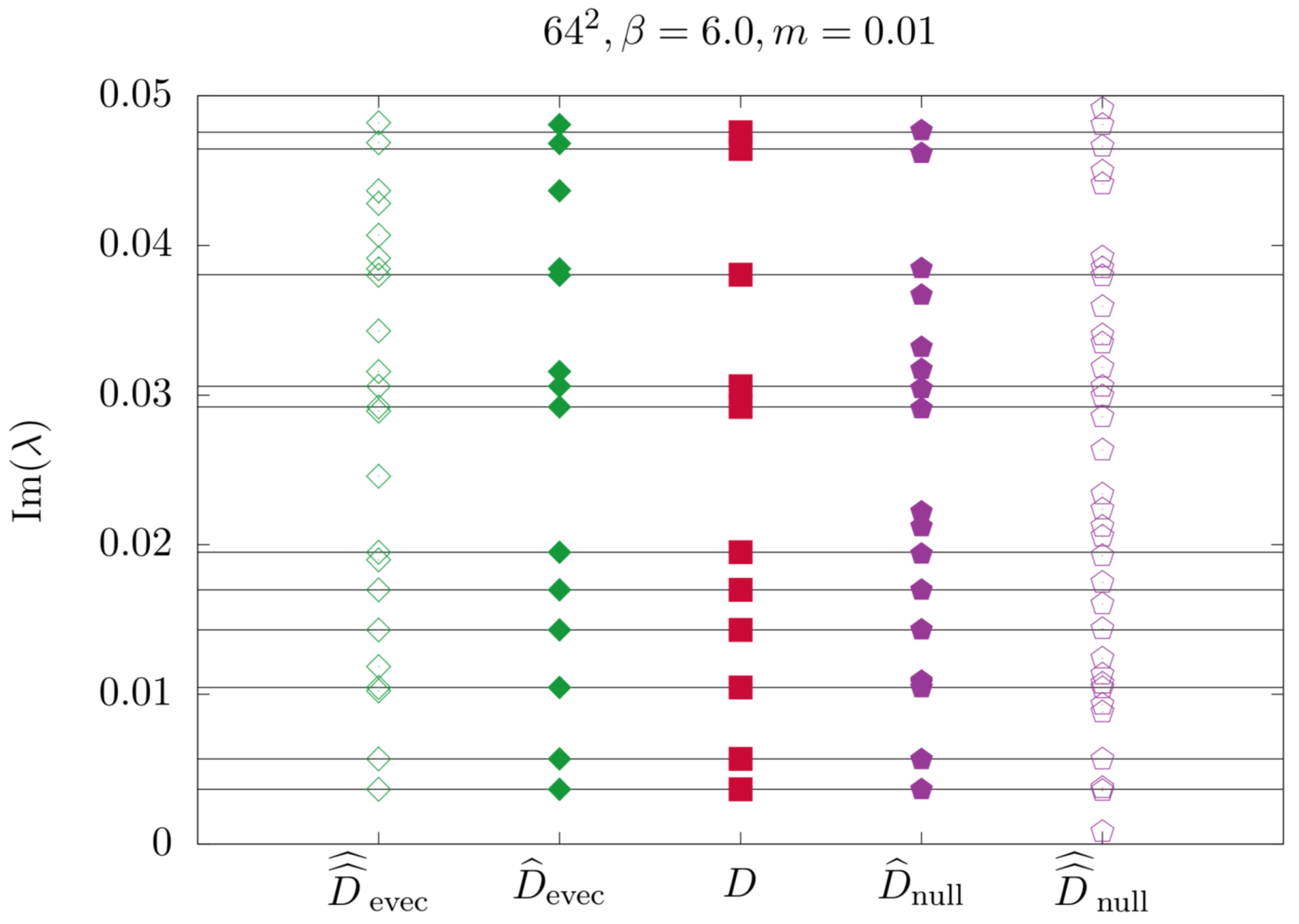
STAGGERED WOES

arXiv:1801.07823

- The staggered operator is maximally anti-Hermitian indefinite (plus a real mass shift)

$$D_{stag}(m) = D_{stag} + m\mathbb{I}$$

- Above some small volumes, a naïve Galerkin projection does not work
- Spurious small eigenvalues appear in the spectrum



THEORY: KAHLER-DIRAC PRECONDITIONING

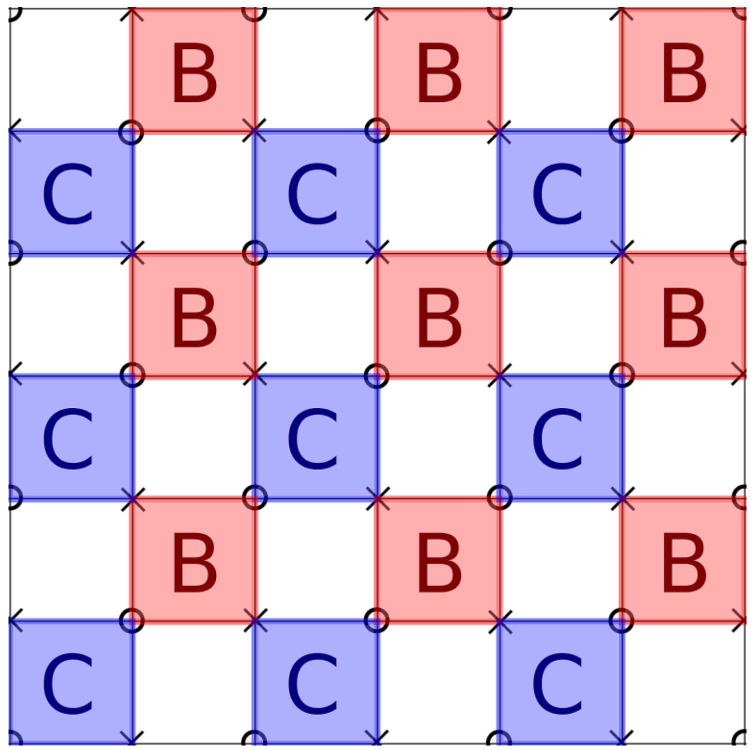
arXiv:1801.07823

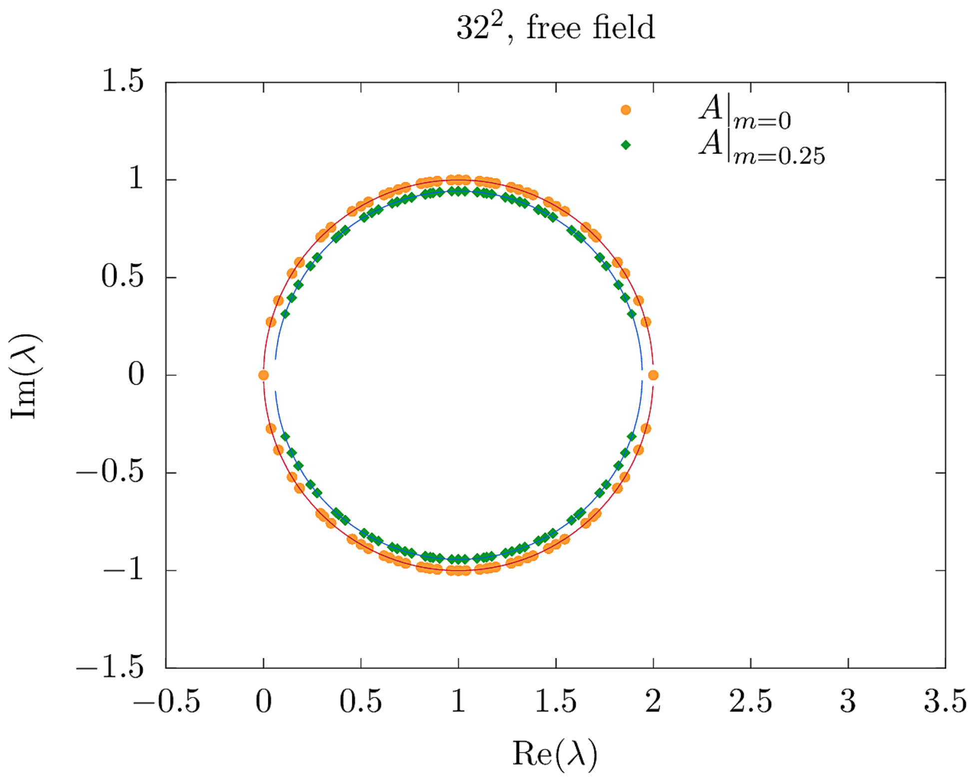
- Key observation: the 2^d hypercube of degrees of freedom is equivalent to a Kahler-Dirac fermion (in the free field)
- Write the staggered operator as a dual-decomposition: $D_{stag} = B + C + m$
    - B: hopping terms within a 2^d block
    - C: hopping terms across blocks
- Perform a block-preconditioning by $(B+m)^{-1}$

$$D_{KD}(m) = (B+m)^{-1}[(B+m)+C]$$
$$= \mathbb{I} + (B+m)^{-1}C$$

Result: overlap-esque spectrum

- Perfect circle in free field
- "Fuzzy" when interactions are enabled





PREVIOUS APPROACH IN QUDA

HISQ Stencil in 4 Dimensions

- Hack the MG machinery in QUDA to manually form an N_{dof} = 48 (2^4 times N_c = 3) degree-of-freedom preserving operator

$$D_{HISQ}(m) = D_{fat} + D_{long} + m\mathbb{I} = \underbrace{(B + m\mathbb{I})}_{X} + \underbrace{C}_{Y} + D_{long}$$

$$D_{KD}(m) = \underbrace{(B + m\mathbb{I})^{-1}}_{Xinv} \left[\underbrace{(B + m\mathbb{I})}_{X} + \underbrace{C}_{Y} + \cancel{D_{long}} \right] = \mathbb{I} + \underbrace{(B + m\mathbb{I})^{-1} C}_{Yhat}$$

- Drop the Naik (three hop) term, justified by via a perturbative argument
- Explicitly forming Y and Yhat ignores a lot of sparsity on the table:
  - Staggered fat link only: 36 complex numbers per fine site
  - Coarse KD hopping term: 1,152 complex numbers per fine site
  - 32x memory bloat!
- This approach is messy and limited per-GPU problem sizes but it worked well enough.

OPTIMIZED APPROACH

If you’ve got it, flaunt it

- Take this expression seriously: $D_{KD} = (B + m\mathbb{I})^{-1} D_{HISQ} = X^{-1} D_{HISQ}(m)$
- Literally implement the KD preconditioned operator as a HISQ stencil application + block-local KD inverse
  - Reuse the fat links from the outer solve
  - Naturally include the long links, take advantage of long link compression
  - Take advantage of NVSHMEM
- The only extra memory overhead is X^{-1}: 144 complex numbers per fine site
  - 4x larger than the fat gauge links
  - 8x *smaller* than Yhat (previous coarse hopping terms)
- Directly coarsening the KD operator becomes *more efficient* because there are fewer memory overheads

THE METHOD IN ACTION

Large MILC Configurations

- **Physical pion mass** configurations courtesy of Carleton DeTar (MILC collaboration)

Volume	β	a (fm)	am_l	am_s	am_c
96^3x192	6.72	0.06	0.008	0.022	0.260
192^3x384	7.28	0.03	0.00415	0.01229	0.1329

- Real workflow: ECP KPP measurement
  - 10 “lighter” masses, including light and strange quarks, traditionally in a multi-shift CG
  - 10 “heavy” masses, traditional CG, solved to a fixed “heavy quark” residual
- Modifications for a HISQ MG test:
  - “Peel” off three lightest masses for MG solve; reuse coarsened links *but update mass*
  - Remaining seven masses remain in a multi-shift solve

THE METHOD IN ACTION

Large MILC Configurations

- **Physical pion mass** configurations courtesy of Carleton DeTar (MILC collaboration)

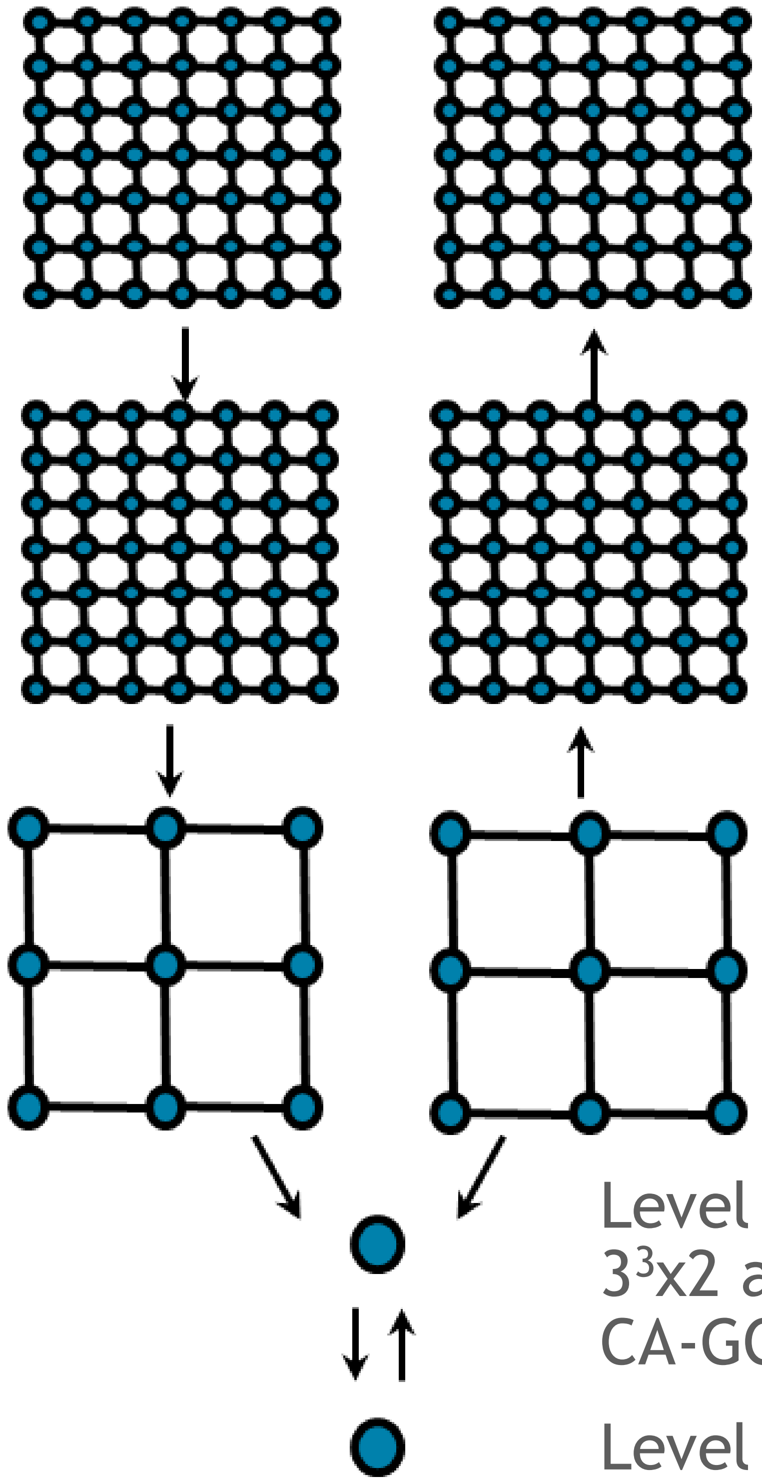
Volume	β	a (fm)	am_l	am_s	am_c
96^3x192	6.72	0.06	0.008	0.022	0.260
192^3x384	7.28	0.03	0.00415	0.01229	0.1329

- Real workflow: ECP KPP measurement
  - 10 “lighter” masses, including light and strange quarks, traditionally in a multi-shift CG
  - 10 “heavy” masses, traditional CG, solved to a fixed “heavy quark” residual
- Modifications for a HISQ MG test:
  - “Peel” off three lightest masses for MG solve; reuse coarsened links *but update mass*
  - Remaining seven masses remain in a multi-shift solve

FIVE LEVEL ALGORITHM

Letting the optimized operator flex

- We use a 24^4 local volume, 512 GPUs
  - Could use as large as a 32^2x48^2 local volume, **72 GPUs**
  - We needed **432 GPUs** two years ago!
- Level 1 is now the "fine" Kahler-Dirac operator
  - Full NVSHMEM support, etc
- We use SVD deflation on the coarsest level
  - Result of the Lanczos developed by Dean Howarth
  - Deflation still has its place!



Level 0: Fine level
3 dof per site
Full HISQ stencil
10^{-6} tolerance

Level 1: **Fine level**
3 dof per site
Full HISQ stencil + KD prec.
GCR to 0.25 tol., max 16 iter.

Level 2: **Intermediate level**
4^3x6 aggregation, 192 dof per site
GCR(16) to 0.25 tol., max 16 iter.

Level 3: **Coarsest level**
3^3x2 aggregation, 192 dof per site
CA-GCR(16) to 0.25 inner tolerance

Level 4: **SVD Deflation**
0, 32, and 1024 singular vectors

PERFORMANCE IMPROVEMENTS

96^3x192 MILC Configuration, physical pion mass, light mass 0.0008

- Times are for ***six*** solves (two HISQ propagators)
- Note: numbers are from NVIDIA's Selene cluster (number 6 on Top500)
  - That said: this run will fit on the NVIDIA V100-16GB on Summit (I couldn't get jobs through the queue in time)
  - This will also fit on Perlmutter's NVIDIA A100-40GB

Method	Coarse Deflation	Setup	Mass 1	Mass 2	Mass 3	Masses 4-10	Heavy Masses	Total time
Multishift	---	0	386 sec				90.4 sec	459 sec
GCR-MG	32	TBA ☹	40.3 sec	30.6 sec	34.5 sec	110 sec	99.3 sec	314 sec
GCR-MG	1024	TBA ☹	19.2 sec	16.9 sec	21.7 sec	111 sec	101 sec	269 sec

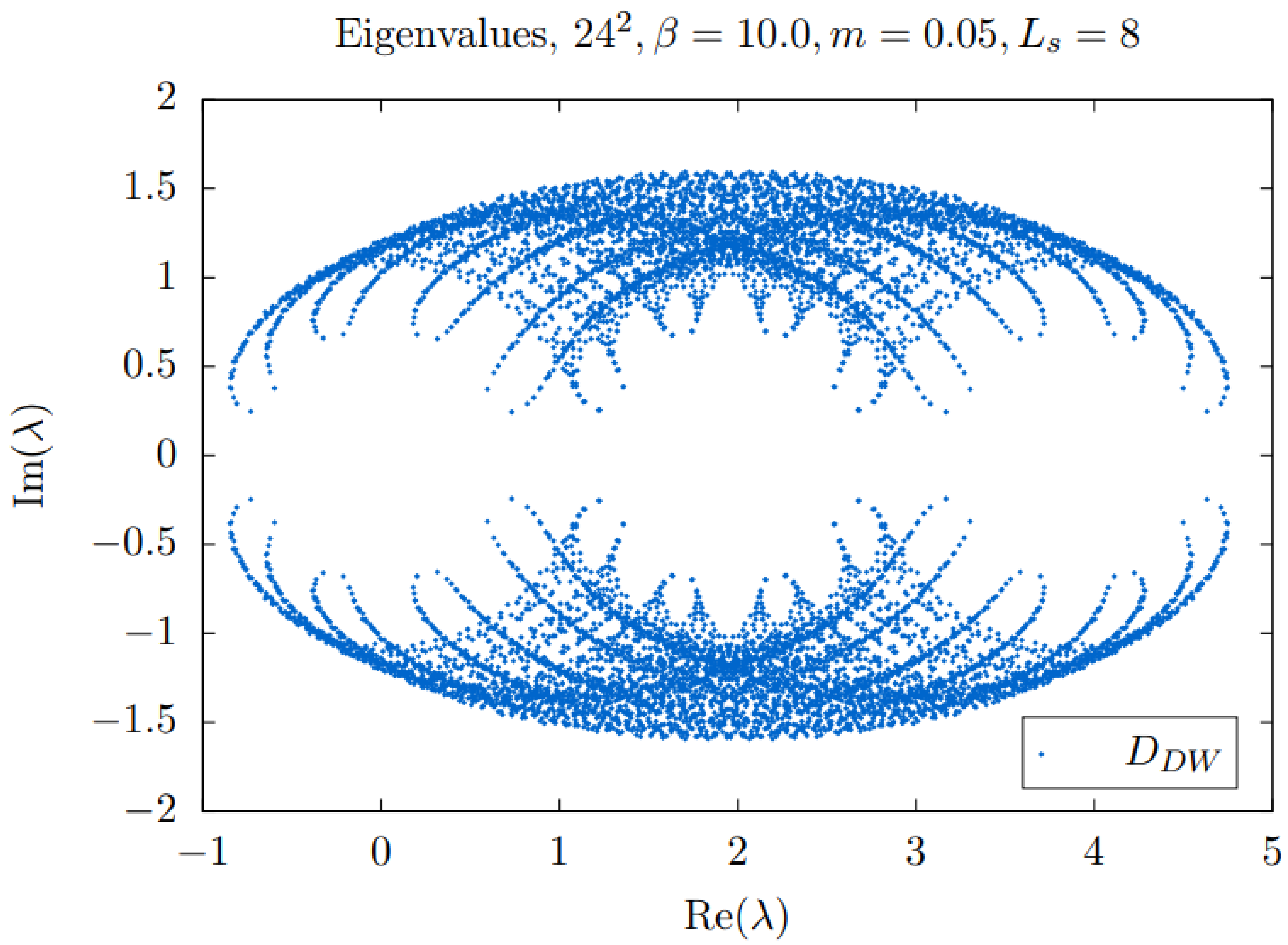
- Took advantage of (almost) all of the magic: gauge compression, but no NVSHMEM (downstream of poor planning)
- Greatly reduced memory overheads: hacky N_{dof}=48 coarse gauge links were ~1.5 GB for a 24^4 local volume
- Greatly reduced setup time (I swear)

DOMAIN WALL, IN BRIEF

DOMAIN WALL WOES

arXiv:2004.07732

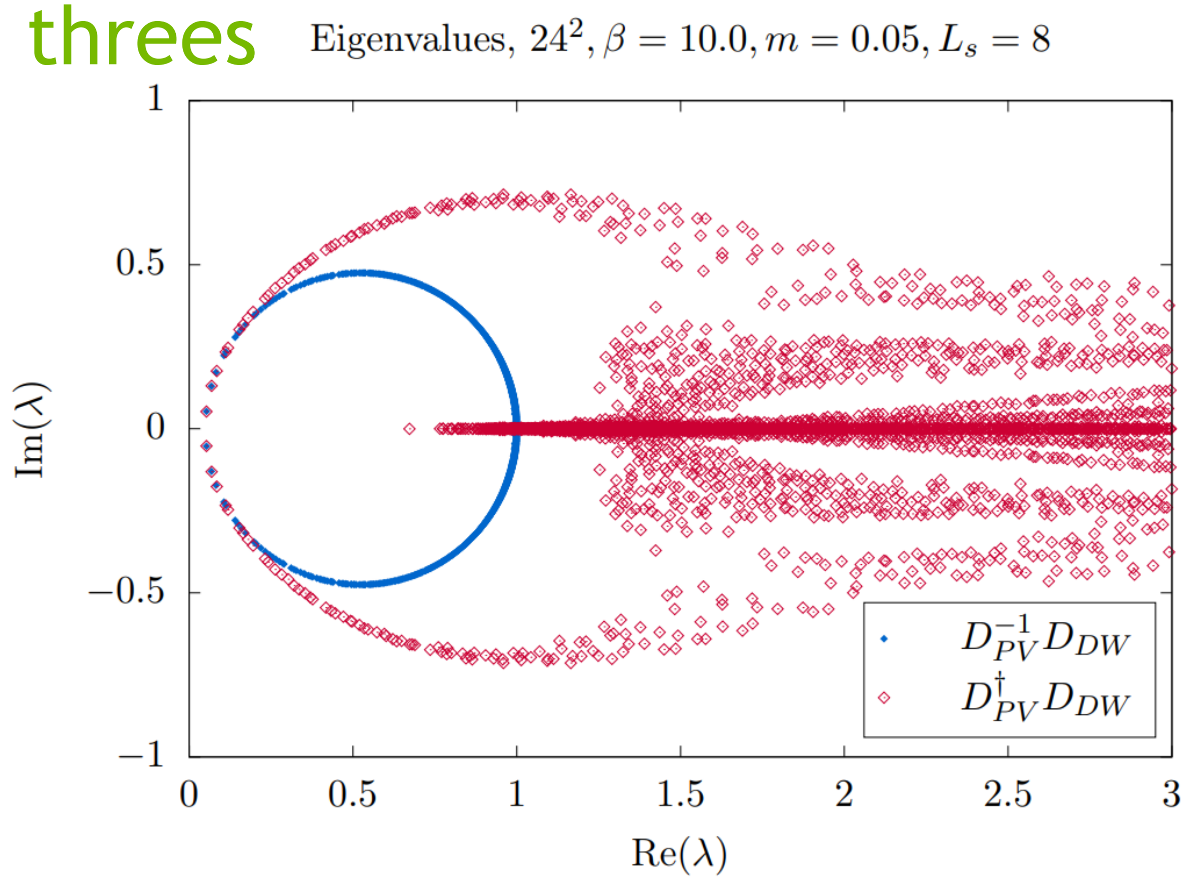
- The domain wall operator is maximally indefinite
- The domain wall operator has the "wrong number" of low modes: N_s x N_c x L_s
- The "normal" approach is the (Schur-preconditioned) normal operator
- Right number of low modes!
- Broadly challenging to apply MG efficiently due to distance > 1 stencil (Cohen 2011, Boyle 2014)
- For some domain wall formulations, "Gamma5" operator is ultra local, MG can be efficiently applied
  - See Peter Boyle, "Algorithms for domain wall Fermions", Algorithms Parallel, July 29 @ 10:45 pm EST

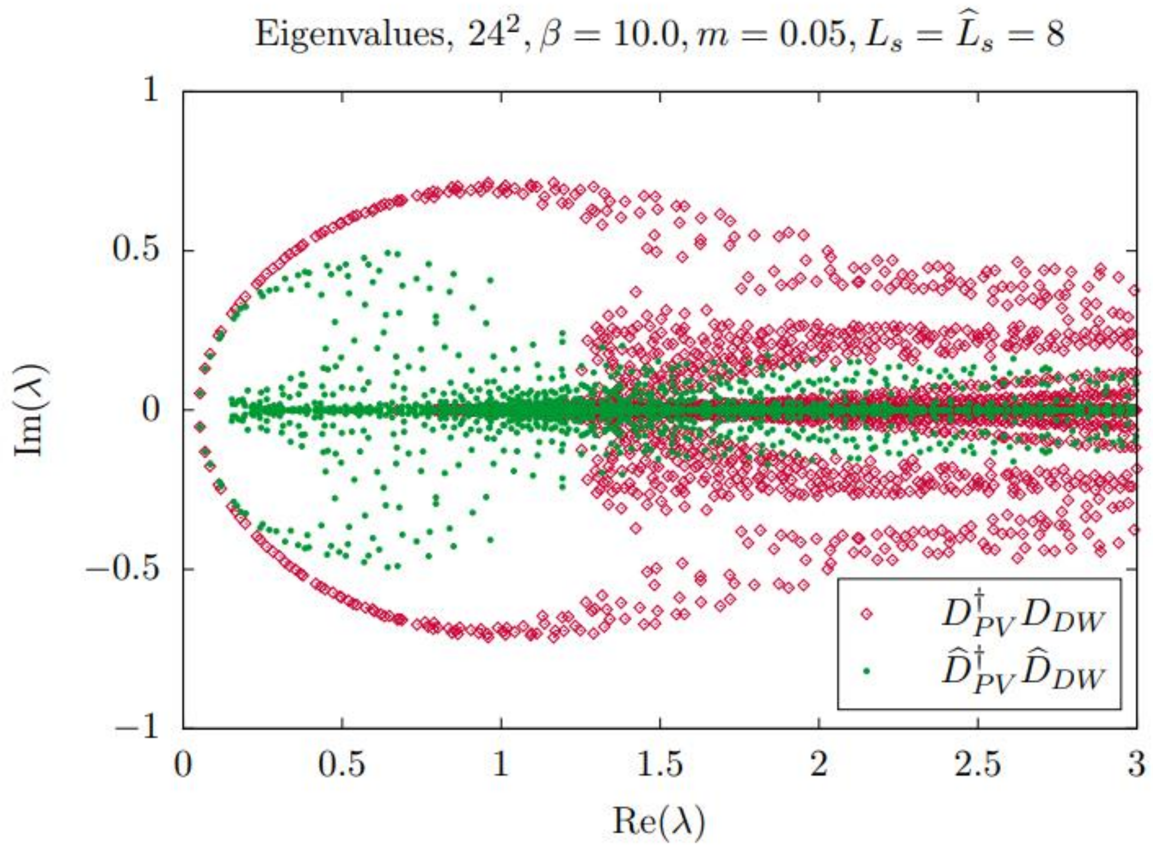


DOMAIN WALL THREE-STEP PLAN

Good things come in threes

- Approximate **Pauli-Villars preconditioning**
  - The overlap operator (Dpvinv Ddw) is the perfect operator
  - Drastic approximation: $D_{PV}^{-1} = D_{PV}^{\dagger}[D_{PV} D_{PV}^{\dagger}]^{-1} \simeq D_{PV}^{\dagger}$
  - Preserves half plane condition
- **Wilson Kernel Galerkin Projection**
  - Coarsen the 4-d kernel, preserve 5-d structure
  - Overlap construction only depends on having a gamma5-Hermitian operator
- **Truncated Prolongation/Restriction**
  - Only prolong/restrict boundary modes
  - Natural thinning of 5-d degrees of freedom





COMMUNITY ASKS AND FUTURE WORK

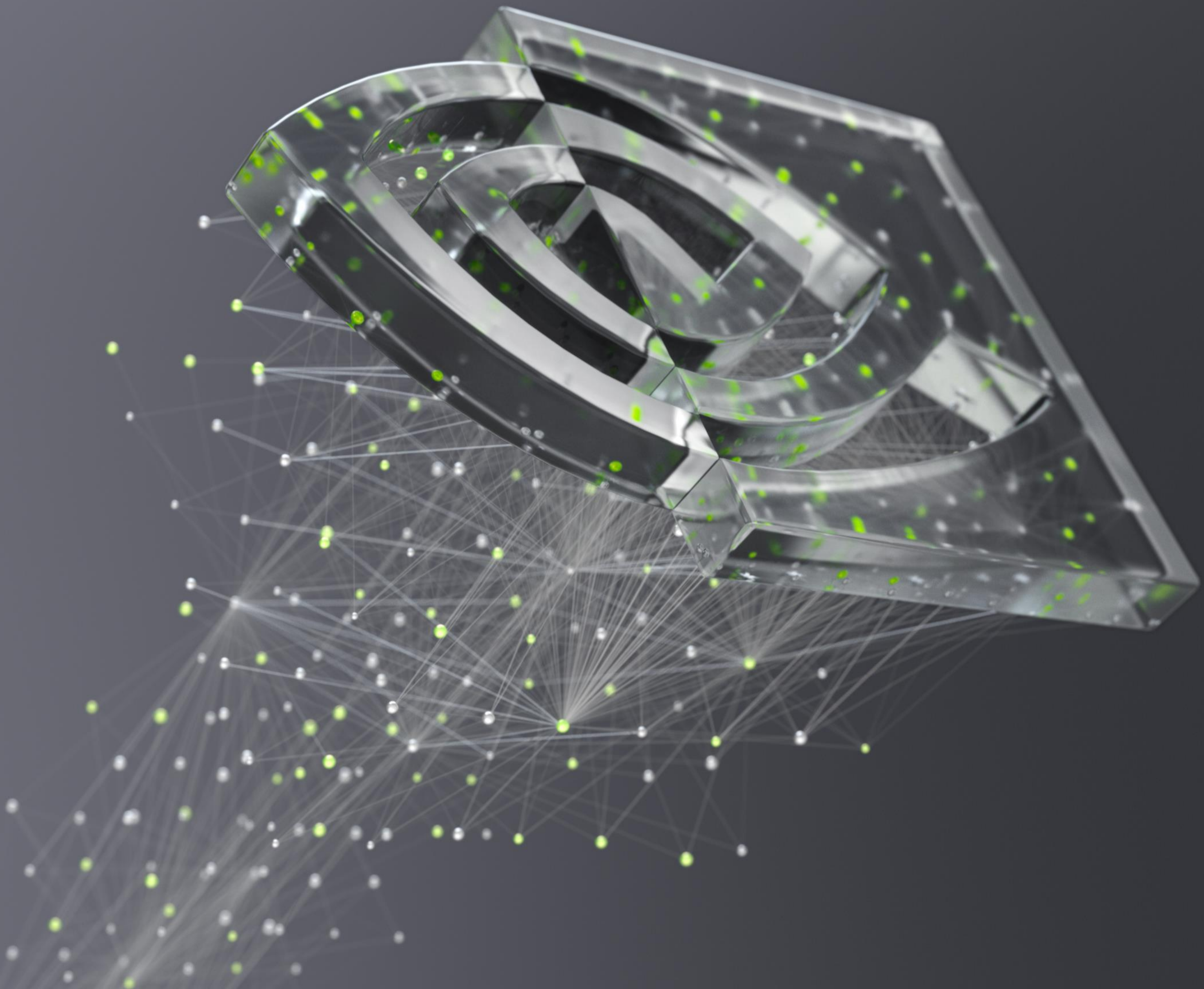
HELP US HELP YOU!

- If you want to try multigrid in your workflow, we want to know and we want to help!
- The best way to reach the QUDA community is our public Slack: quda.slack.com
- If you want to use HISQ MG in particular…
    - *I* want to know and I want to help!
    - I’ve learned a lot about tuning, but that doesn’t mean I know much

FUTURE WORK

There's always something else to do…

- HISQ Multigrid
  - Updating this to the Generic Kernels framework
    - Kate Clark, "Preparing for QUDA 2.0", Software Development Parallel, July 28 @ 1 pm EST
  - Formalizing MILC support (merging into the mainline develop branch)
  - Fused HISQ + KD inverse operator for improved overlap of compute and communications
  - Better guidance on tuning the algorithm
- More broadly…
  - Improved scale calculations for 16-bit fixed point coarse links in QUDA 2.0
  - NVSHMEM support for the coarse operators
  - Domain wall/Mobius MG



NVIDIA