Random Ruminations

Bálint Joó – ORNL, Science Engagement Section

SciDAC-5 Kickoff Meeting

Jefferson Lab, Newport News, VA

Dec 2. 2022

ORNL is managed by UT-Battelle LLC for the US Department of Energy

"Conventional HMC" – a reminder

- Procedure:
  - refresh momenta & pseudo-fermions: { $U(0)$, $\pi(0)$ } and φ_i
  - perform "Molecular Dynamics" : { $U(\tau)$, $\pi(\tau)$ } with fixed φ_i
    - nested 4th order Force-Gradient Integrator with $\tau = N\,\delta\tau$
    - N = number of outermost steps, dt outermost step size
    - constructed of reversible combination of symplectic steps
  - Metropolis Update step: $\delta H = H(U(\tau),\pi(\tau),\varphi_i) - H(U(0),\pi(0),\varphi_i)$
  - Acceptance test: $P_{acc} = \min\{1, \exp(-\delta H)\}$
  - If we accept: $\{U,\pi\} = \{U(\tau),\pi(\tau)\}$; otherwise $\{U,\pi\}=(U(0), \pi(0)\}$
  - Rinse & Repeat until enough U-s are produced
- Typically (apocryphally?):
  - 95% of gauge generation wallclock spent in MD forces...
  - 60-70% of wallclock spent in Linear Solvers

Hasenbusch's Trick and Determinant decomposition

Pseudofermions: $$\det\left(M^\dagger M\right) = \int d\phi^\dagger d\phi \;\exp\left\{-\phi^\dagger\left(M^\dagger M\right)^{-1}\phi\right\}$$

Hasenbusch's Trick:

$$\det\left(M^\dagger M\right) = \left[\prod_{i=0}^{N-1}\frac{\det\left(M_i^\dagger M_i\right)}{\det\left(M_{i+1}^\dagger M_{i+1}\right)}\right]\det\left(M_N^\dagger M_N\right) \quad \text{with} \quad M_0 = M$$

In the exponential:

$$\phi^\dagger\left(M^\dagger M\right)^{-1}\phi = \sum_{i=0}^{N-1}\phi^\dagger M_{i+1}\left(M_i^\dagger M_i\right)^{-1}M_{i+1}^\dagger\phi + \phi^\dagger\left(M_N^\dagger M_N\right)^{-1}\phi$$

Forces:

Linear Solves needed

$$\frac{d}{d\tau}\left[\phi^\dagger\left(M^\dagger M\right)^{-1}\phi\right] = -\phi^\dagger\left(M^\dagger M\right)^{-1}\left[\dot{M}^\dagger M + M^\dagger \dot{M}\right]\left(M^\dagger M\right)^{-1}\phi$$

Hasenbusch Trick works because M_i and M_{i+1} are similar:

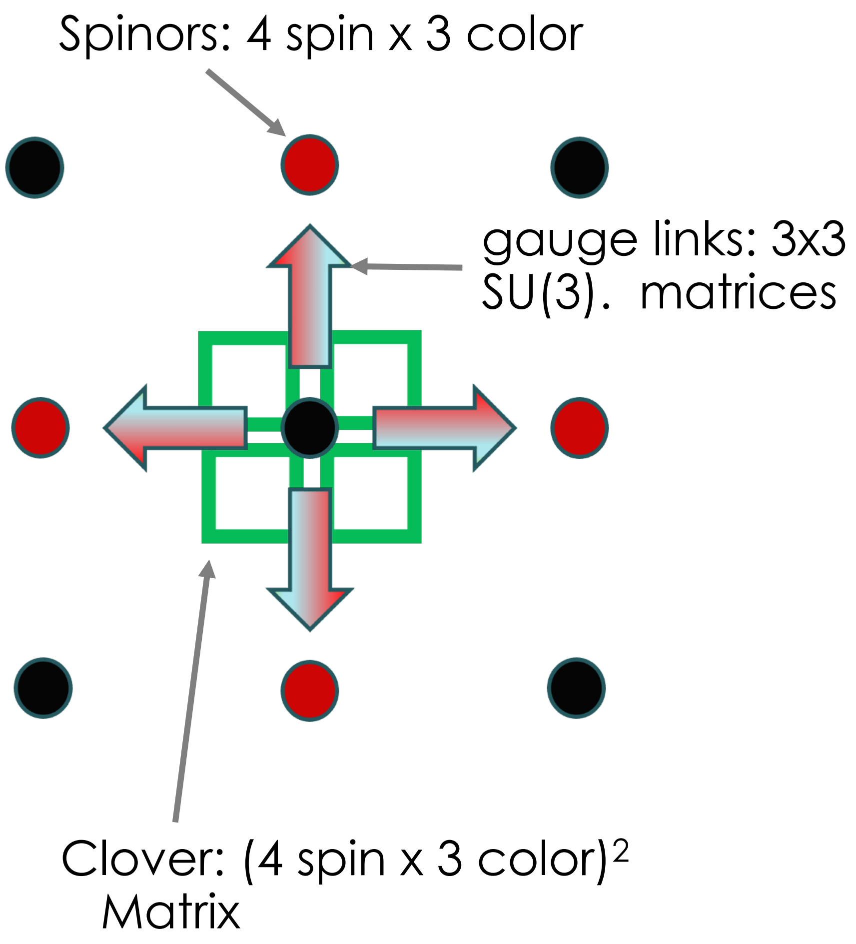
$$M_{i+1} = M_0(m_q + \epsilon_{i+1}) \quad \text{or} \quad M_{i+1} = M_0 - i\mu_{i+1}\gamma_5$$

So

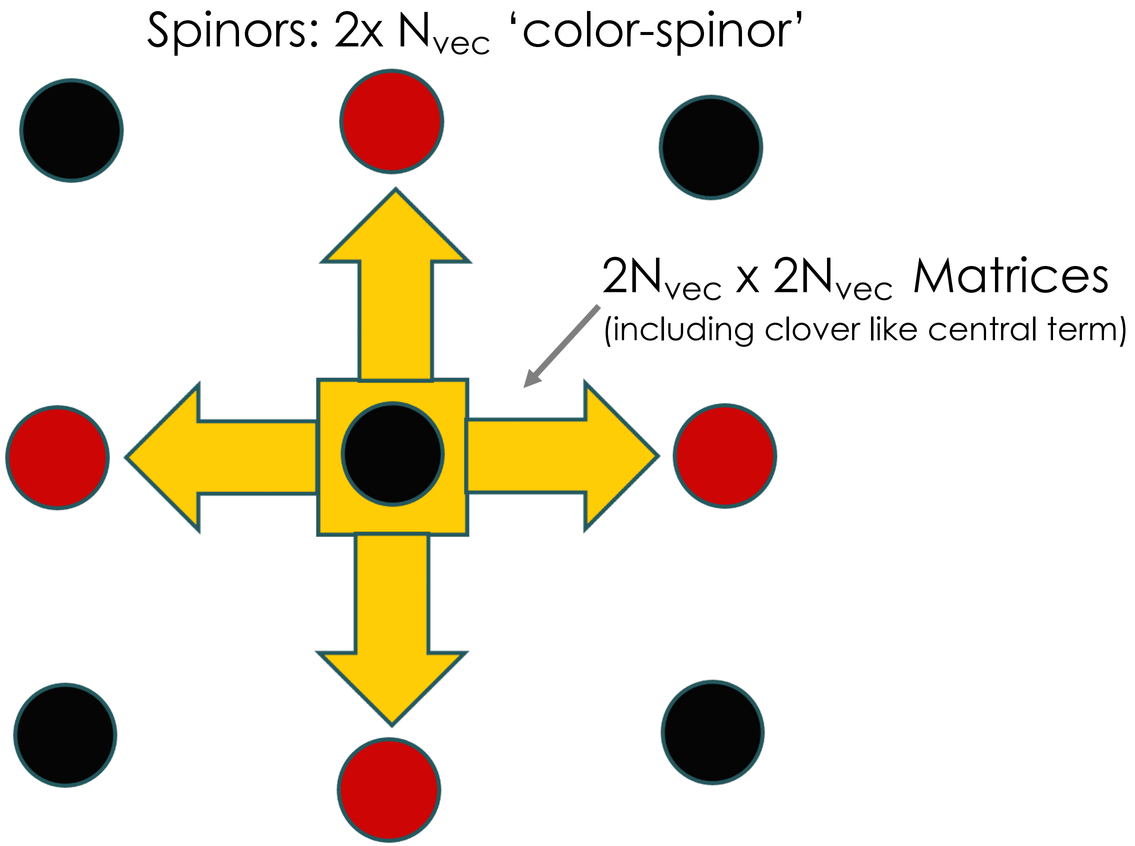
$$M_{i+1}M_i^{-1} \approx 1 + \delta M \quad \Rightarrow \quad F \approx \frac{d\left(\delta M\right)}{d\tau}$$

Control force size with range of masses ϵ_i or μ_i

Operators used in Solving the Dirac equation



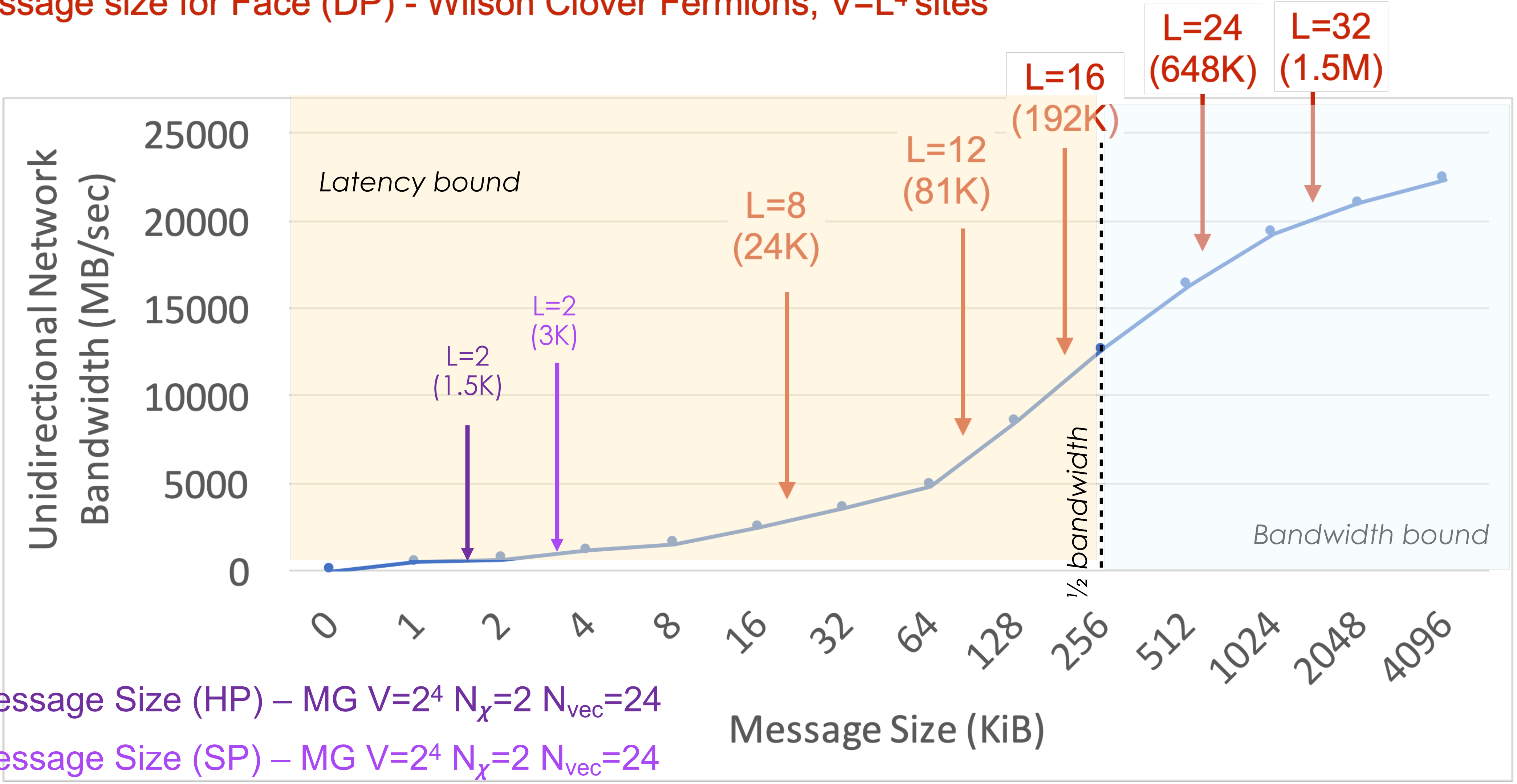
Wilson-Clover Dirac Operator: Fine Lattice, Full Volume: e.g. 128^3x256 global ~14^4-16^4 local (per process)



Multigrid Coarse Operator: Coarse Lattice, Full Volume: e.g. 8^3x16 global ~2^4-4^4 local (per process)

The curse of latency…

→ Message size for Face (DP) - Wilson Clover Fermions, V=L^4 sites

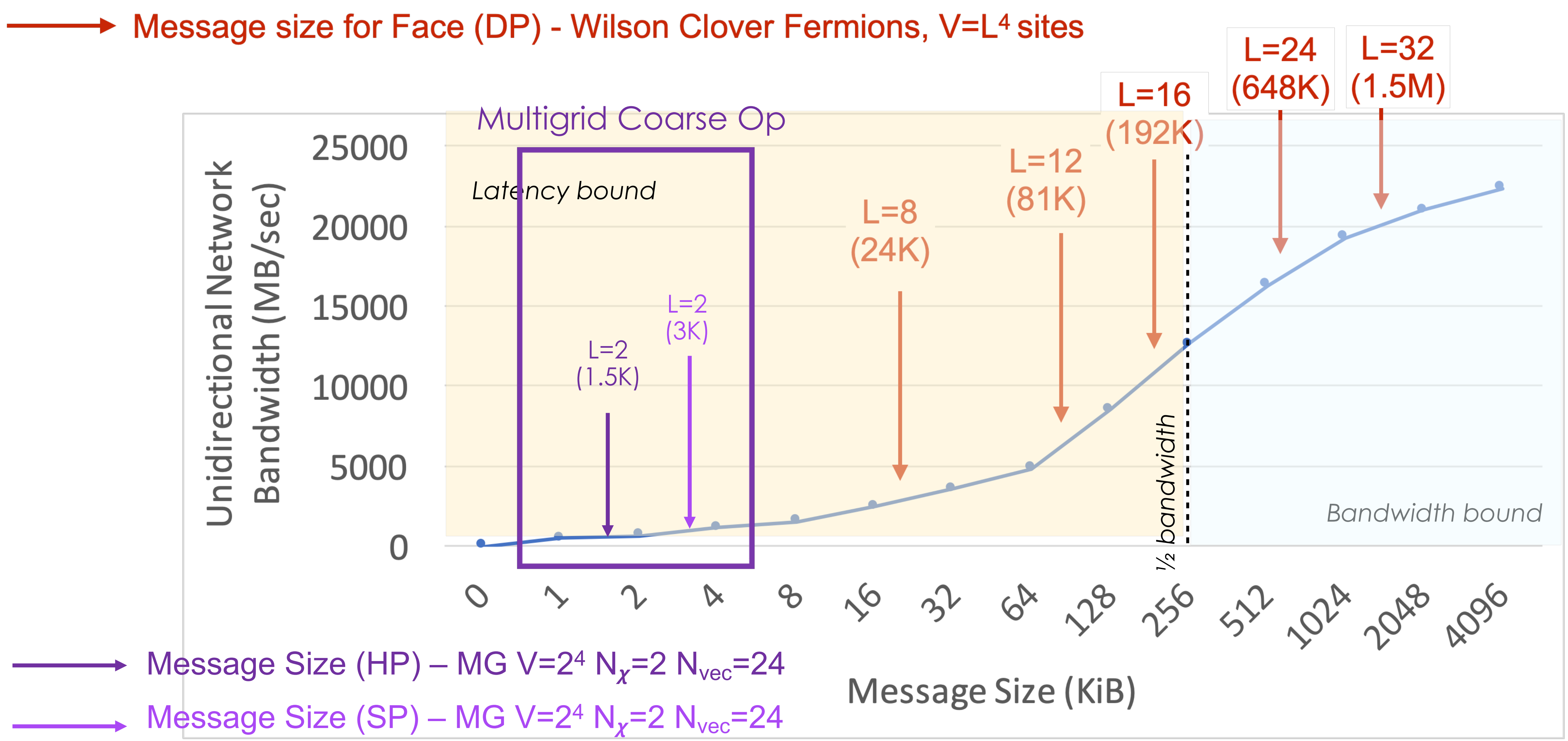


→ Message Size (HP) – MG V=2^4 N_χ=2 N_{vec}=24

→ Message Size (SP) – MG V=2^4 N_χ=2 N_{vec}=24

Data for Summit from SummitWorkshop-SummitNodePerformane-WJ talk by W. Joubert

The curse of latency…



Data for Summit from SummitWorkshop-SummitNodePerformane-WJ talk by W. Joubert

SciDAC 4 attempt to improve Latency issues

- Write $M_j = M - i\mu_j \gamma_5$
- Then: $(M_j^\dagger M_j) = M^\dagger M + \mu_j^2$
- Solve for all terms at once: $X_j = (M_j^\dagger M_j)\phi_j = (M^\dagger M + \mu_j^2)\phi_j$
  - Need Multi-RHS solve
  - For Multigrid we need a 2 step solve. $M_j^\dagger Y_j = \phi_j$, $M_j X_j = Y_j$
  - Multi-RHS => message aggregation => push away from latency bound region in the fabric
  - Block GMRES/GCR will need modification: add +/- $i\mu_j \gamma_5$ to M as needed for solution j
- Other possible trick: Neglect term without $-i\mu_j \gamma_5$ in MD
  - bound spectrum from below: no spikes!!!!!
  - This is kind of like reweighting but a small μ_j may make it feasible
  - Twisted mass messes with operator eigenspectrum, complicates multigrid...
- This is unfinished work from SciDAC-4. Wei Sun made good progress but then COVID etc. etc.

General Problems for Wilson-Clover fermions

- Critical slowing down in Solvers with quark mass
  - mostly cured with Multi-Grid
- Critical slowing down with lattice spacing
  - not yet solved: ML? Otherwise? DD?
- MD Instability
  - Limits on step size (CFL like instability at large step sizes)
    - keep a small step size so that. $|| F || \delta\tau < C$ (where C is some critical value)
  - (Near) Zero modes can develop in the integration: drive $|| F ||$ to be large
- Other numerics:
  - smaller ‘a’ -> **Larger Lattices -> Master Field Simulations**
  - global sums & error accumulation: will we need quad precision for big enough V?
  - dH calculations can suffer rounding error from sums: error in P_{acc} ?

Large Lattice Simulation Tricks (from P. Fritzsch Lat'22)

- Exponential Clover Action:
  - Treat diagonal term of Wilson Clover as first order truncation of an exponential
  - Switch to using the full exponential

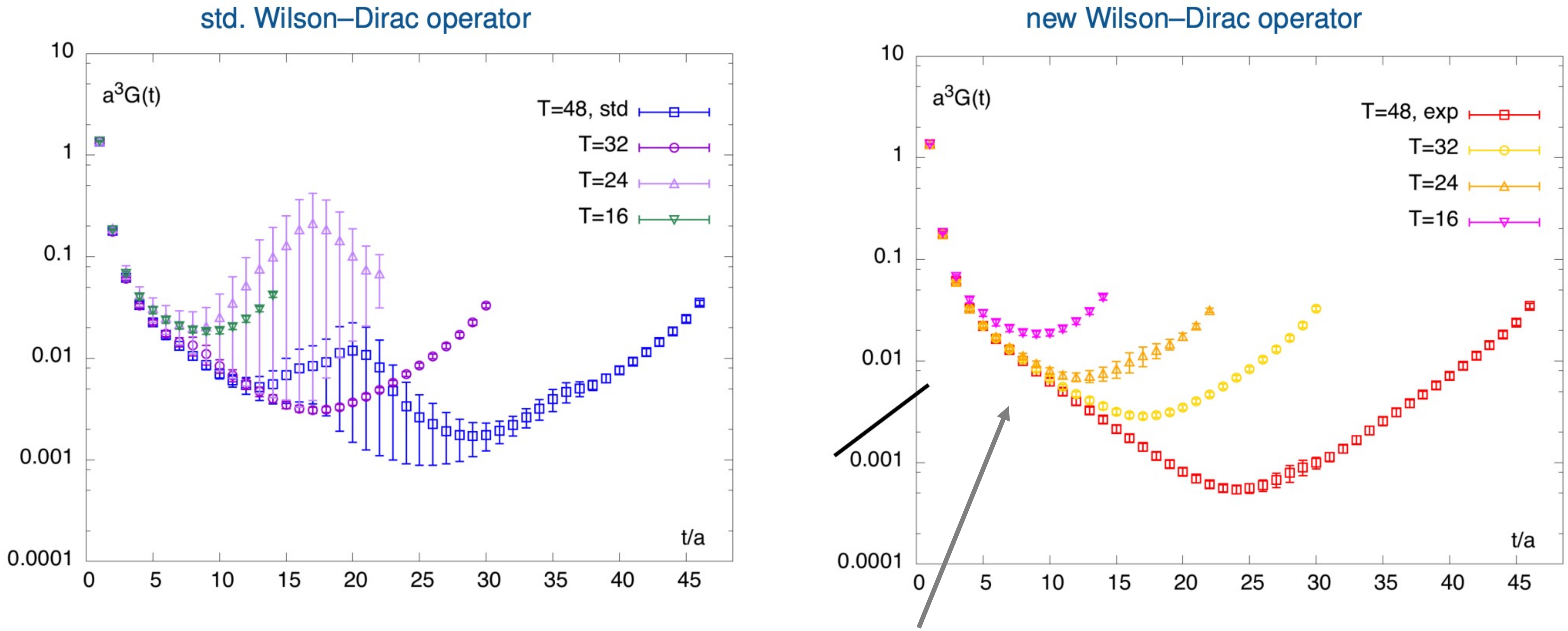
$$(N_d + M) + \frac{i}{4} c_{\mathrm{SW}} \sigma_{\mu\nu} F_{\mu\nu} \rightarrow (N_d + M) \exp\left\{ \frac{i}{4} \frac{c_{\mathrm{SW}}}{(N_d + M)} \sigma_{\mu\nu} F_{\mu\nu} \right\}$$

  - Claimed benefits:
    - regulates UV fluctuations
    - guarantees invertibility
    - Still a valid Symanzik improvement
  - Simple(?) to implement – use Cayley Hamilton theorm to exponentiate
    - Also need to compute forces and Tr Log
    - NB: Currently 'Clover' implementation is a 'special' in Chroma (outside of QDP++)
- Stochastic MD algorithm instead of HMC

Exponential Clover (cont'd): Quenched approx testing

from P. Fritzsch Larrice'22 presentation

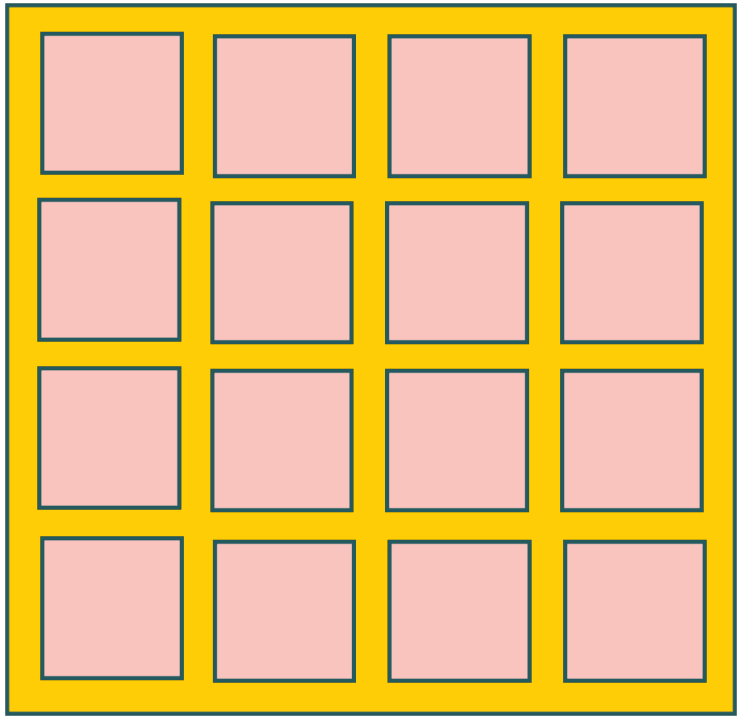
- Impact best seen in pure gauge theory ($N_\mathrm{f} = 0$, quenched; i.e. same gauge background). Ill-defined theory for fermionic observables. $\Rightarrow$ exceptional problems
- Different lattices $L/a \in \{16, 24, 32, 48\}$ and same gluon action ($\beta = 6.0$, $a = 0.094\,\mathrm{fm}$).
- pion correlator $G(t) \propto e^{-m_\pi t}$ at zero momentum, $m_\pi \approx 220\,\mathrm{MeV}$



exponential clover: no exceptional configs, clean pion correlator

Thoughts about DD and Hierarchical DD

- Regular 'DD' a-la Leuscher is feasible with suitable support for Domains in QDP++ (Frank's talk)
- To some degree, everything we have done up to now will be restricted to domains
  - large domains -> map to GPUs (following P. Boyle)
- Biggest pain point for DD & Hierarchical DD likely to be dealing with domain boundaries and 'window frame' areas (yellow)
  - Luescher style DD will still need solves on boundaries.
  - reduced dimension compared to global lattice volume (good)
  - For Hierarchical methods: may need to simulate with the Multiboson algorithm. (Is this practical?)
- Historical perspective: Multibosons were being investigated when I was a graduate student as a competitor to HMC.
  - Multibosons were tought complicated to implement and a nightmare to tune back in the late 1990s. HMC 'won'.
  - Do we really want Multibosons again?



Random Summary Thoughts from Me / Discussion points...

- This is the last presentation of the workshop .... so time to poke the bear !
- Do we want to finish the leftover work from SciDAC-4? Is it worth it or is it a distraction from the DD world?
- I think the future may well be Domain Decomposed (even if not hierarchical)
  - good match for accelerated architectures
  - we might as well add in things like Exp. Clover and SMD (will we need them for DD?)
  - BUT: I do note that there is a large ensemble of Iso Clover ensembles in USQCD right now to work with. Changing actions is 'changing horses' mid race...
- Does the DD world still fit Chroma?
  - Chroma = HMC + Props + Basic measurements
  - Hierarchical approach will need observables folded into the Gauge Generation? Need all new measurement code too.
  - How does DD work with ML etc?

... a new software stack for Lattice QCD?

- Designed ground up to support Domains natively?
  - for Multigrid, DD methods, etc?
- NOT using QMP - but going straight to MPI
- NOT using QIO – but using e.g. HDF5 (C++ interface)
- Written in terms of a performance portable abstraction (Kokkos? C++/std::par? Multiple Back Ends?)
- ... or is this package simply called future QUDA / Grid ?
- ... will this software integrate the ML work? or be completely separate?

Acknowledgement

- This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725
- This research is supported by SciDAC-5