

algorithms & machines

Xiao-Yong Jin - Argonne National Laboratory - July 27, 2026 - Lattice 2026 - UMD

machines

machines in the 1st lattice conference

31 contributions, 102 participants

Gauge Theory on a Lattice: 1984

April 5-7, 1984

CONF-8404119--

DE84 014856

**Proceedings
of the Argonne National Laboratory
Workshop**

Lattice Gauge Theories on a Hypercube Computer

Steve W. Otto

California Institute of Technology, Pasadena, California, 91125

PARALLEL SUPERCOMPUTERS AND LATTICE GAUGE THEORIES

Anthony Ierrano
Physics Department
Columbia University
New York, NY 10027

DISCUSSION PAPER FOR A HIGHLY PARALLEL ARRAY PROCESSOR-BASED MACHINE

Ray Hagstrom, Gary Bolotin, and John Dawson
Argonne National Laboratory
High Energy Physics Division
Argonne, Illinois 60439

the 3rd lattice conference

30 contributions, 98 participants

This volume presents the contributions to the international workshop entitled "Lattice Gauge Theory - a Challenge in Large Scale Computing" that was held in Wuppertal from November 4 to 7, 1985. This meeting was the third in a series of European workshops in this rapidly developing field.

field theories has grown considerably during the past years. In fact, quite a few prominent colleagues of ours have even gone into machine building, both in industry and research institutions: more parallelism, and more dedicated computer architecture are their design goals to help them boost the Megaflop rate in their simulation processes. The workshop contained several interesting seminars with status reports on such supercomputer projects like the Italian APE (by E. Marinari), the IBM project GF-11 (by D. Weingarten), and the Danish projects MOSES and PALLAS (by H. Bohr).

MEMORY AND LEARNING IN A CLASS OF NEURAL NETWORK MODELS
D.J. Wallace
Physics Department,
The University of Edinburgh
Edinburgh EH9 3JZ, U.K.

lattice 90

QCD MACHINES — PRESENT AND FUTURE *

Norman H. CHRIST

Department of Physics, Columbia University, New York, NY 10027

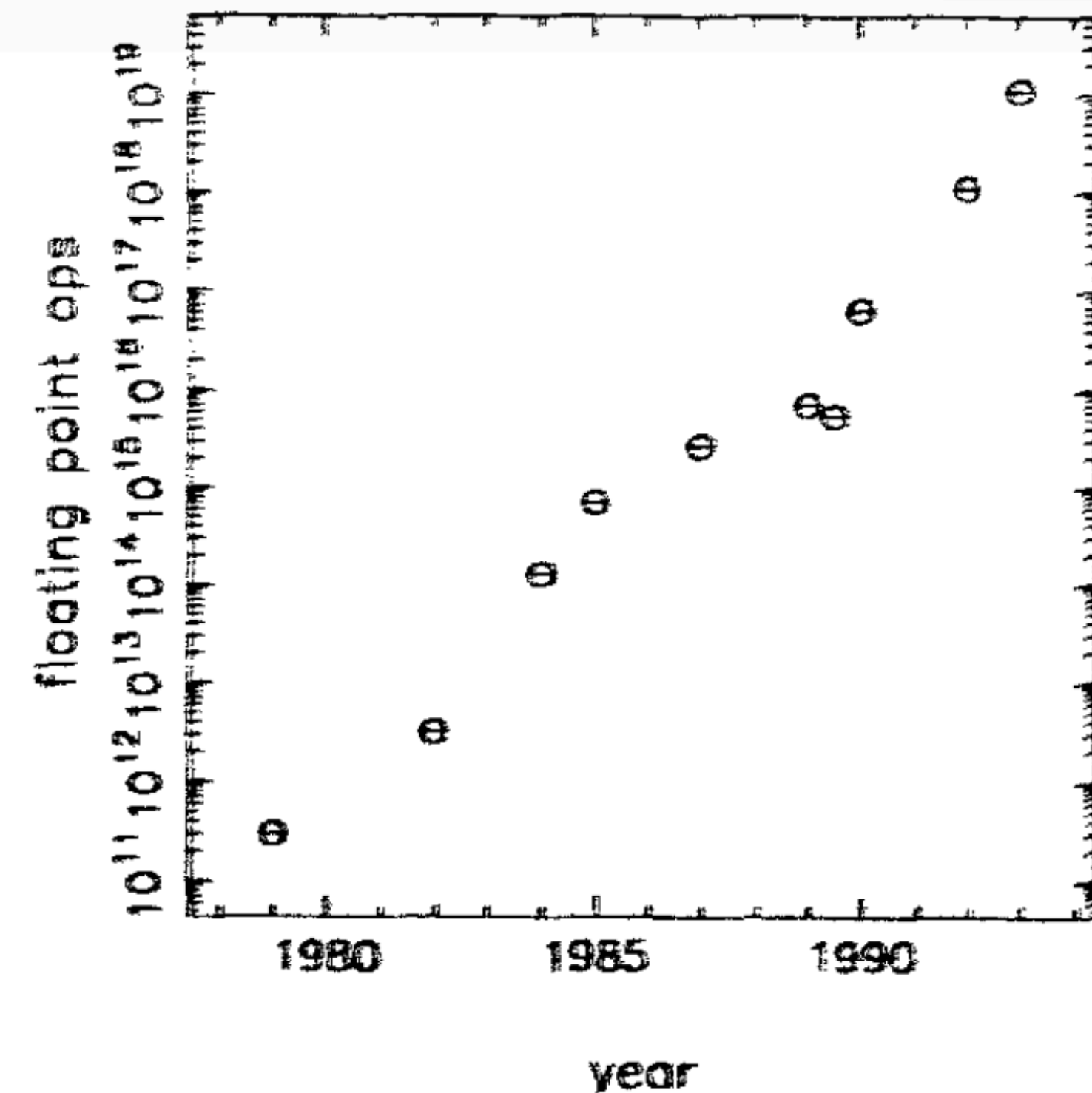


Figure 1: The size of a typical, large-scale, lattice QCD calculation (represented by total floating operations performed) is plotted versus year.

tion of the time. While the early growth in computer power shown in the figure reflects calculations that used an increasing fraction of the currently available computer power, later increases required advances in supercomputer performance and the appearance of special purpose, parallel machines. The exponential growth shown in the figure is maintained through 1993 by the final two hypothetical points corresponding to APE100 and the QCD Teraflop Project.

lattice 91

Parallel QCD Machines

Don Weingarten

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

machine	Mflops	cost (M\$)	flops/\$	ops/year	years/ 10^{17} ops
Vax 11/780	0.25	0.25	1.	$6. \times 10^{12}$	16000
CDC 7600	2.0	2.0	1.	$6. \times 10^{13}$	1600
Cray 1	40.	10.	4.	1.2×10^{15}	80

Table 1: Estimates of the sustained performance, cost and work per year of typical machines used for numerical QCD in the early 1980's

machine	year	nodes	Mflops	regs.	Mby sram	Mby dram	utilization
Columbia	1985	16	16	0	0.128	1.0	20%
	1987	64	16	18	0.128	2.0	20%
	1989	256	64	192	2.0	0	40%
GF11	1991	470	20	256	0.064	2.0	80%
APE	1986	4	64	128	0	8	20%
	1988	16	64	128	0	16	20%

Table 2: Completion date, node count, node arithmetic peak, node memory characteristics, and ratio of sustained to peak arithmetic rate for first generation QCD machines.

machine	year	nodes	Mflops	regs.	Mby sram	Mby dram	utilization
ACPMAPS	1989	16	20	32	0	8.0	20%
	1991	256	20	32	0	8.0	20%
QCDPAX	1991	432	30	2	2.0	4.0	20%

Table 3: Completion date, node count, node arithmetic peak, node memory characteristics, and ratio of sustained to peak arithmetic rate for second generation QCD machines.

machine	architecture	peak	sustained	memory	cost	flops/\$
		Gflops	Gflops	Gbytes	M\$	
CM2	SIMD	32	6.5	2	7	900
Meiko	MIMD	5	1.5	0.5	1	1500
AP1000	MIMD	4	1.5	8	?	?
iPSC/860	MIMD	5	1.	0.5	1.9	500

Table 4: Estimates of the characteristics of some commercial machines used for numerical QCD

lattice 93

Computers for Lattice Field Theories

Y. Iwasaki ^a

^aInstitute of Physics, University of Tsukuba, Ibaraki 305, Japan

Table 1
List of dedicated QCD computers

Project	Peak speed Gflops	year
Columbia 16	0.25	1985
64	1.0	87
256	16	89
APE 4	0.25	86
16	1.0	88
QCDPAX	14	90
GF11	11	91
ACPMAPS	5	91
APE100	6($\rightarrow$ 100)	92($\rightarrow$ 94)
ACPMAPS upgraded	50	93
Teraflops	1,600	96
CP-PACS	$\geq$ 300	96
0.5Teraflops	800	95
APE1000	$\sim$ 1000	

Table 2
Characteristics of dedicated QCD computers I

Project	Columbia	APE	QCDPAX	GF11	ACPMAPS
peak speed	16 Gflops	1 Gflops	14 Gflops	11 Gflops	5 Gflops
processors	256	16	480	566	256
network	2d torus	linear array	2d torus	Memphis switch	crossbar and hypercube+
arichi- tecture	MIMD	SIMD	MIMD	SIMD	MIMD
CPU	80286	—	68020	—	Weitek
FPU	80287	Weitek1032 $\times$ 4	LSI Logic	Weitek1032 $\times$ 2	XL8032
	Weitek3364 $\times$ 2	Weitek1033 $\times$ 4	L64133	Weitek1033 $\times$ 2	chip set
SRAM	2MB	—	2MB	64KB	2MB
DRAM	8MB	16MB	4MB	2MB	10MB
speed/ processor	64Mflops	64Mflops	32Mflops	20Mflops	20Mflops
host	VAX11/780	μ VAX	Sun 3/260	3090	μ VAX

Table 3
Characteristics of dedicated QCD computers II

Project	APE100	ACPMAPS
processors	2048	612
arichi- tecture	SIMD	MIMD
CPU	MAD (custom)	i860
memory	4MB	32MB
speed/ processor	50 Mflops	80 Mflops
network	3d torus	crossbar hypercube+
host	SUN WS	SGI
peak speed	100 Gflops	50Gflops
arithmetic	32 bit	32 (64) bit

Table 4
Characteristics of dedicated QCD computers III

Project	Teraflops	CP-PACS	0.5Tflops
processors	8K	1–1.5K	16K
arichi- tecture		MIMD	MIMD
CPU		enhanced PA-RISC	DSP TI
memory	32MB	64MB	2MB
speed/ processor	200–300 Mflops	200–300 Mflops	50 Mflops
network		hypercrossbar	4d torus
host		main frame	SUN WS
peak speed	$\geq$ 1.6Tflops	$\geq$ 300Gflops	0.8Tflops
arithmetic	64bit	64 bit	32 bit

lattice 95

Computers for Lattice Simulation

James C. Sexton ^a

^aSchool of Mathematics, Trinity College, Dublin 2, Ireland

Table 1

A summary of special purpose lattice QCD machines presently in active development. Performance estimates are only approximate, and the cost per Gf number refers to sustained performance.

	APE 100	APE Mille	CP-PACS	QCDSP
Architecture	SIMD	SIMD	MIMD	MIMD
Processor	MAD	Tarzan/Jane	PA-RISC	DSP
Clock	25 MHz	50 MHz	150 MHz	25 MHz
Peak/Proc	50 Mf	200 Mf	300 Mf	50 Mf
Precision	32	64	64	32
Memory/Proc	1 Mw	4/8 Mw	8 Mw	0.5 Mw
Memory/Flop	0.02 Mw/Mf	.02/.04 Mw/Mf	0.027 Mw/Mf	0.01 Mw/Mf
Network	3D NN	3D "Global"	3D crossbar	4D NN
Max Procs	2048	4096	1024	16384
Peak	102 Gf	820 Gf	307 Gf	820 Gf
Cost	\$12M to buy	\$13M	\$15M	\$3M
Performance	55%	55%	60%	40%
Cost per Gf	\$0.2M/Gf	\$0.03M/Gf	\$0.08M/Gf	\$0.01M/Gf
Commercial?	yes	yes	yes	no

Table 2

Modern supercomputers at a glance.

Company	Machine	Proc	Proc Peak	Maximun Procs	Machine Peak	Maximum Memory	Network
Cray	C90	vector	1 Gf	16	16 Gf	8 Gb	shared memory
Cray	T90	vector	1.8 Gf	32	58 Gf	8 Gb	shared memory
Fujitsu	VPP500	vector	1.6 Gf	220	352 Gf	32 Gb	full crossbar
Cray	T3D	Alpha 21064	0.15 Gf	2048	300 Gf	32 Gb	3D
IBM	SP2	Power 2	0.264 Gf	128	34 Gf	256 Gb	special
Convex	SPP1000	PA-RISC	0.2 Gf	128	25 Gf	32 Gb	shared memory
Cray	J90	custom	0.2 Gf	32	6.4 Gf	8 Gb	shared memory
Digital	Alpha 8400	Alpha 21164	0.6 Gf	12	7.2 Gf	14 Gb	shared memory
SGI	Power	R8000	0.36 Gf	18	6.48 Gf	16 Gb	shared memory

Table 3

Rough estimates of machines and computer power presently available to a representative selection of lattice QCD groups around the world.

Country	Group	Machines	Power
Germany	various	3 QH2	21 Gf
Italy	Roma I	4 QH4	55 Gf
	Roma II	QH4	14 Gf
	Pisa	QH4, 2 QH1	21 Gf
	Parma	2 QH1	7 Gf
Japan	JLQCD	VPP500(80)	50 Gf
	Tsukuba	various	10 Gf
	Riken	VPP500(30)	
UK	UKQCD	various	2.5 Gf
US	Columbia	256 node	7 Gf
	GF11	GF11	5 Gf
	Fermilab	ACPMAPS	5 Gf
	LANL	CM5	5 Gf
	SCRI	CM2	5 Gf
	MILC	various	5 Gf
	NRQCD	various	1 Gf

lattice 99

A summary of computer resource presently available for lattice QCD. These are rough estimates which attempt to quantify available computer resources in average, sustained Gigafllops.

Country	Group	Lattice '95		Lattice '99	
		Machine	Power	Machine	Power
<u>Germany:</u>	various	3QH2	21	QH4+QH2	21
<u>Italy:</u>	RomaI	4QH4	55	4QH1+2QH2	28
	RomaII	QH4	14	1QH1	3
	Pisa	QH4+2QH1	21	2QH1	7
	Other	2QH1	7	NQH1	16
<u>Japan:</u>	Tsukuba	various	10	CP-PACS	300
	JLQCD	VPP500(80)	50	VPP500(80)	50
	RIKEN	VPP500(30)	10	VPP800E	30
<u>UK:</u>	UKQCD	various	2.5	T3E/152	41
<u>US:</u>	Columbia	256-node	7	QCDSP	120
	RIKEN/BNL	-----		QCDSP	180
	Fermilab	ACPMAPS	5	ACPMAPS	5
	IBM	GF11	5	-----	
	LANL	various	5	Origin2000	3.6
	MILC	various	5	various	6

Computers for Lattice QCD

Norman H. Christ^a *

^aDepartment of Physics, Columbia University, New York, NY 10027 USA

lattice 2003

Recent Development of Machines for Lattice QCD

Thomas Lippert

Department of Physics
University of Wuppertal

7. Overview (large systems)

USA	JLAB	FNAL	QCDOC
Q2 2003	128 XEON nodes Myrinet	128 +48 XEON nodes Myrinet	prototype CPU and daughterboard
end 2003	256 nodes GigE grid	256 nodes GigE grid or Myr	5k nodes
2004	256 nodes	256 nodes	5 susTflops (RIKEN), 2.5 susTflops Columbia, funding for 10 susFlops for US community (Brookhaven)
2005	3-4 susTflops		
2006	5-6 susTflops	5-6 susTflops?	
UK			QCDOC
2004			5 susTflops
Japan	KEK	CCP	ES Center
Q2 2003	64 nodes Hitachi SR8000/ F1 (768 Gflops for QCD)	CPPACS 614 Gflops + Hitachi SR800/ G1 172.8 Gflops + Fujitsu VPP5000, 230 Gflops for QCD	Earth simulator, 640 Gflops for QCD
2006	upgrade to 10- 20 Tflops		
2007		upgrade to 10- 20 Tflops	
Italy	APEmille	apeNEXT	
Q2 / 2003	1024 nodes	sign - off processor	
end 2003		256 nodes	
2004		Several susTflops	
Germany	APEmille	apeNEXT	Cluster (Wuppertal)
Q2 / 2003	1024 nodes		128 proc AliCE
end 2003			1024 proc AliCEnext
end 2004		12 susTflops	

Lattice 2003, Tsukuba, July 15-19, 2003

Thomas Lippert, Wuppertal University

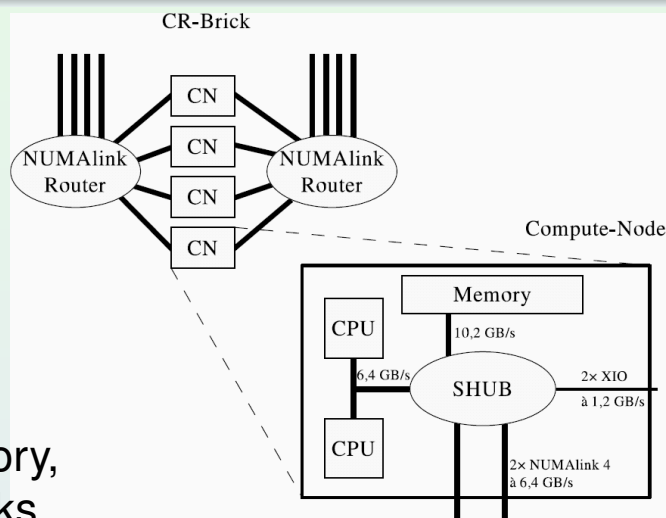
lattice 2005

Performance of machines for lattice QCD simulations

Tilo Wettig

Institute for Theoretical Physics
University of Regensburg

SGI Altix

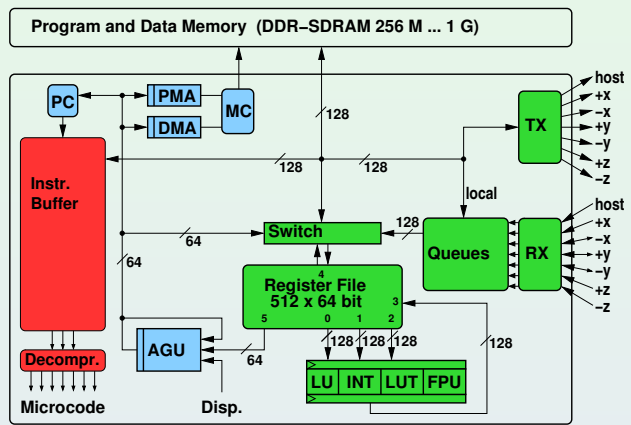
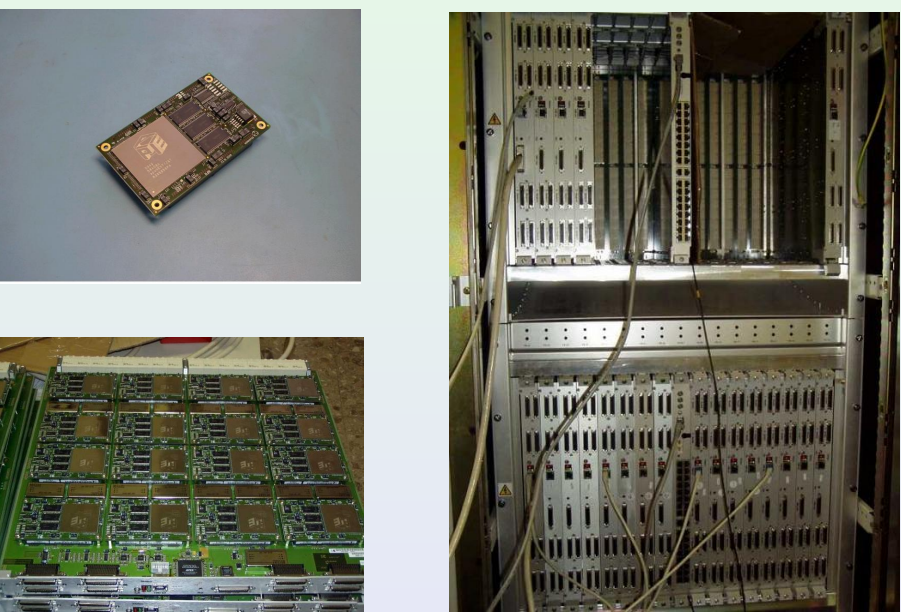


- based on Itanium-2 processor
- compute node: 2 CPUs, 8 GB memory, S-HUB, ccNUMA links
- connected by fat tree (shmем up to 512 CPUs)
- 3.2 μ s SGI-MPT latency
- weak scaling results for Wilson-Dslash ($V_{\text{local}} = 4^4$, fits in L3 cache)

# CPUs	V_{global}	sustained perf.
8	$8^3 \times 4$	31%
16	8^4	26%
32	$8^3 \times 16$	30%
64	$8^3 \times 32$	28%

source:
Thomas Streuer

apeNEXT system



- successor to APEmille
- collaboration of INFN/DESY/Orsay
- custom-designed processor (J&T)
- 8 Flops per cycle (complex $a \times b + c$) at 160 MHz $\rightarrow$ 1.3 GFlop/s peak
- 4 kB on-chip register file
- memory controller and communications hardware on chip
- 3-d torus network, DMA

Large PC cluster installations

existing lattice QCD clusters with **more than 1 TFlop/s peak**

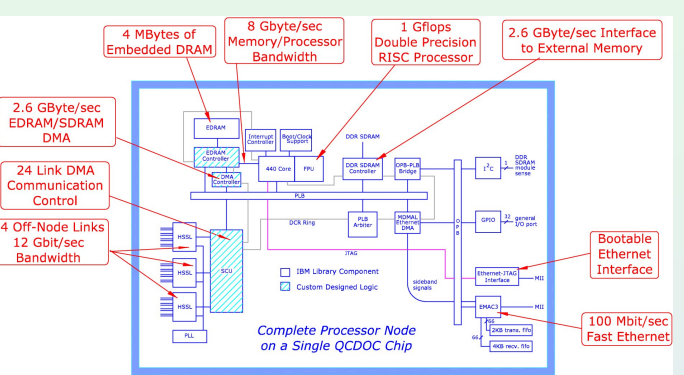
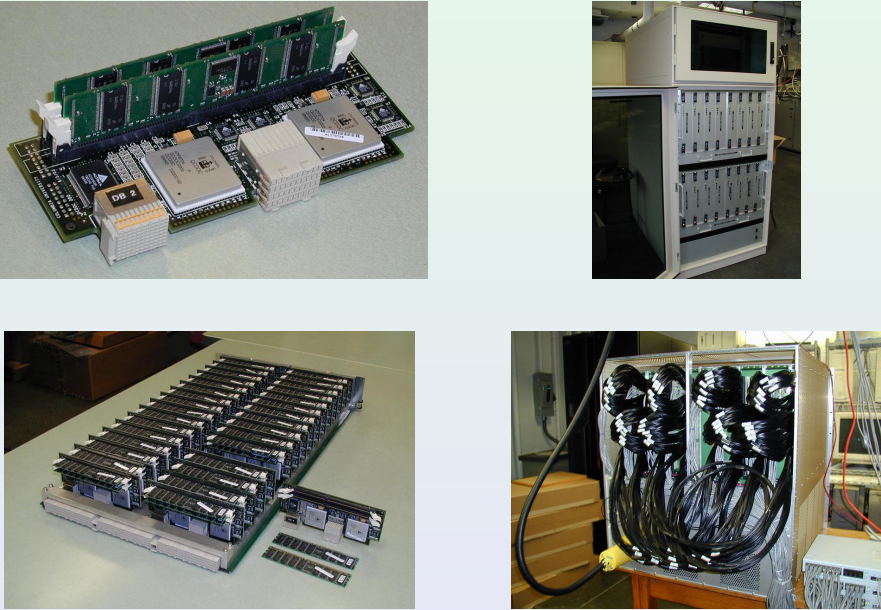
	CPU	Network	Peak (TFlop/s)	Name
Wuppertal	1024 Opteron	Gig-E (2d)	3.7	ALICEnext
JLAB	384 Xeon	Gig-E (5d)	2.2	4G
JLAB	256 Xeon	Gig-E (3d)	1.4	3G
Fermilab	260 (520) P4	Infiniband	1.7 (3.4)	Pion
Fermilab	256 Xeon	Myrinet	1.2	W

- NB:
- peak numbers are double precision
 - all benchmarks are single precision

BlueGene/L overview

- currently #1 and #2 on the Top 500 list (183 TFlop/s peak at LLNL, 115 TFlop/s peak at IBM Watson)
- grew out of the QCDOC project
- system-on-a-chip design (ASIC)
 - 2 PowerPC 440 cores and 4 FPUs at 700 MHz $\rightarrow$ 5.6 GFlop/s peak
 - 32+32 kB L1 cache (I/D) per core, 2 kB L2 cache per core (prefetch), 4 MB shared L3 cache
- distributed memory (512 MB DDR per chip)
- network:
 - 3-d torus with nearest-neighbor links and virtual cut-through routing
 - global tree network for global operations
 - no DMA
- two modes of operation:
 - **co-processor mode**: one CPU for computation, one for communication $\rightarrow$ peak performance 2.8 GFlop/s per chip
 - **virtual-node mode**: both CPUs for computation and communication $\rightarrow$ 5.6 GFlop/s, but communication cannot overlap with computation

QCDOC system



- successor to QCDS
- collaboration of Columbia/UKQCD/RBRC/IBM
- custom-designed ASIC
- PowerPC 440 core + 64-bit FPU
- 2 Flops per cycle at 400 MHz $\rightarrow$ 0.8 GFlop/s peak
- 4 MB on-chip memory
- memory controller and communications hardware on chip
- 6-d torus network, DMA

lattice 2008

GPUs play the game

Recent algorithm and machine developments for lattice QCD

GPGPU

- “Lattice QCD as a video game”,

G.I.Egri, Z.Fodor, S.D.Katz, D.Nogradi,

K.K.Szabo, hep-lat/0611022.

- NVIDIA G80 arch. > 300 GFlops(SP)
- Lattice Wilson kernel > 30 GFlops
- Difficult to program using Graphic API (OpenGL)

- NVIDIA provides HPC GPGPU language

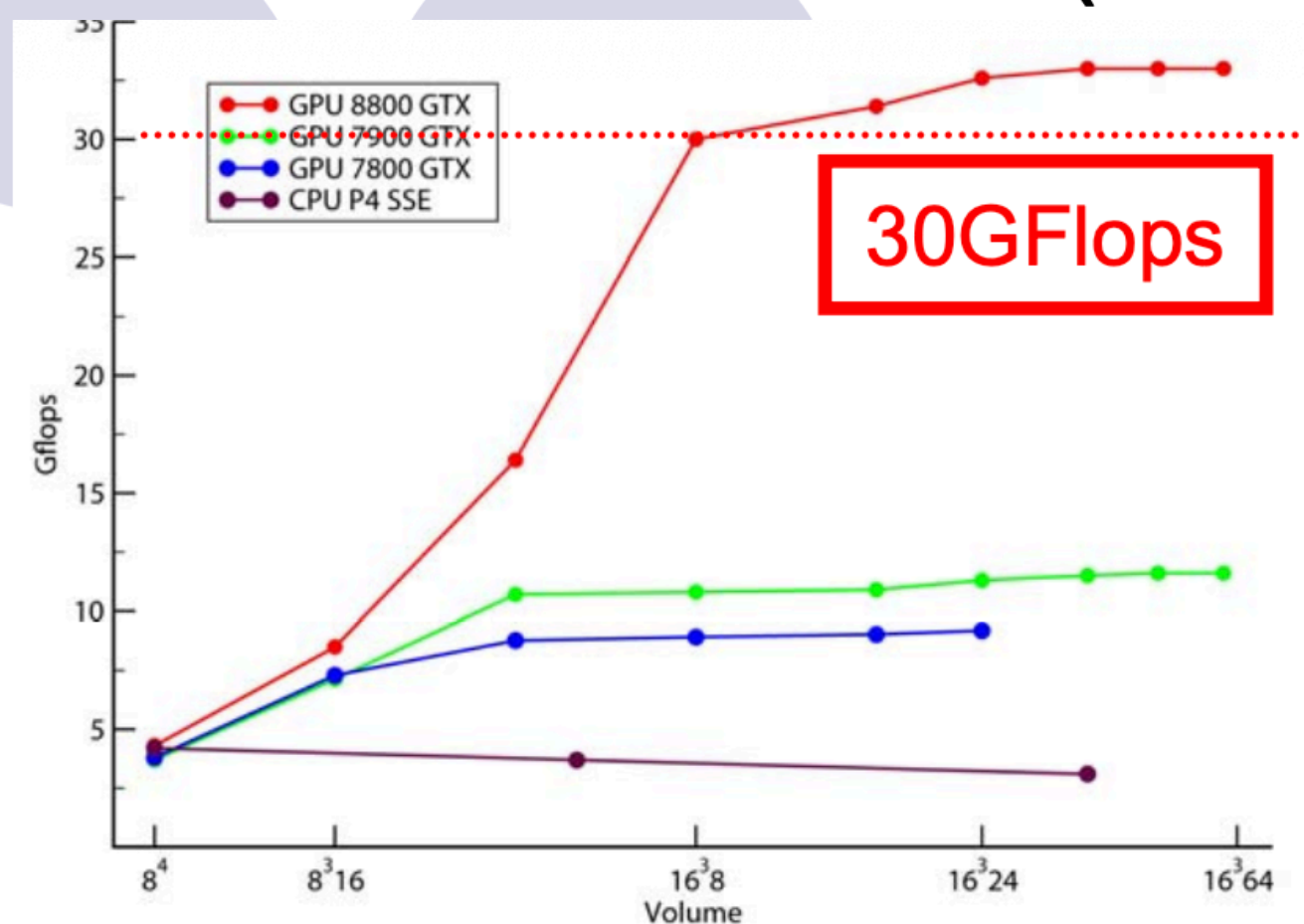
- CUDA (a C/C++ simple extension)
- Easy to learn, but requires hardware/memory model knowledge

- My experience with CUDA (GeForce 8800 GTX)

[**NO WARRANTY CUDA code**:http://theo.phys.sci.hiroshima-u.ac.jp/~ishikawa/CUDA/CudaQCDSolver_0.06.tar.gz]

- Hopping matrix mult (16^4) can also achieve > 40 GFlops.

Ken-Ichi Ishikawa (Hiroshima Univ.)



[Poster by C. Rebbi, “Blastign Through Lattice Calc. using CUDA”
talk by F. Di Renzo, “GPU computing for 2-d spin systems: CUDA vs OpenGL”]

C. Rebbi (Poster):
Wilson Dirac 62GFlops!
with Nvidia Tesla C870

lattice 2016

Peter Boyle
University of Edinburgh
Alan Turing Institute
Brookhaven National Laboratory

System attributes	NERSC Now	OLCF Now	ALCF Now	NERSC Upgrade	OLCF Upgrade	ALCF Upgrades	
Name Planned Installation	Edison	TITAN	MIRA	Cori 2016	Summit 2017-2018	Theta 2016	Aurora 2018-2019
System peak (PF)	2.6	27	10	> 30	150	>8.5	180
Peak Power (MW)	2	9	4.8	< 3.7	10	1.7	13
Total system memory	357 TB	710TB	768TB	~1 PB DDR4 + High Bandwidth Memory (HBM) +1.5PB persistent memory	> 1.74 PB DDR4 + HBM + 2.8 PB persistent memory	>480 TB DDR4 + High Bandwidth Memory (HBM)	> 7 PB High Bandwidth On-Package Memory Local Memory and Persistent Memory
Node performance (TF)	0.460	1.452	0.204	> 3	> 40	> 3	> 17 times Mira
Node processors	Intel Ivy Bridge	AMD Opteron Nvidia Kepler	64-bit PowerPC A2	Intel Knights Landing many core CPUs Intel Haswell CPU in data partition	Multiple IBM Power9 CPUs & multiple Nvidia Voltas GPUS	Intel Knights Landing Xeon Phi many core CPUs	Knights Hill Xeon Phi many core CPUs
System size (nodes)	5,600 nodes	18,688 nodes	49,152	9,300 nodes 1,900 nodes in data partition	~3,500 nodes	>2,500 nodes	>50,000 nodes
System Interconnect	Aries	Gemini	5D Torus	Aries	Dual Rail EDR-IB	Aries	2 nd Generation Intel Omni-Path Architecture
File System	7.6 PB 168 GB/s, Lustre [®]	32 PB 1 TB/s, Lustre [®]	26 PB 300 GB/s GPFS [™]	28 PB 744 GB/s Lustre [®]	120 PB 1 TB/s GPFS [™]	10PB, 210 GB/s Lustre initial	150 PB 1 TB/s Lustre [®]

Tesla Products	Tesla K40	Tesla M40	Tesla P100
GPU	GK110 (Kepler)	GM200 (Maxwell)	GP100 (Pascal)
SMs	15	24	56
TPCs	15	24	28
FP32 CUDA Cores / SM	192	128	64
FP32 CUDA Cores / GPU	2880	3072	3584
FP64 CUDA Cores / SM	64	4	32
FP64 CUDA Cores / GPU	960	96	1792
Base Clock	745 MHz	948 MHz	1328 MHz
GPU Boost Clock	810/875 MHz	1114 MHz	1480 MHz
Peak FP32 GFLOPs ¹	5040	6840	10600
Peak FP64 GFLOPs ¹	1680	210	5300
Texture Units	240	192	224
Memory Interface	384-bit GDDR5	384-bit GDDR5	4096-bit HBM2
Memory Size	Up to 12 GB	Up to 24 GB	16 GB
L2 Cache Size	1536 KB	3072 KB	4096 KB
Register File Size / SM	256 KB	256 KB	256 KB
Register File Size / GPU	3840 KB	6144 KB	14336 KB
TDP	235 Watts	250 Watts	300 Watts
Transistors	7.1 billion	8 billion	15.3 billion
GPU Die Size	551 mm ²	601 mm ²	610 mm ²
Manufacturing Process	28-nm	28-nm	16-nm FinFET
¹ The GFLOPS in this chart are based on GPU Boost Clocks.			

lattice 2018

Algorithms and Machines Review

Meifeng Lin
Brookhaven National Laboratory

Top 20 of TOP 500 (June 2018)

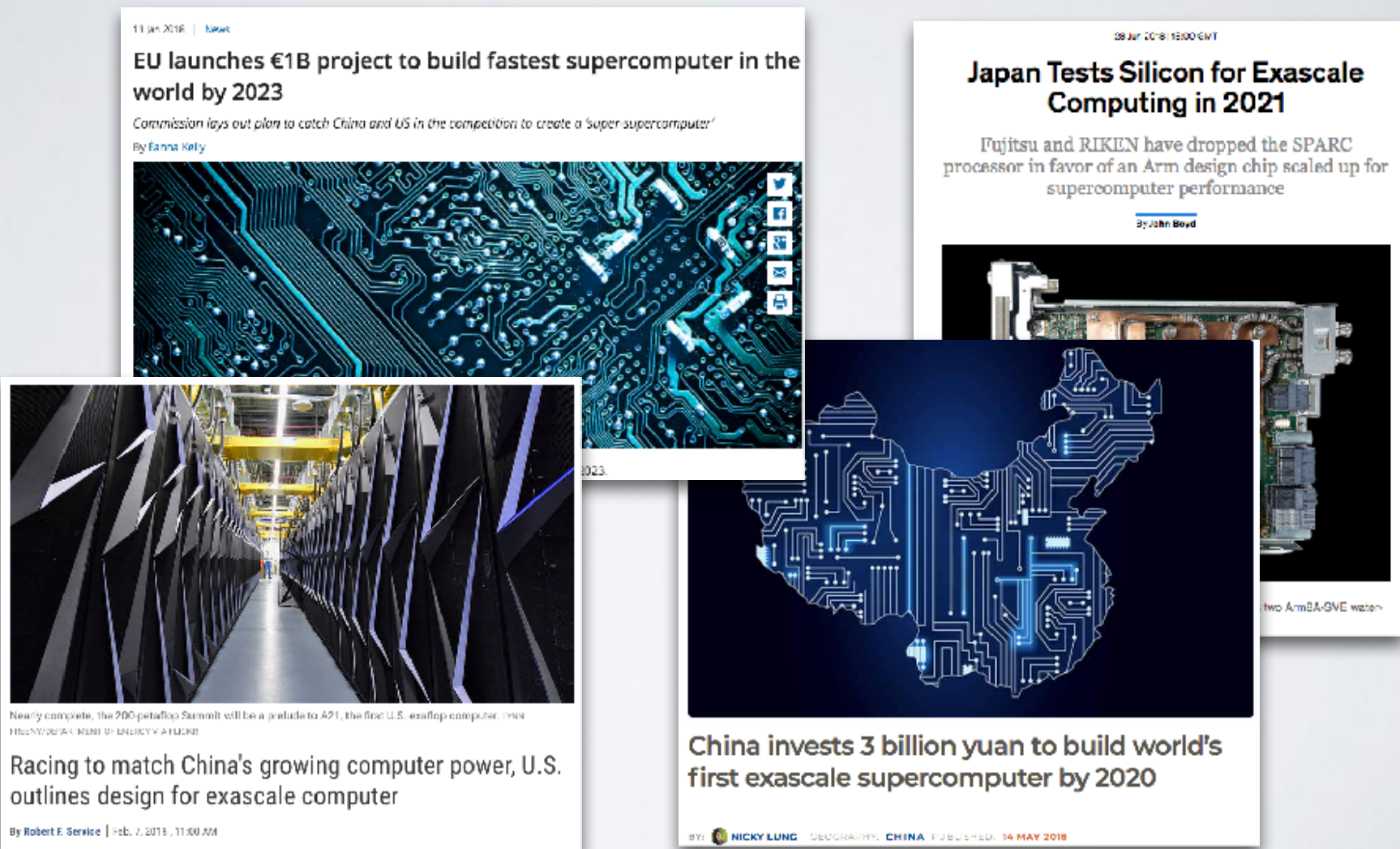
- Top 500 list has a blend of GPU-based and multi/many-core based architectures.
- In heterogeneous configurations, host CPUs are dominantly Intel Xeon CPUs.
- Several Intel Xeon Phi-based systems made it to the Top 20.

Rank	Name	Architecture	Site
1	Summit	IBM Power9 + NVIDIA GPUs	DOE/SC/Oak Ridge National Laboratory
2	Sunway TaihuLight	Sunway	National Supercomputing Center in Wuxi
3	Sierra	IBM Power9 + NVIDIA GPUs	DOE/NNSA/LLNL
4	Tianhe-2A	Intel Xeon + Matrix	National Super Computer Center in Guangzhou
5	AI Bridging Cloud Infrastructure	Intel Xeon + NVIDIA GPUs	National Institute of Advanced Industrial Science and Technology
6	Piz Daint	Intel Xeon + NVIDIA GPUs	Swiss National Supercomputing Centre (CSCS)
7	Titan	AMD + NVIDIA GPUs	DOE/SC/Oak Ridge National Laboratory
8	Sequoia	BQC	DOE/NNSA/LLNL
9	Trinity	Intel Xeon Phi	DOE/NNSA/LANL/SNL
10	Cori	Intel Xeon Phi	DOE/SC/LBNL/NERSC
11	Nurion	Intel Xeon Phi	Korea Institute of Science and Technology Information
12	Oakforest-PACS	Intel Xeon Phi	Joint Center for Advanced High Performance Computing
13	HPC4	Intel Xeon + NVIDIA GPUs	Eni S.p.A.
14	Tera-1000-2	Intel Xeon Phi	Commissariat a l'Energie Atomique (CEA)
15	Stampede2	Intel Xeon Phi	Texas Advanced Computing Center/Univ. of Texas
16	K Computer	SPARC	RIKEN Advanced Institute for Computational Science (AICS)
17	Mira	BQC	DOE/SC/Argonne National Laboratory
18	Marconi Intel Xeon Phi	Intel Xeon Phi	CINECA
19	TSUBAME3.0	Intel Xeon + NVIDIA GPUs	GSIC Center, Tokyo Institute of Technology
20		Intel Xeon	United Kingdom Meteorological Office

Data from top500.org

The race to exascale

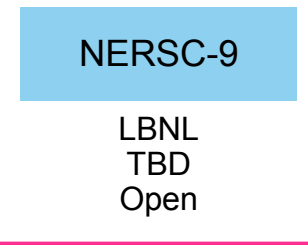
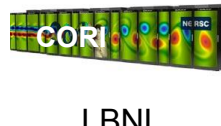
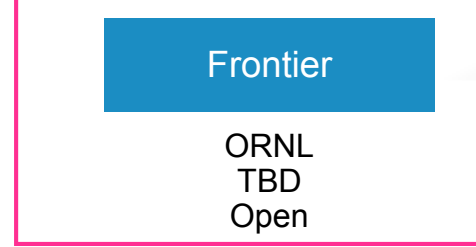
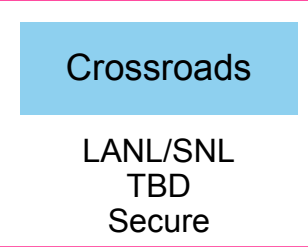
- China, EU, Japan and US are all developing exascale supercomputers.



Future Pre-Exascale and Exascale Systems in the US

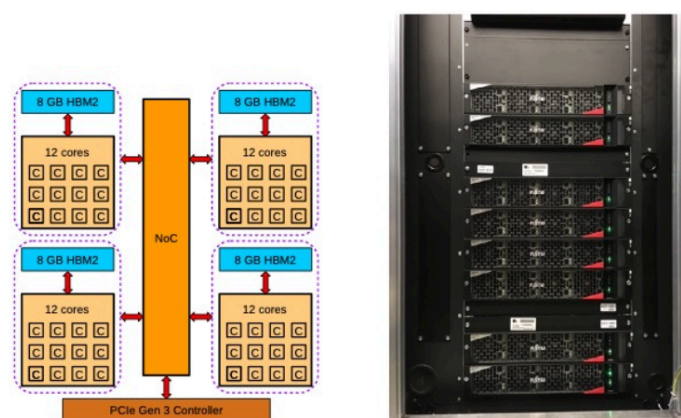
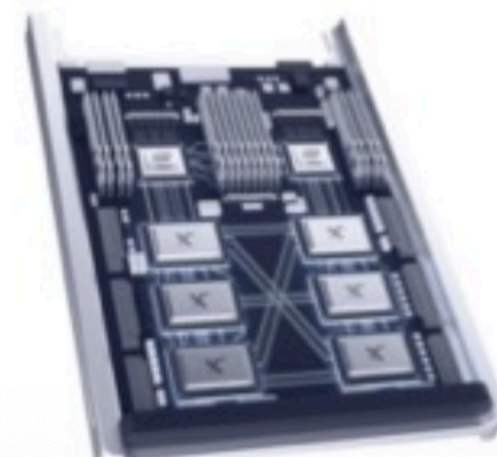
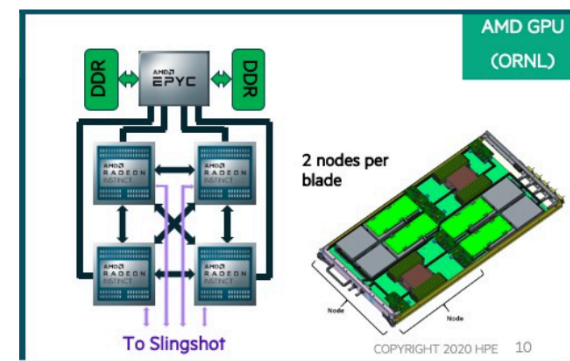
- Architectures on these future systems are still unknown

Relevant Pre-Exascale and Exascale Systems for ECP

Pre-Exascale Systems				Exascale Systems
2013	2016	2018	2020	2021-2022
 Argonne IBM BG/Q Open	 Argonne Intel/Cray KNL Open	 ORNL IBM/NVidia P9/Volta Open	 LBNL TBD Open	 Argonne Intel/Cray TBD Open
 ORNL Cray/NVidia K20 Open	 LBNL Cray/Intel Xeon/KNL Unclassified			 ORNL TBD Open
 LLNL IBM BG/Q Secure	 LANL/SNL Cray/Intel Xeon/KNL Secure	 LLNL IBM/NVidia P9/Volta Secure	 LANL/SNL TBD Secure	 LLNL TBD Secure

Doug Kothe and Stephen Lee, ECP update @ ASCAC 2017

lattice 2021

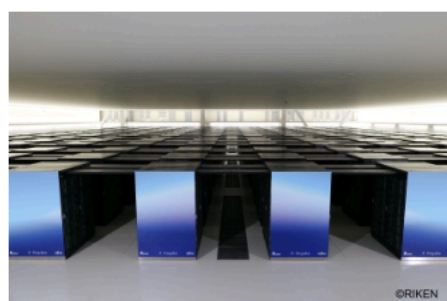


A64FX

QPACE 4

Node Programming Models: Pick your standard

- Vendor options
 - NVIDIA CUDA: the CUDA you know, evolved..
 - Support for Tensor cores in cuBLAS, cuDNN and CUDA (wmma:: namespace)
 - ecosystem support: cuBLAS, cuFFT, cuRAND etc.
 - C++ Standard Library (cuda::std::)
 - Strong push to implement **ISO C++ standard** features in full.
 - AMD HIP
 - very similar to NVIDIA's CUDA core library (replace 'cuda' with 'hip')
 - **open source implementations for ROCm, also for CPUs** (HIP CPU)'
 - LLVM/Clang based compiler
 - ecosystem support: hipBLAS, hipFFT/rocFFT, hipRAND/rocRAND, etc
 - access to CDNA features through compiler
 - Intel oneAPI
 - based on **SYCL Standard from the Khronos Group, with OpenCL heritage**
 - Compiler in oneAPI SDK and in Intel's contributions to LLVM
 - ecosystem support: oneMKL, oneDNN, ...
 - oneAPI Tutorial at this conference by P. Steinbrecher
- Directive based approaches: **OpenMP-offload and OpenACC**
 - OpenMP: `#pragma omp teams distribute parallel for simd`
 - all vendors committed to supporting OpenMP-5.0 offload
 - OpenACC supported in NVIDIA HPC-SDK and GCC (work by Mentor Graphics)
 - OpenMP is likely the best option for Fortran (although nvfortran can offload DO CONCURRENT)
- **OpenCL** is available through all vendors as well
 - but beware vendor specificity in the implementations...



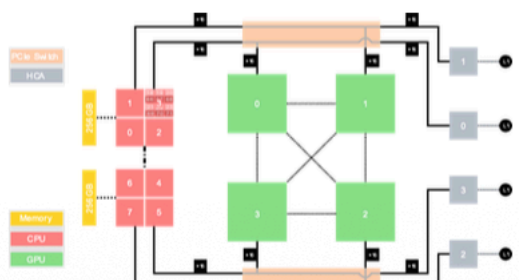
Supercomputer Fugaku



OLCF Summit



TaihuLight



Juwels Booster Node Architecture

lattice 2025

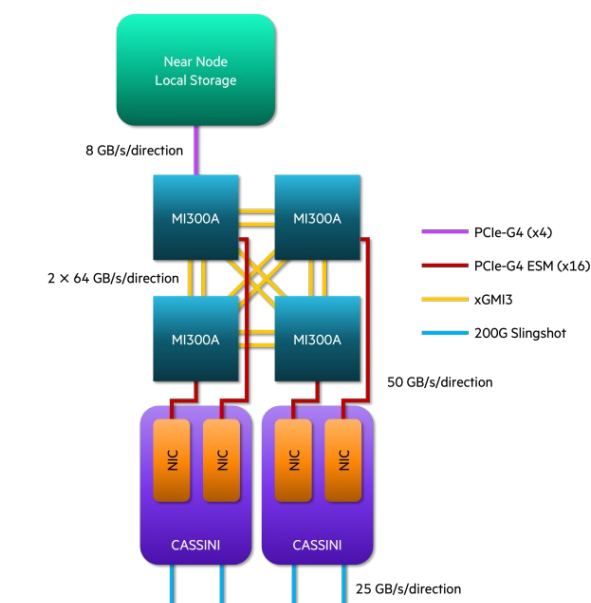
MACHINES AND EXASCALE COMPUTING IN LQCD

A review of current hardware trends in HPC

LATTICE 2025, MUMBAI – 06/11/2025 – S. KRIEG

EL CAPITAN

LLNL, USA



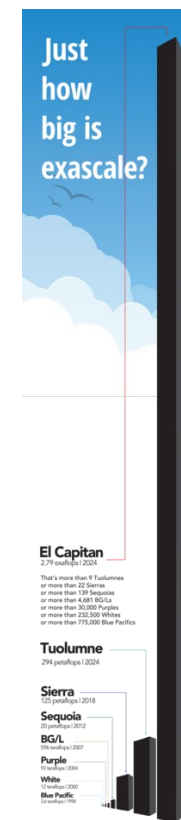
CPU: AMD EPYC "Genoa"
GPU: **AMD Instinct MI300A** (4 per node)
APU (CPU GPU bundle) 128 GB HBM3 each
Performance peak: 1.74 Eflops (29.6 MW)
Deployment: 2024
#1 in TOP500



Member of the Helmholtz Association

06/11/2025

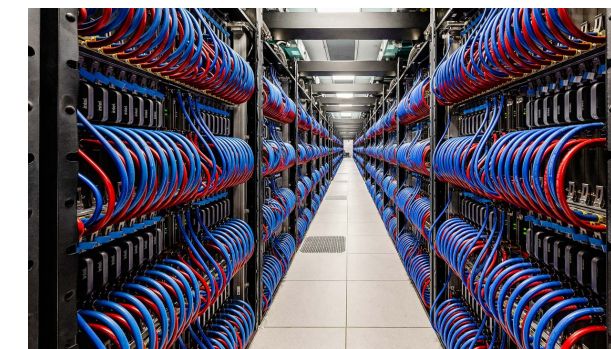
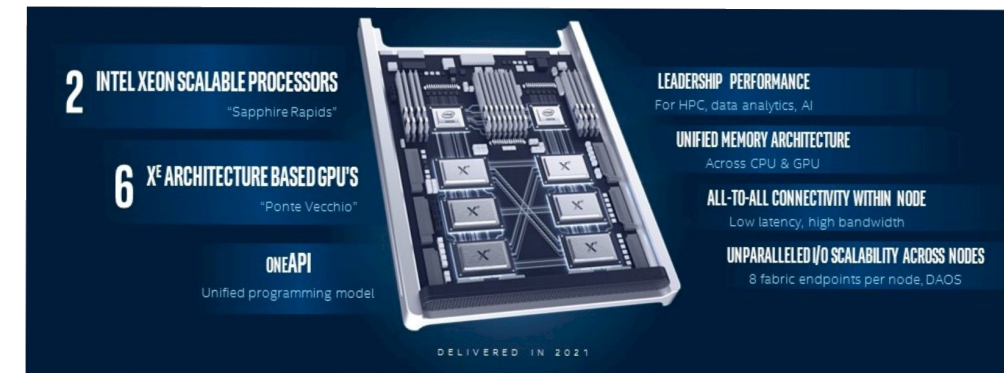
32



AURORA

ANL, USA

CPU: 2 x Intel Xeon Max, 128 GB HBM, 1024 GB DDR5
GPU: **Intel DC GPU Max** (6 per node, 128 GB HBM each)
Performance peak: 1.01 Eflops (38.7 MW)
Deployment: 2023
#3 in TOP500



CPU-CPU: UPI, CPU-GPU: PCIe-G5, GPU-GPU: Xe link, CPU-NIC: PCIe-G5 switched to G4

Member of the Helmholtz Association

06/11/2025

34



AN INCREASINGLY DIVERSE LANDSCAPE



JUPITER@JSC (Germany, EU)
CPU: ARM Grace (GH200)
GPU: **H100** (GH200)
Performance peak: 0.79 Eflops
Deployment: 2025
#4 in TOP500



LUMI@CSC (Finland, EU)
CPU: AMD EPYC
GPU: **AMD MI250**
Performance peak: 0.38 Eflops
Deployment: 2023
#9 in TOP500



Fugaku @ Riken (Japan)
CPU: ARM
GPU: **none**
Performance Peak: 0.44 Eflops
Deployment: 2021
#7 in TOP500



Leonardo@CINECA (Italy, EU)
CPU: Intel Xeon
GPU: **Nvidia A100**
Performance peak: 0.17 Eflops
Deployment: 2022
#10 in TOP500



ALPS@CSCS (Switzerland)
CPU: ARM Grace (GH200)
GPU: **H100** (GH200)
Performance peak: 0.43 Eflops
Deployment: 2024
#8 in TOP500

June 2025: 4/10 top systems use Nvidia GPUs, 4/10 AMD GPUs, 1/10 Intel GPUs, 1/10 no GPUs
November 2022: 5/10 top systems use Nvidia GPUs, 3/10 no GPUs, 2/10 AMD GPUs, 0/10 Intel GPUs
November 2021: 7/10 top systems used Nvidia GPUs, 3/10 no GPUs, 0/10 AMD or Intel GPUs

FRONTIER

ORNL, USA

CPU: AMD EPYC "Trento", 512 GB
GPU: **AMD MI250** (4 per node, 128 GB each)
Performance peak: 1.35 Eflops (24.6 MW)
Deployment: 2022
#2 in TOP500



Member of the Helmholtz Association

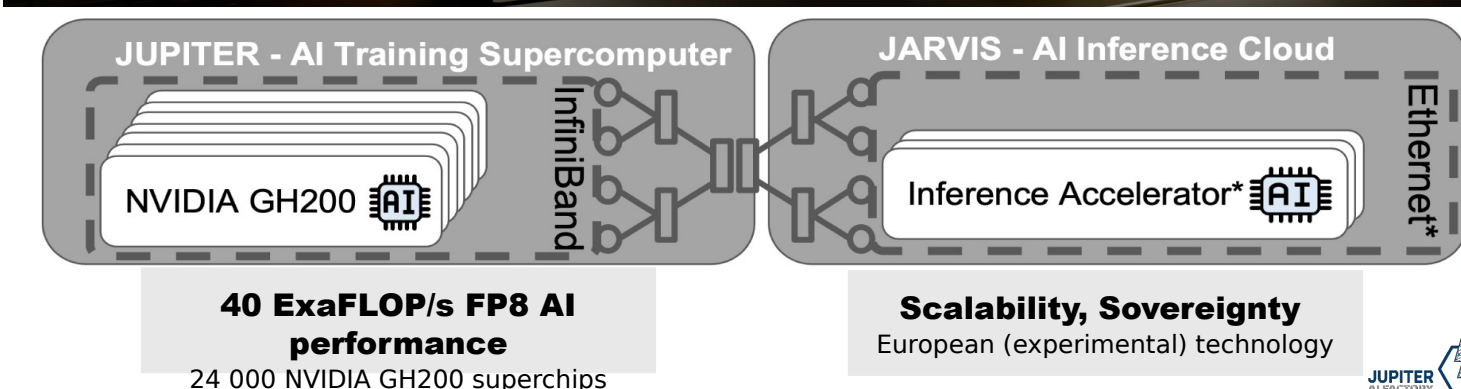
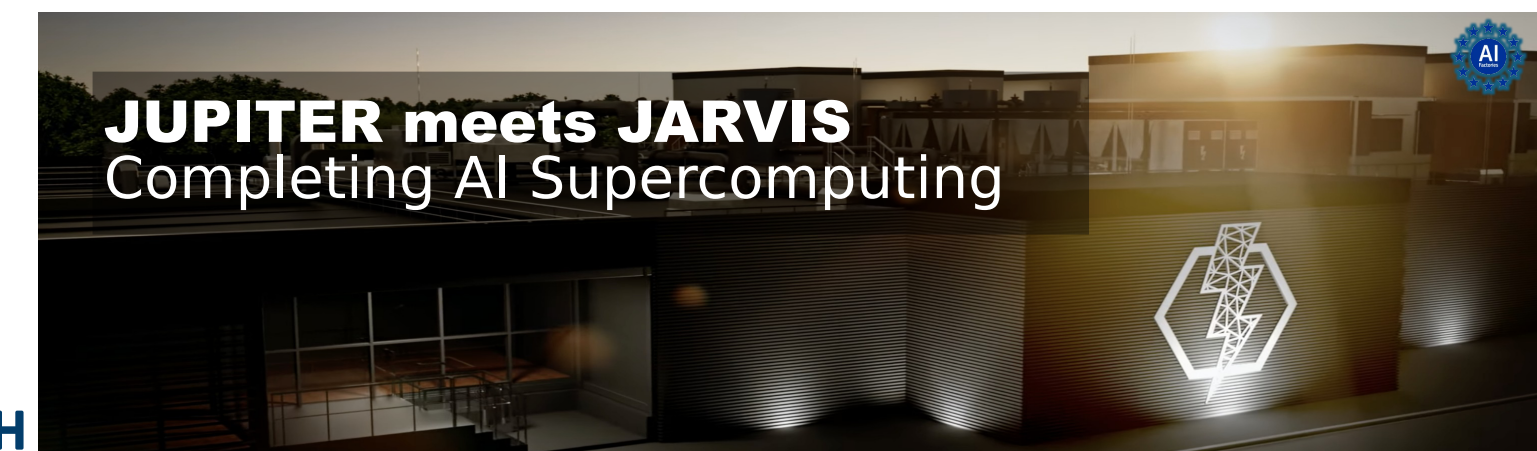
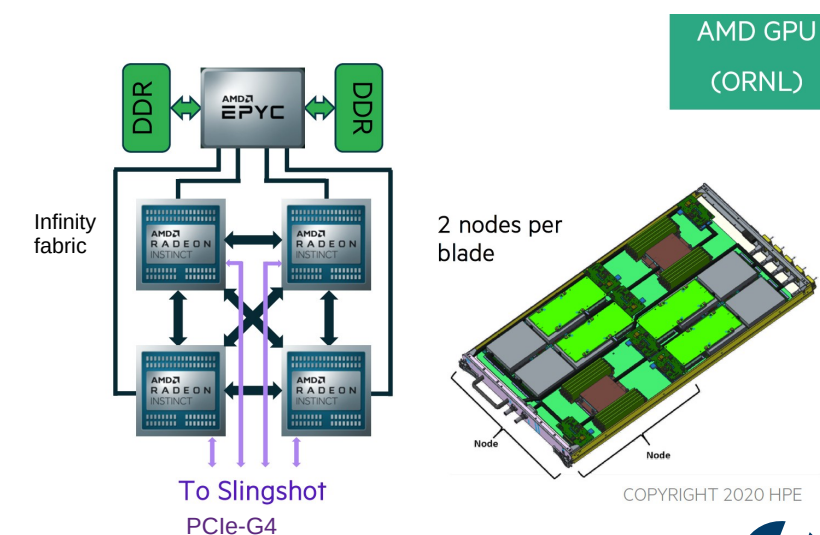
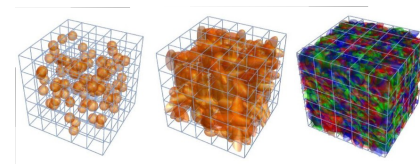
06/11/2025

33



QCD Constraints on Isospin-Dense Matter and the Nuclear Equation of State

Phys. Rev. Lett. 134, 011903 – Published 6 Janua 2025



the newcomer

- LineShine TOP500 #1

- ARM with SVE and SME

- LingQi high-speed network with a dual-plane multi-rail fat-tree

- JUPITER Booster #5

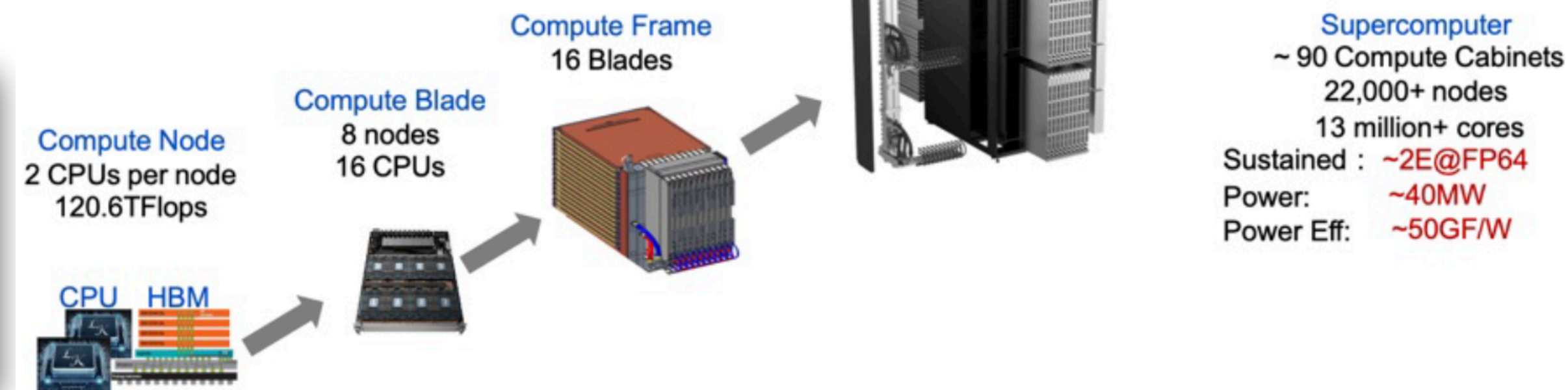
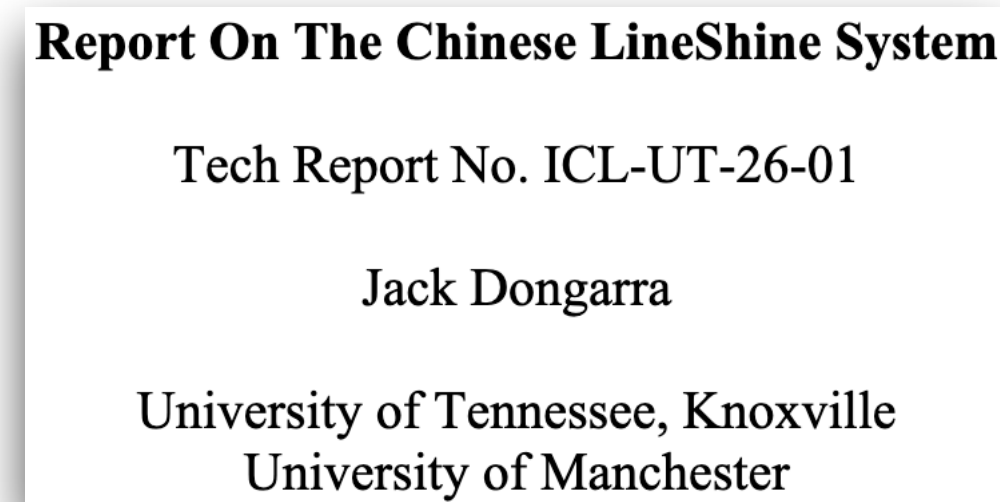
- NVIDIA GH200 Superchip

- Quad-Rail NVIDIA InfiniBand NDR200,

- HPC7 TOP500 #6

- AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A

- Slingshot-11



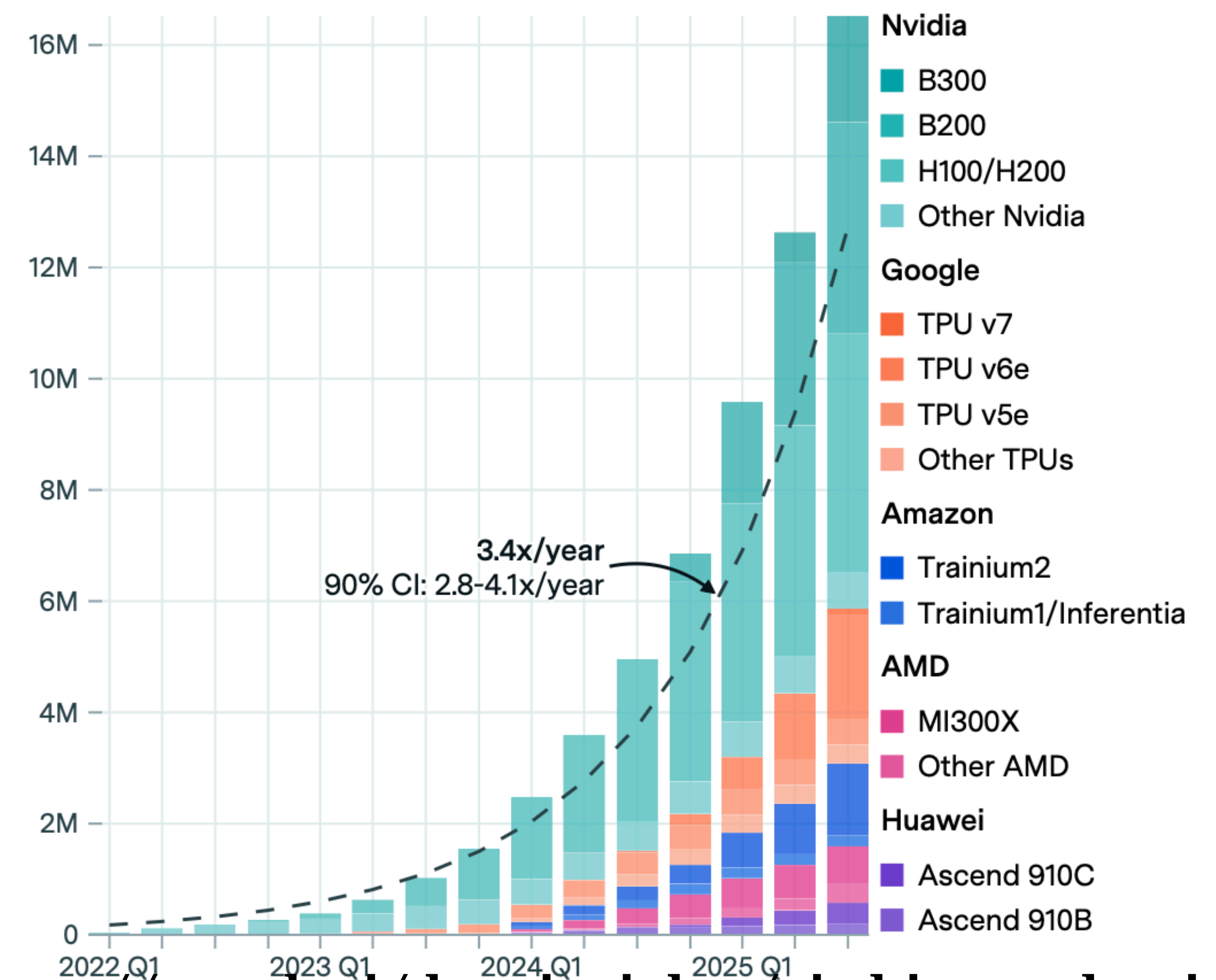
the coming-soon

- Doudna, NERSC
 - NVIDIA Vera Rubin CPU-GPU platform
 - Quantum-X800 InfiniBand
- Alice Recoque
 - AMD EPYC Venice CPUs and Instinct MI430X GPUs
 - Eviden BXI v3 fabric
- FugakuNEXT
 - Fujitsu MONAKA-X Arm CPUs and NVIDIA GPU subsystem

the missing

Total available computing capacity from AI chips across all major designers has grown by approximately 3.3x per year since 2022, enabling larger-scale model development and consumer adoption. NVIDIA AI chips currently account for over 60% of total compute, with Google and Amazon making up much of the remainder.

Cumulative compute capacity (H100e)



<https://epoch.ai/data-insights/ai-chip-production>

FP8 is All You Need (Part 1): Debunking Hardware FP64 as the HPC Holy Grail*

A Tensor–Memory Equilibrium Model and Implementation Strategy
for Ozaki Scheme II on Memory-Bound Workloads
in the Post-FP64 Era

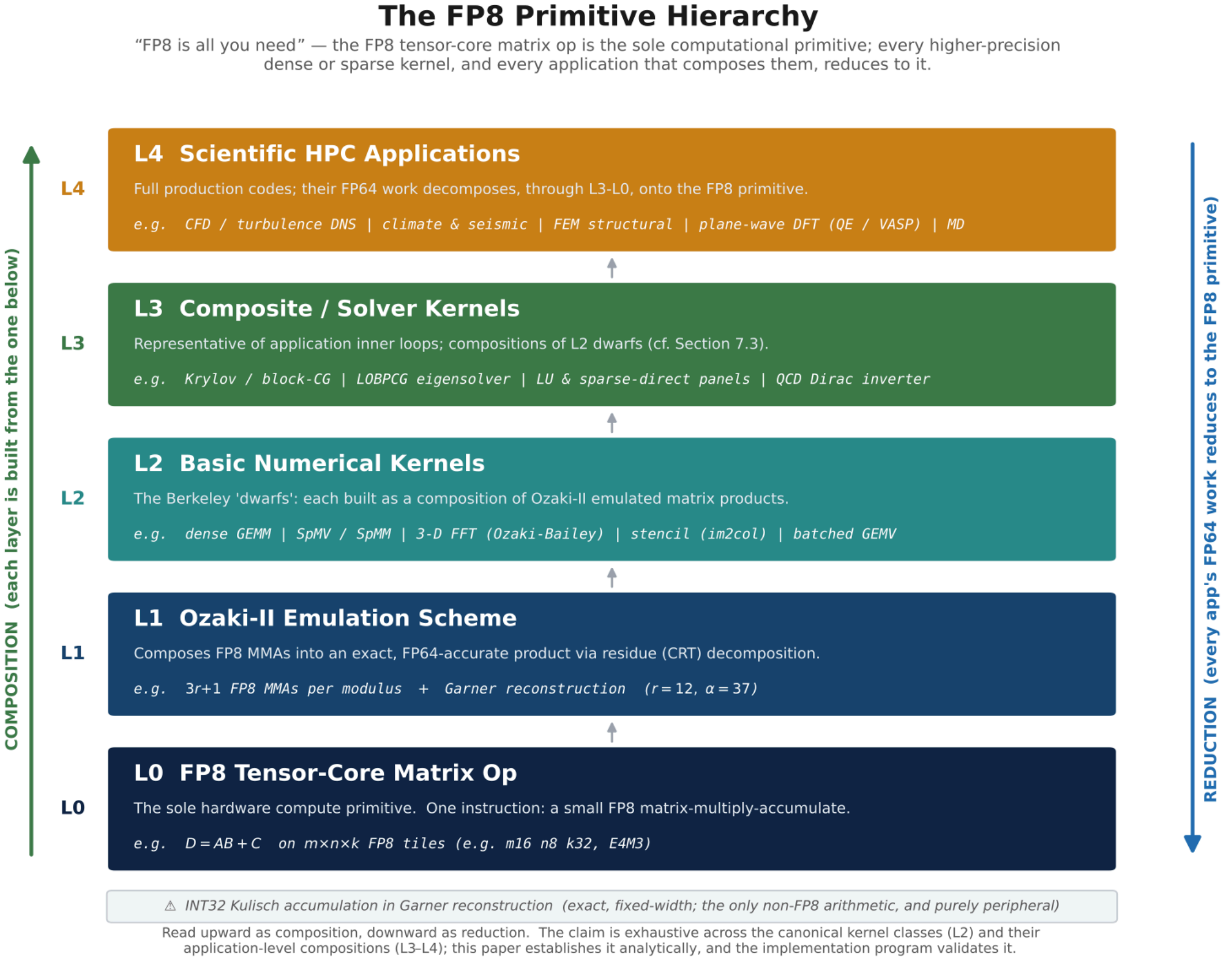
Satoshi Matsuoka[†]
Director, RIKEN Center for Computational Science (R-CCS)
Kobe, Hyogo, Japan

Version July 3, 2026

Table 6: Coverage of the Berkeley dwarf classes and their compositions. Each computational dwarf is examined through a concrete kernel whose work reduces to an L2 primitive (dense GEMM, SpMV/SpMM, FFT, or stencil) and thence, via Ozaki II or the Kulisch FFT route, to the L0 FP8 matrix op. The right column gives the recovery verdict on B300 at full FP64 accuracy; “FP32+K” denotes the FP32-with-Kahan path for BLAS-1 reductions. The enumeration is closed over the dwarf taxonomy, not a sample.

Dwarf class	Examined via	Reduces to	B300 verdict
Dense linear algebra	GEMM, LU/ <code>trsm</code> panel	dense GEMM (L2)	memory roof (§6)
Sparse linear algebra	SpMV, SpMM, sparse-direct	SpMV/SpMM (L2)	memory roof (App. G)
Spectral methods	3-D FFT	FFT (L2) [24]	memory roof (Kulisch)
Structured grids	stencil sweeps	stencil/im2col (L2)	memory roof (§6)
Unstructured grids	FEM SpMV, dual-mesh	SpMV/SpMM (L2)	memory roof / corner (§9)
Dense reductions (BLAS-1)	<code>ddot</code> , <code>dnrm2</code>	FP32+K	not binding (<5%)
Structured stencil (QCD)	Dirac inverter	stencil/GEMM (L2)	memory roof (FP32 inner)
Monte Carlo / particle	FCI-QMC, beam dynamics	— (algorithmic)	corner (§9)

(e) **Lattice QCD Dirac inversion.** The Dirac operator [14] is a structured stencil with 4×4 or 3×3 complex-matrix-valued links, $\text{OI}\approx 1$, GEMM-decomposable. Production LQCD codes already use mixed precision (FP32 inner, FP64 polish) with iterative refinement [5]; Ozaki II FP32 mode addresses the FP32 inner solve at memory-roof speed, and the FP64 polish is BLAS-2 and falls under (a).



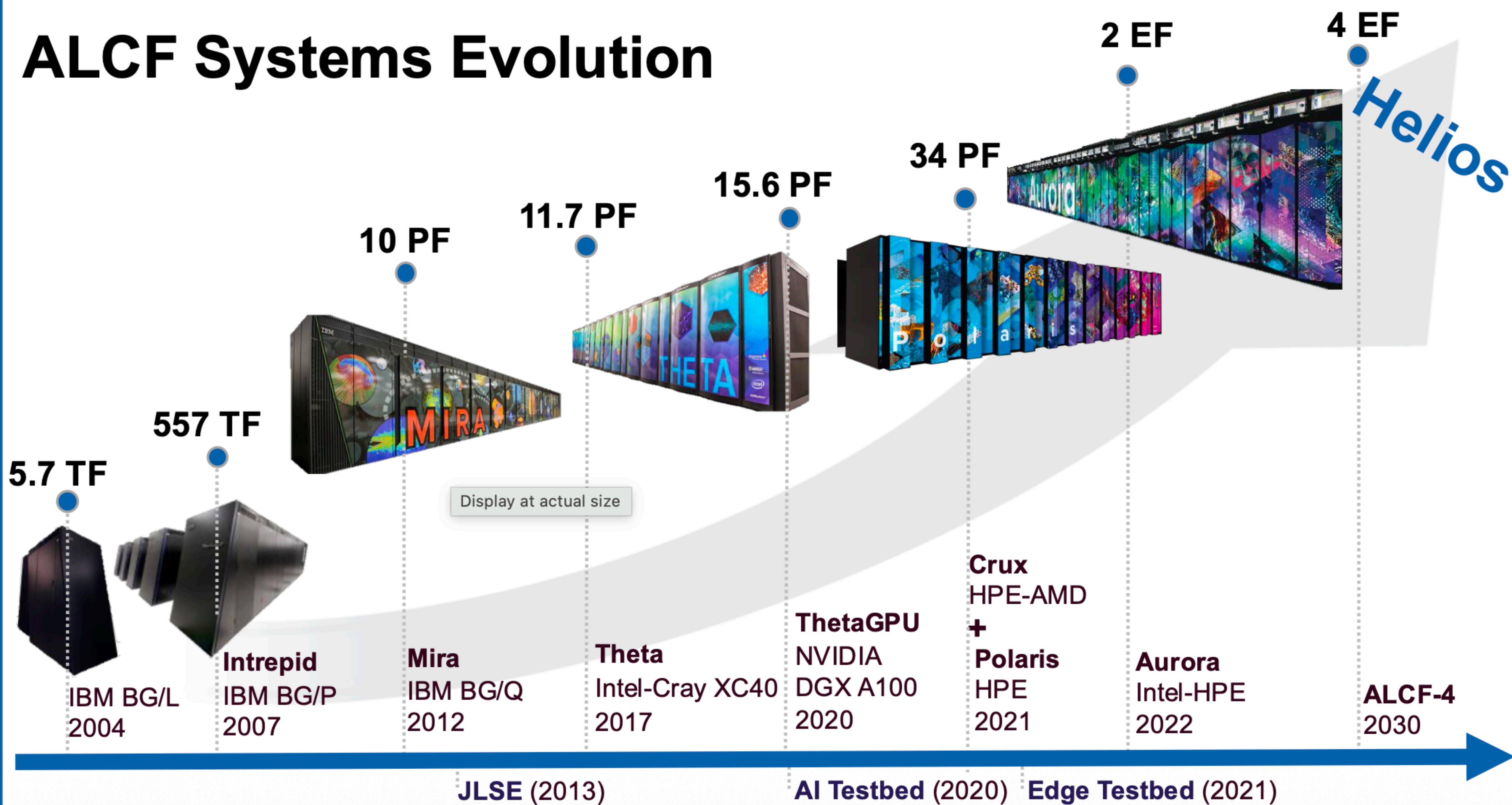
ALCF



U.S. DEPARTMENT OF
ENERGY

Argonne 
NATIONAL LABORATORY

ALCF Systems Evolution



U.S. DEPARTMENT
of ENERGY

Argonne National Laboratory is a
U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC.



Argonne Leadership
Computing Facility

ALCF AI Testbed Deployments



Cerebras (CS-3)
4x CS-3 nodes

Supports:

AI training and inference of massive AI models.

Status:

Upgraded software stack. In production for pre-training and finetuning.



Metis
SambaNova SN-40L nodes

Supports:

AI Inference for models developed by Argonne researchers.

Status:

4x nodes being integrated with ALCF Inference service. Adding 2 nodes for a total of 6x.



Tenstorrent
2x Galaxy Nodes
(64 wormhole cards)

Supports:

AI Inference (one node) and SDK development (one node).

Status:

Two nodes configured and available to Argonne users. Will be made available to ALCF users.



GroqRack
72 accelerators (9 nodes x 8 accelerators per node)

Supports:

AI inference using Language Processing Unit (LPU) technology.

Status:

Available for allocation requests.



Graphcore Bow Pod64
64 accelerators (4 nodes x 16 accelerators per node)

Supports:

ML and AI workloads.

Status:

Available for allocation requests.

Upcoming Deployments



Workforce Development Janus

NVIDIA 16x H100 GPU Cluster
(NVIDIA BlueField DPUs)



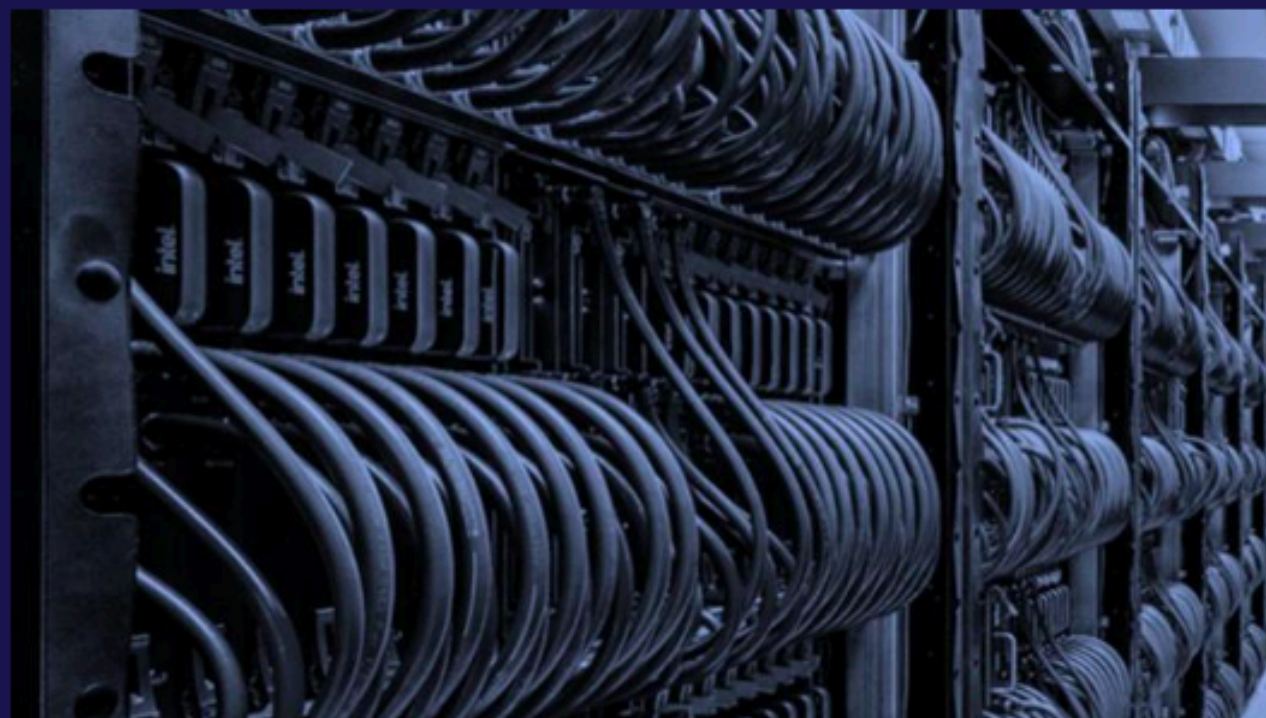
AI Inference System

Minerva
NVIDIA 64x B200



AI Inference System

Tara
NVIDIA 2688x GH200



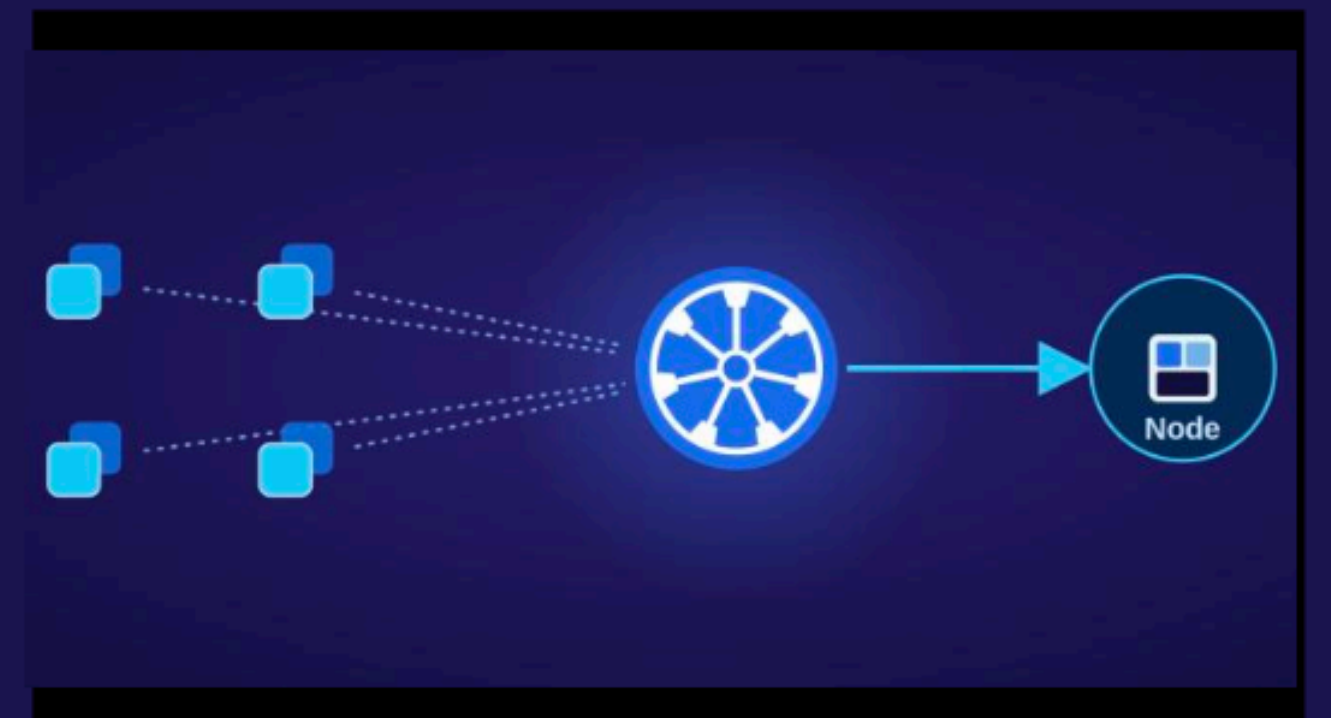
Networking

400GbE Ethernet



Storage

Global Filesystem

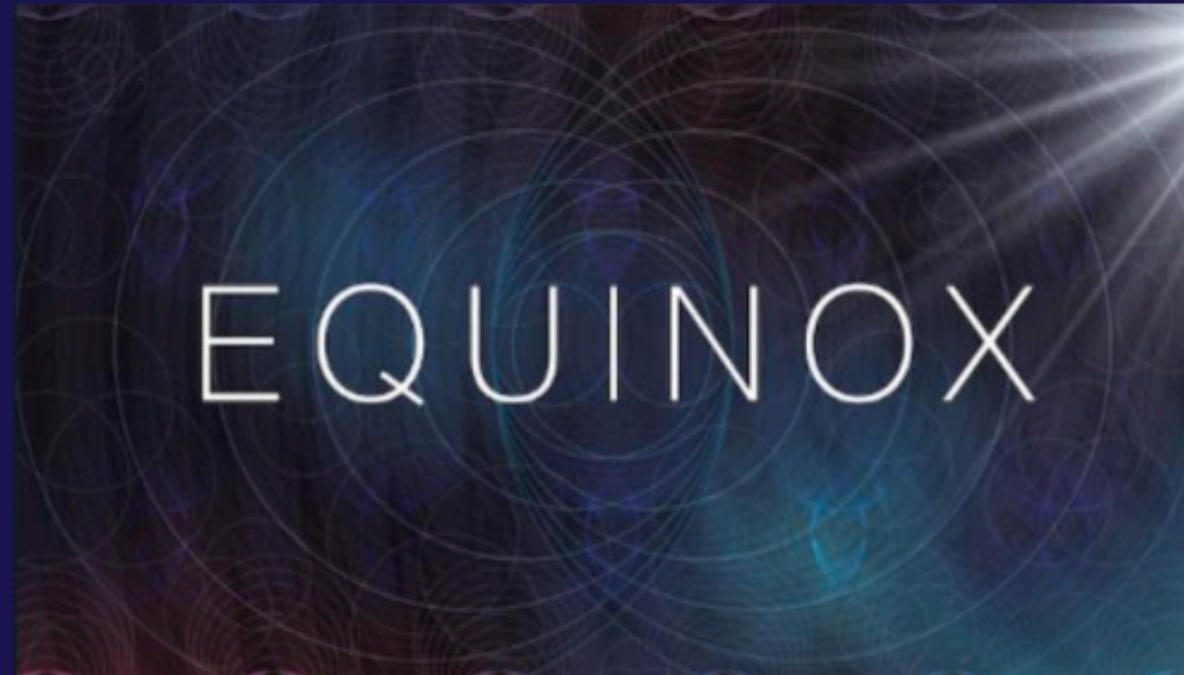


Containers Platform

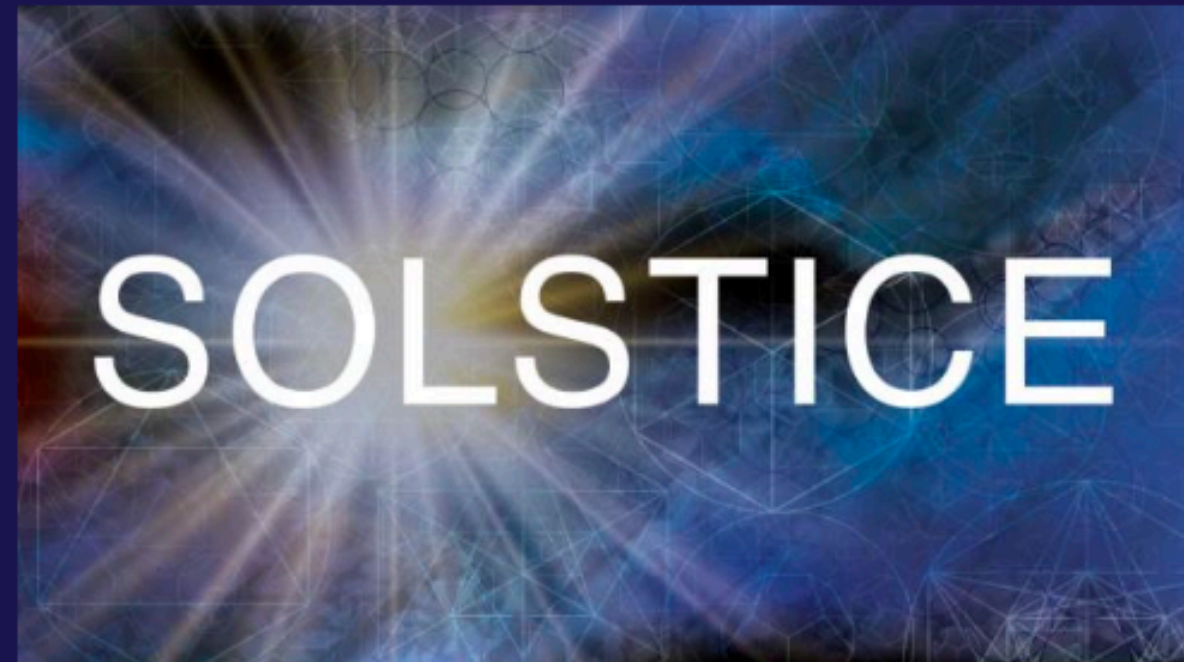
Artemis
Kubernetes Cluster

New Computing Partnerships and Key Activities

DOE · Argonne · NVIDIA · Oracle



- 2,016 NVIDIA GB300 GPUs
- Access provided in 2026



- 100,000 NVIDIA mix of GPUs
- Delivery in 2027 / 2028

Leading Efforts

- **AI Services** for the American Science Cloud (AmSC)
- **Model Services** AI model catalog & infrastructure
- **At-Scale Services** across ASCR facilities
- **AmSC Interfaces** programmable access to AmSC
- **Intelligent Operations** facility data for AI-driven analysis
- **IRI Integration** for programmable HPC workflows
- **User Engagement** training, docs, demos, support

Partnering Efforts

- **AmSC Technical Leadership** helping define priorities
- **GlobusAuth Integration** with the Virtual Organization team
- **Globus Transfer** ↔ **AmSC catalog** with Data Services
- **ModCon Liaison** align requirements in AmSC development

softwares & algorithms (non-ML)

accelerator use

- Frank Winter: Portability without Code Changes: Scaling QDP-JIT/Chroma via LLVM across NVIDIA, AMD, and Intel GPUs, 7/27 2:40 PM
- Justus Kuhlmann: Accelerating the HMC in openQxD using QUDA, 7/27 4:30 PM

Chroma + QDP-JIT on NVIDIA, AMD and Intel GPUs

Chroma + QDP-JIT

- Chroma algorithms remain unchanged
- QDP-JIT provides the portable execution layer
- Full GPU support now available for:
 - NVIDIA GPUs
 - AMD GPUs
 - Intel GPUs

Production HMC

- Wilson Clover RHMC runs stably on DOE exascale systems
- Demonstrated on Frontier and Aurora

Why JIT?

- Avoids dependence on vendor C++ offload compilers
- No reliance on SYCL, OpenMP offload, or oneAPI front-ends
- No “bloated” executable

Next Step

- Adaptive algebraic multi-grid features are being added
- Support includes variable lattice sizes, mixed-size expressions, and blocked reductions

software and library

- Xiangyu Jiang: Use QUDA with Python, 7/27 4:10 PM
- Nikolai Husung: opbasis: automated construction of minimal operator bases for lattice QCD, 7/27 5:10 PM
- Max Hansen: AmPyL: A python toolkit for three-particle finite-volume spectra and scattering amplitudes, 7/28 4:50 PM
- Gunnar Bali and Giovanni Pedriva: First experiences operating and using ILDG 2.0, 7/27 3:20 PM
- Constantinos Costa: Efficient Storage and Compression of Intermediate Data in Lattice Calculations, 7/27 4:50 PM
- Alexis Harilaos Verney-Provatas: Easier parameter scans with hmcdj, 7/28 4:10 PM

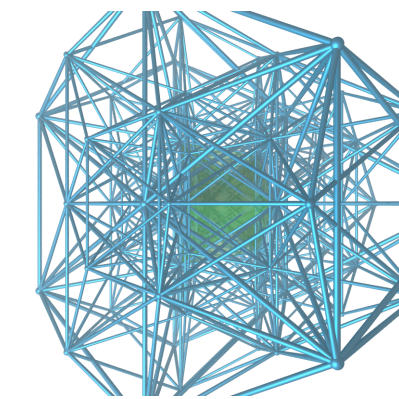
autodifferentiation

you don't need to use ML to use it

- GPT/GRID (Christoph, Peter, and others) supports it at the Python level
- JuliaQCD supports it at the LLVM level (local node for now)
 - Lattice Gauge Theory via LLVM-Level Automatic Differentiation
Yuki Nagai, Akio Tomiya, Hiroshi Ohno
<https://arxiv.org/abs/2602.20516>
 - See 3 talks using JuliaQCD by Shunsuke Yasunaga, Akio Tomiya, Okuto Morikawa
- QEX (James Osborn) supports it
 - James' previous lattice proceedings
 - X Jin's talk, 7/28 3 PM
 - <https://github.com/jcosborn/qex/tree/devel/src/experimental/graph/examples>

alternative formulations

- Shailesh Chandrasekharan: Projected Density Matrix Sampling for Quantum Lattice Hamiltonians, 7/28 4:10 PM
- Rohan Misra: The Affine Conjecture: Using Numerical Methods to Tune Geometries for Non-integrable Lattice Field Theories, 7/30 5:10 PM
- Daniel Negradi: Lattice QCD on the 16-cell honeycomb, 7/31 3:20 PM
- Katsumasa Nakayama: Tensor renormalization group for four-dimensional $SU(N)$ gauge theories, 7/31 4:10 PM
- Takaaki Kuwahara: Tensor renormalization group analysis of correlators in Yang-Mills theory, 7/31 4:30 PM
- Benjamin Izett: Simulating the string formulation of $SU(3)$ lattice gauge theory, 7/31 4:50 PM



Tensor renormalization group study of cold and dense QCD in the strong coupling limit

Yuto Sugimoto^{1,*}, Shinichiro Akiyama^{2,3,†} and Yoshinobu Kuramashi^{2,‡}

¹Department of Physics, Tohoku University, Sendai 980-8578, Japan

²Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

³Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan



(Received 29 January 2026; accepted 8 May 2026; published 11 June 2026)

We study the phase structure of the $(3 + 1)$ -dimensional cold and dense QCD with the Kogut–Susskind quark in the strong coupling limit using the tensor renormalization group method. The chiral and nuclear transitions are investigated by calculating the chiral condensate and the quark number density as a function of the chemical potential. For a fixed temporal extent $N_\tau = 8$, we determine the critical quark masses m_c^χ and m_c^n for the chiral condensate and the quark number density, respectively, at which the first-order phase transition terminates with the vanishing discontinuity in thermodynamic quantities. We find that both quantities at the same quark mass exhibit a discontinuity at the same chemical potential, and the resulting critical quark masses are consistent with each other. We also compare our results for the critical quark masses with those obtained from the Monte Carlo simulation in the dual formulation and from the mean-field analysis. We further confirm the first-order phase transition at finite quark mass on a 1024^4 lattice, which is essentially in the thermodynamic limit at zero temperature, as expected from the mean-field analysis.

DOI: [10.1103/v5v2-7mq5](https://doi.org/10.1103/v5v2-7mq5)

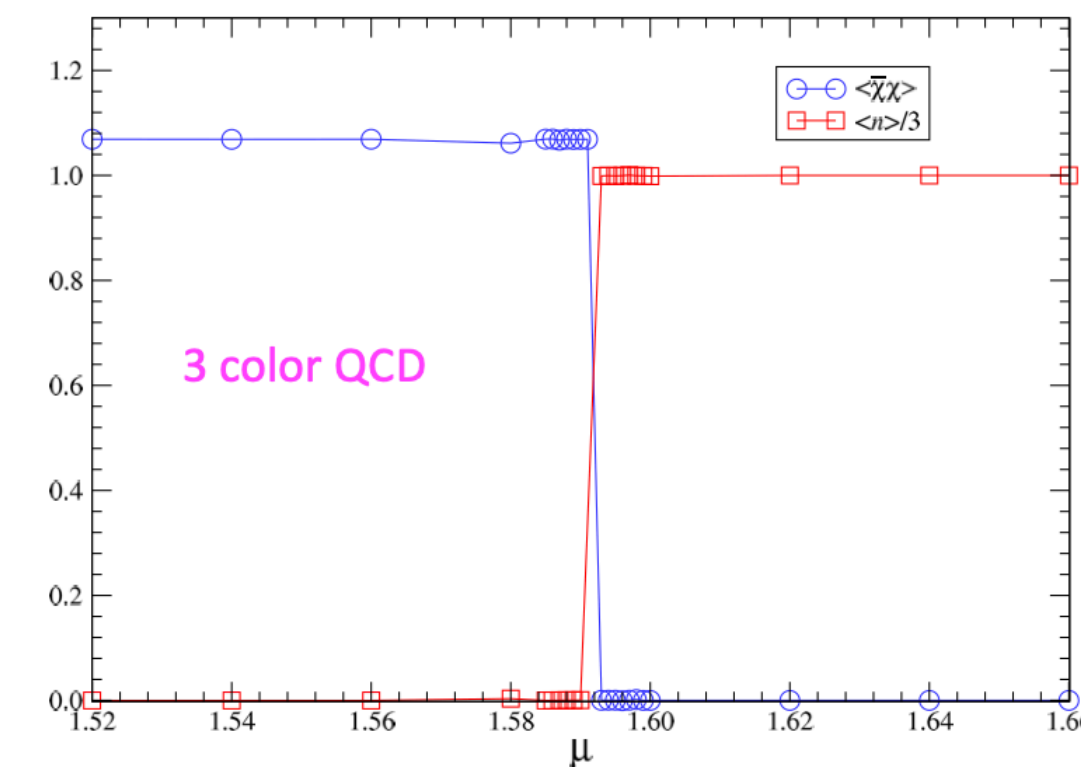


Phase Transition at T=0

Sugimoto-Akiyama-YK, PRD113(2026)114509

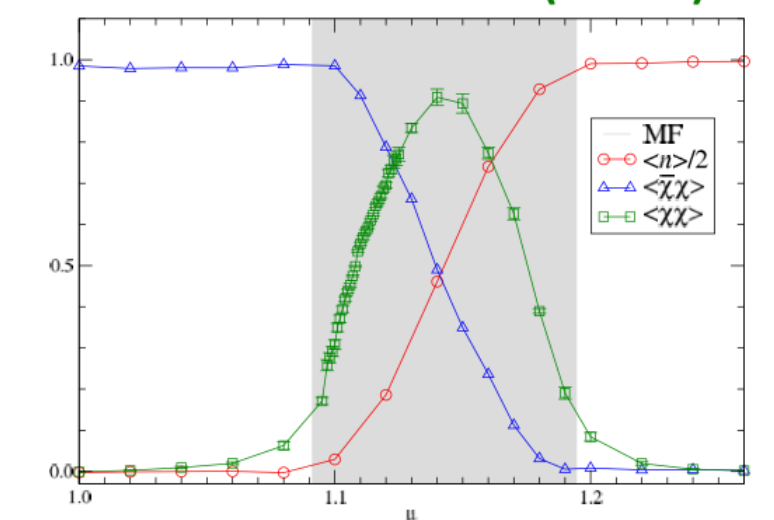
What happens w/ $m=2.07$ ($\approx m_c(N_\tau = 8)$) at T=0?

$V=1024^4 \Rightarrow$ thermodynamic limit at zero temperature



$\langle \bar{\chi}\chi \rangle$: chiral condensate
 $\langle n \rangle$: quark number density
 $\langle \chi\chi \rangle$: diquark condensate (2 color QCD)

cf. 2 color QCD PRD113(2026)034503



Nuclear transition and chiral transition are the first-order at the same chemical potential

First successful study of phase structure of finite density QCD at T=0

$\Rightarrow$ Next step is finite g case

critical slowing down and topological charge

- Peter Boyle: algorithms for domain wall fermions, 7/27 3:20 PM
- Gianluca Fuwa: Parallel tempered Metadynamics in full QCD, 7/27 4:50 PM
- Victor Granados-Pinto: Exploration of Parallel Tempering on Boundary Conditions, 7/27 5:10 PM
- Xiao-Yong Jin: Crafting the change of variables for HMC, 7/28 3 PM
- Shuhei Yamamoto: Comparison of Smearing Kernels for Field-Transformation HMC via the Master-Field Technique, 7/31 5:10 PM

AI/ML-inspired RMHMC kernels—CPN model study

Nobu Matsumoto's poster

w/ R. Abbott, P. Boyle, L. Jin,
C. Kelly, C. Lehner; in prep.

Riemannian manifold HMC (RMHMC)

~Fourier acceleration

$$H_{\text{HMC}} \equiv \frac{1}{2} \pi^T K(U) \pi + S_{\text{QCD}}(U) - \log \det K(U)$$

- Simple game:
Design $K(U)$ so that the HMC accelerates
- Combined w/ multilevel integration,
the overhead becomes additive
to the fermion cost
- Complementary to
the field transform HMC (FTHMC)

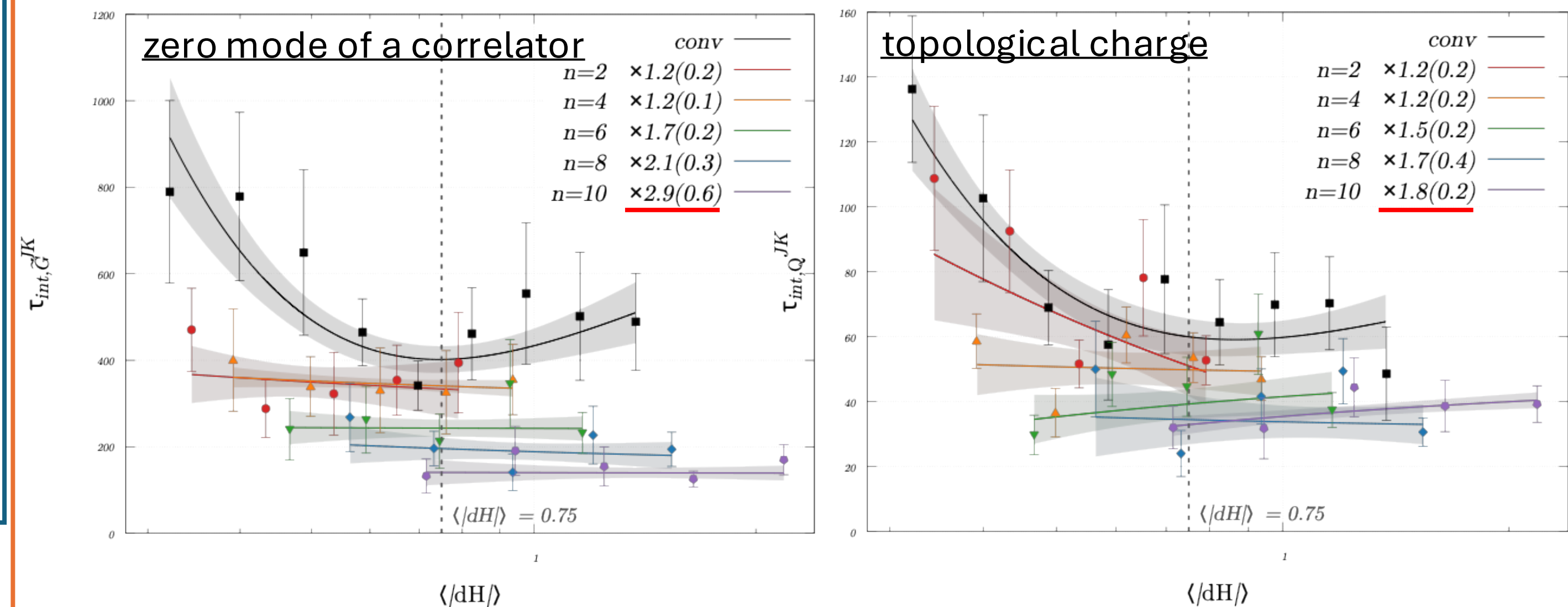
This work

Import the coupling layer from AI/ML:

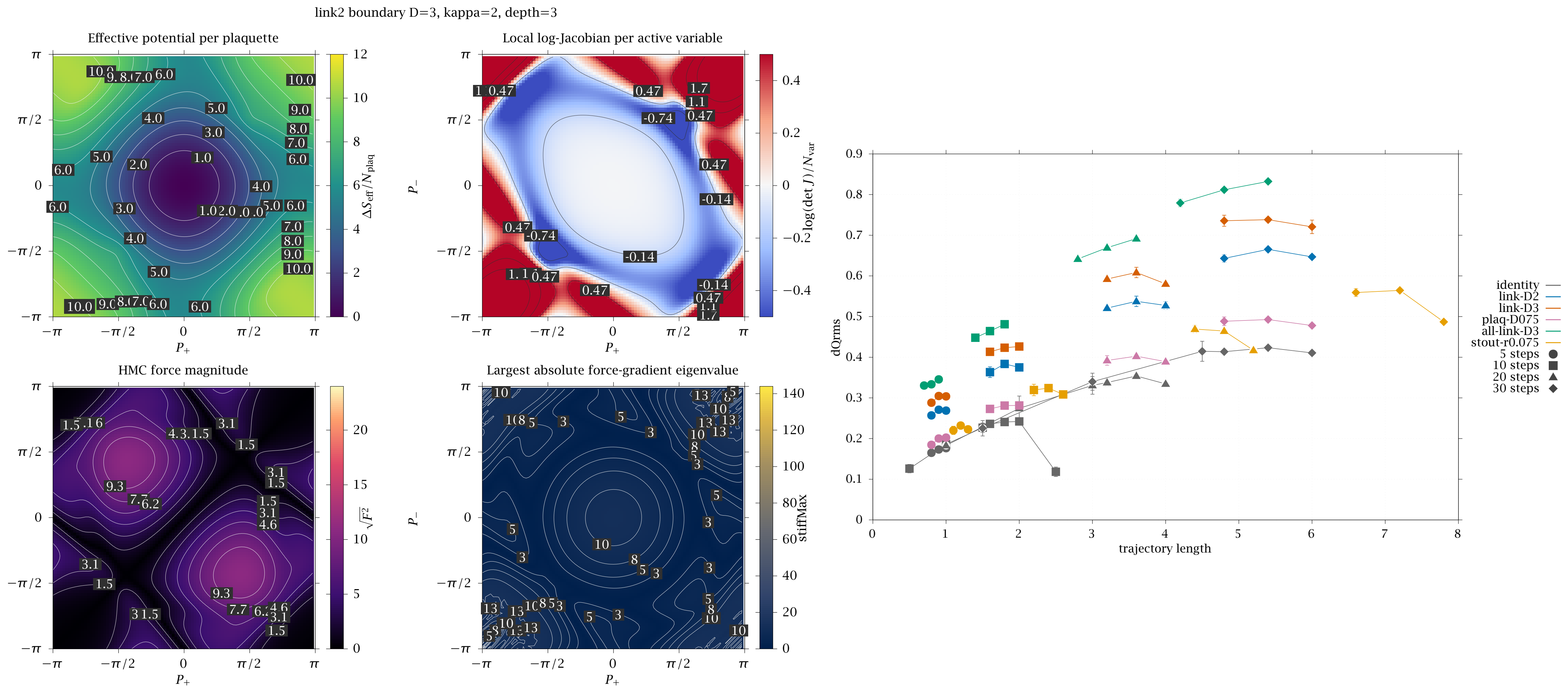
- The inverse of K is easy to take; no internal CG
→ overhead massively reduced
- No force from the determinant

Even the simple smearing kernel works:

speed-up over conventional HMC at fixed accuracy $\langle |dH| \rangle = 0.75$ speed-up over conventional HMC at fixed accuracy $\langle |dH| \rangle = 0.75$



Xiao-Yong Jin: Crafting the change of variables for HMC, 7/28 3 PM



Parallel Tempered Metadynamics for full QCD

Timo Eichhorn ^{*}, Gianluca Fuwa [†], Christian Hoelbling [‡] and Lukas Varnhorst [§]

Department of Physics, University of Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany

(Dated: July 24, 2026)

the swap probabilities are given by

$$A_{1\leftrightarrow 2}(U_1, U_2) = \min \left\{ 1, \frac{\exp[V(Q_{\text{CV}}(U_2))]}{\exp[V(Q_{\text{CV}}(U_1))]} \right\}.$$

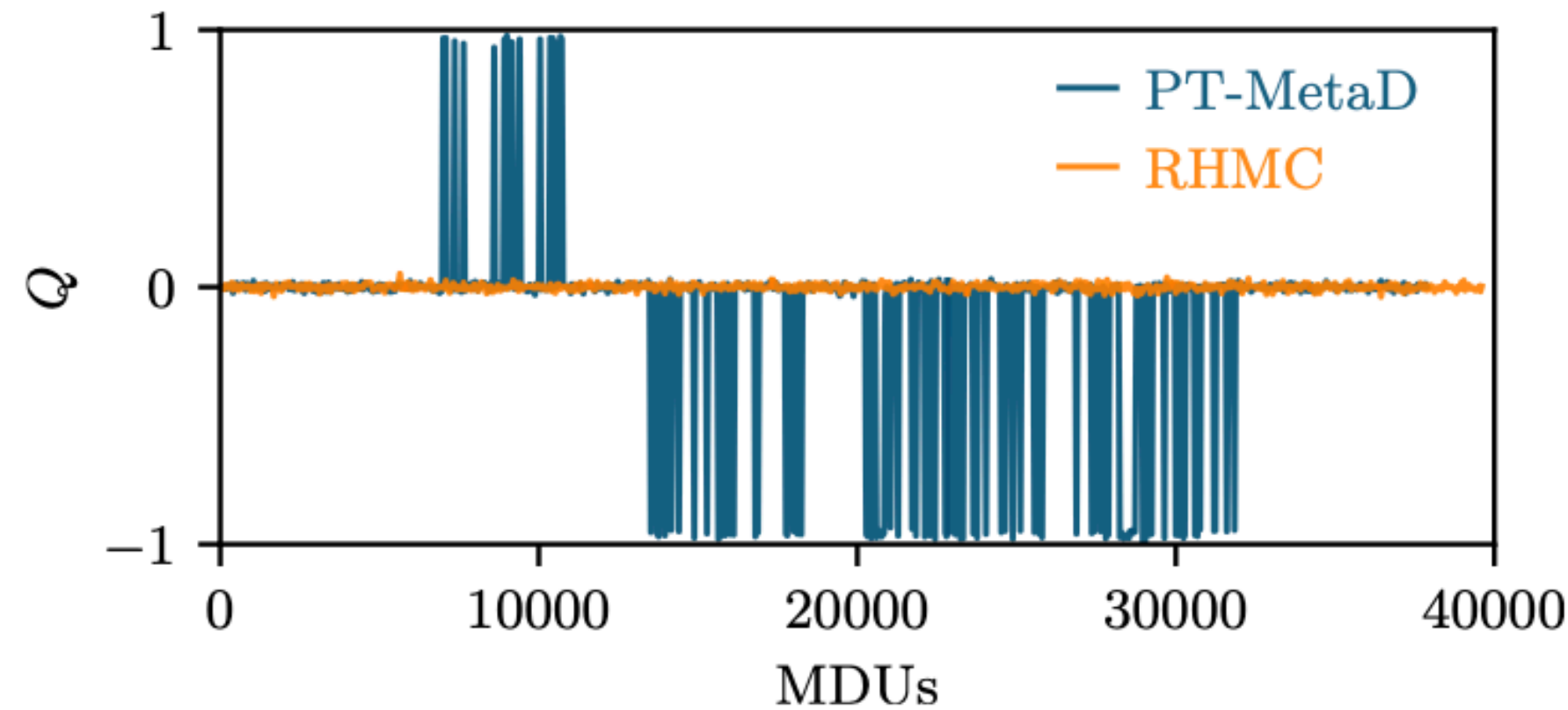


FIG. 2. Time series of the topological charge at $\beta = 1.15$ using the RHMC and the PT-MetaD algorithm. The RHMC algorithm remains within the zero sector, whereas PT-MetaD repeatedly samples multiple sectors.

Gianluca Fuwa 7/27 4:50 PM

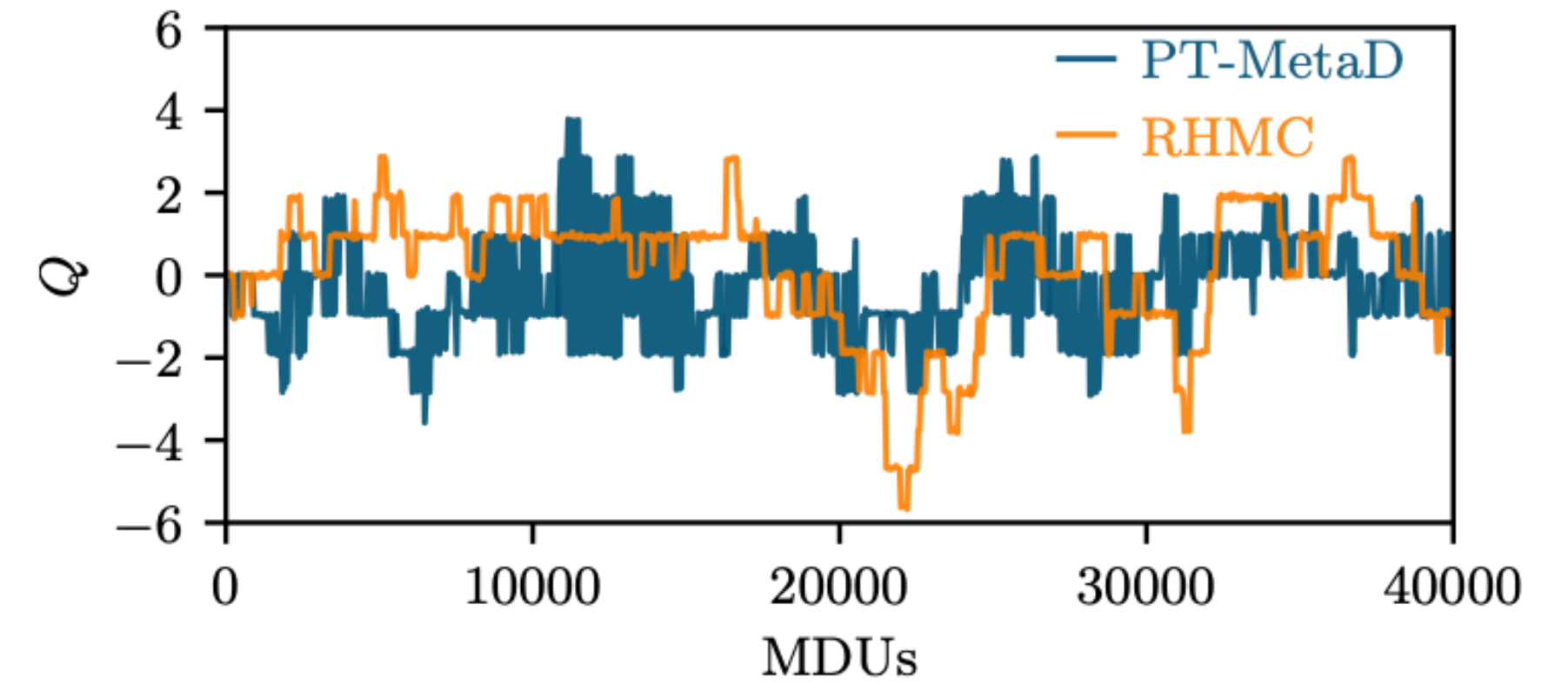
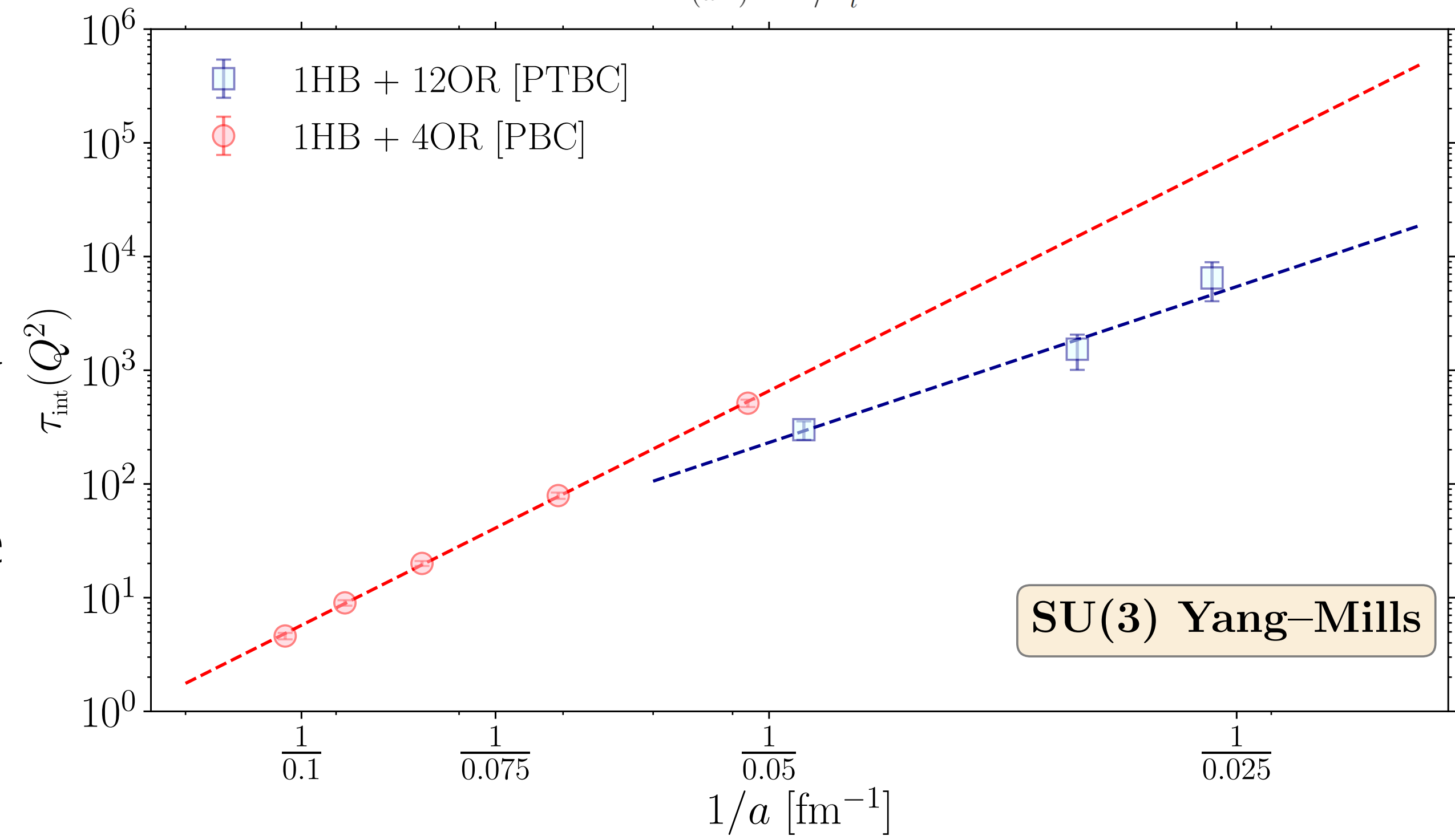
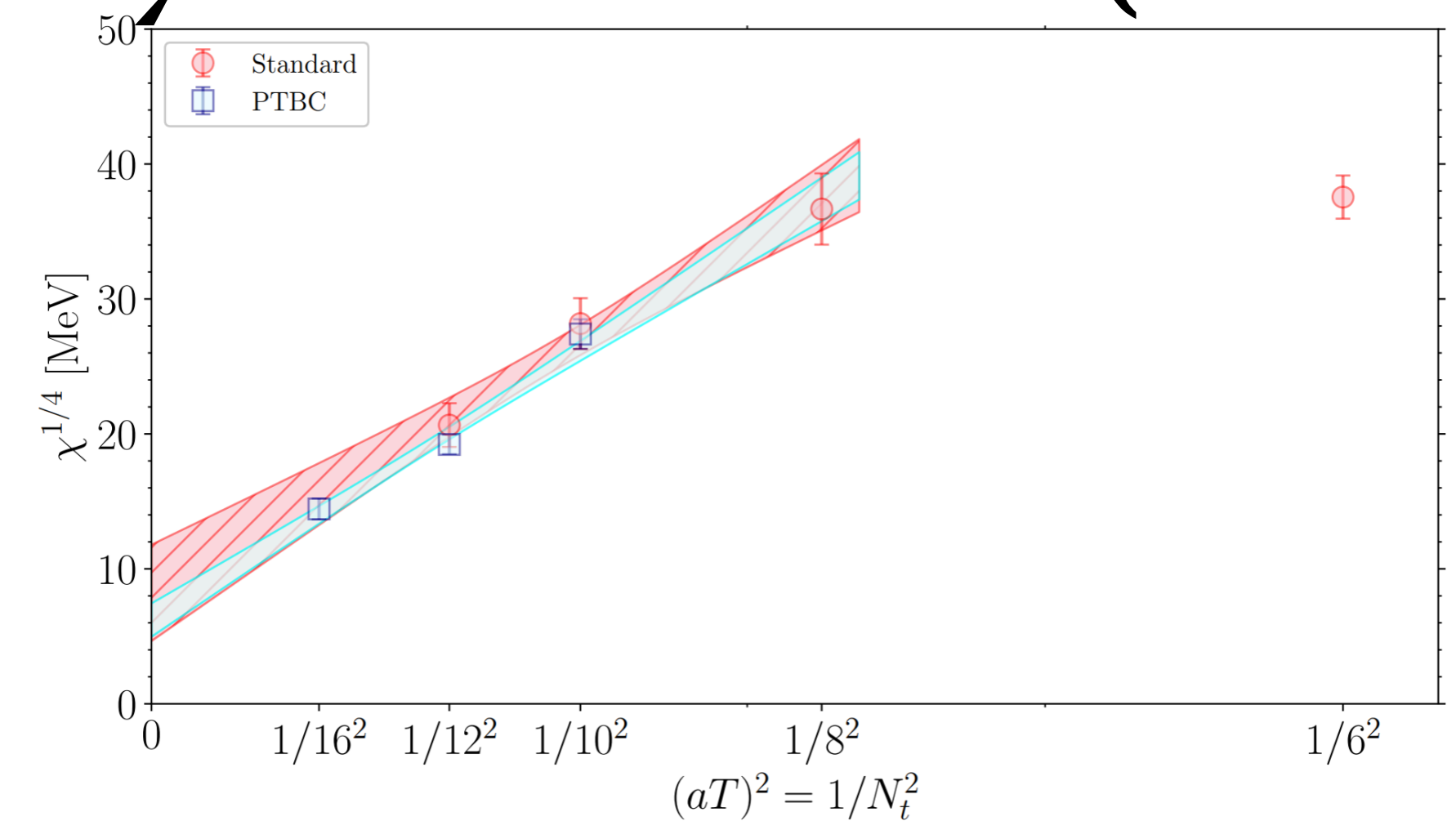


FIG. 3. Time series of the topological charge at $\beta = 0.95$ using the RHMC and the PT-MetaD algorithm. Both algorithms sample multiple sectors, but PT-MetaD exhibits more frequent transitions and a shorter integrated autocorrelation time of Q^2 (see [Table III](#)).

Parallel Tempering on Boundary Conditions (PTBC)

to mitigate topological freezing

- 2d CP(N-1):
<https://arxiv.org/abs/1706.04443>
- 4d gauge theories:
<https://arxiv.org/abs/2012.14000>
- full QCD: <https://arxiv.org/abs/2404.14151>
- strong coupling without topological freezing <https://arxiv.org/abs/2403.13607>
- large-N limit of the topological susceptibility <https://arxiv.org/abs/2510.08006>
- large-N limit of the topo susceptibility slope <https://arxiv.org/abs/2606.29885>



Inverse blocking with conditional normalizing flows

Anna Hasenfratz, Ethan Neil, Letizia Parato, Noah Schwartz

Department of Physics, University of Colorado Boulder

Letizia Parato, 7/30/26, 3:20 PM
Anna Hasenfratz, poster

Approximate perfect blocking

Perfect Blocking:

Require that the critical action blocks into itself, $\mathbf{S}_b(\kappa_{cr}) = \mathbf{B}[\mathbf{S}_f(\kappa_{cr})]$:

- Choose a translationally invariant finite-range kernel K
- Form $\psi = K\phi$ and retain the even-even sites ψ_{ee}
- Compare ψ_{ee} with directly generated coarse fields ϕ_c
- If $S_c = B[S_f]$, all operator expectation values of ψ_{ee} and ϕ_c must agree
- Optimize K using a covariance-weighted mismatch of local observables

A five-parameter 5×5 kernel already gives close coarse-distribution matching.

Conditional upscaling flow

The fine proposal is parameterized by a coarse field and detail variables:

$$\tilde{\phi}_f = K^{-1}F_{\theta}(\phi_c, z), \quad z : \text{Gaussian random}$$

Optimize the initial inverse map $F_{\theta}(\phi_c, z)$ in stages:

- Embed ϕ_c , generated with $S(\kappa_{cr})$, on the even-even sites
- Generate edge sites $(1, 0)$ and $(0, 1)$ first, then body-centered sites $(1, 1, 1)$
- Apply the invertible kernel K^{-1}
- Train by matching local observables of native and proposed fine ensembles
- Add tail-sensitive loss terms to reduce rare large deviations

The action-density distribution rapidly converges to the native target.

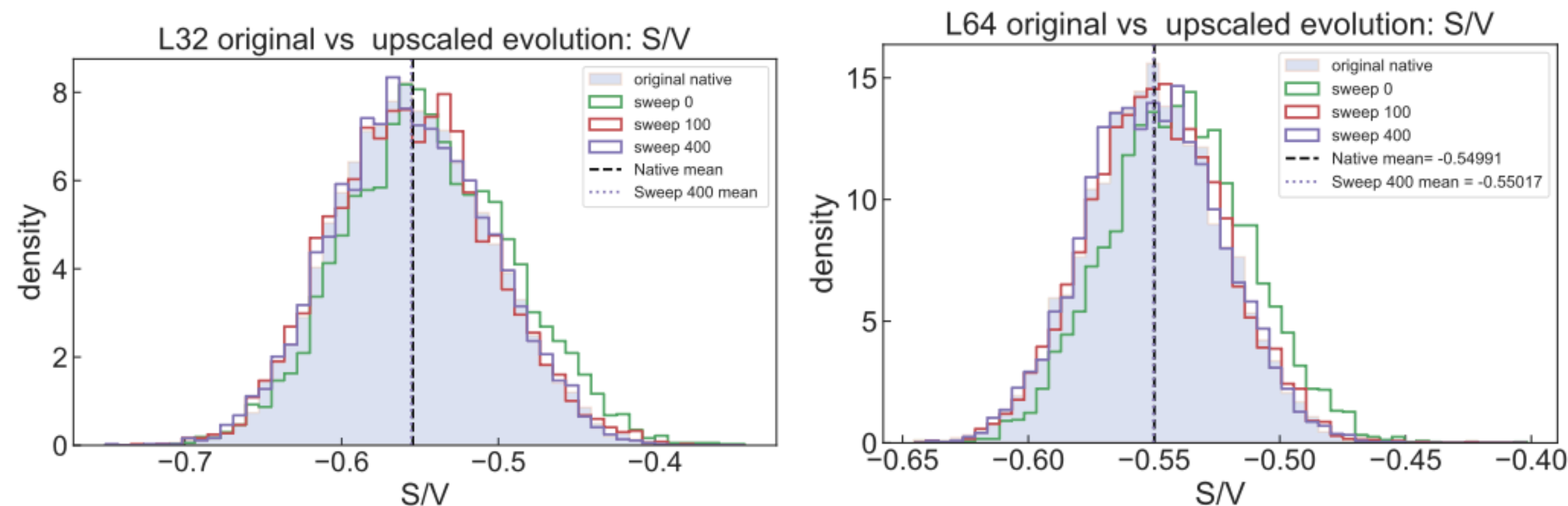


Figure 4. Rethermalization of the action-density distribution after 0, 100, and 400 sweeps.

Proposed update steps:

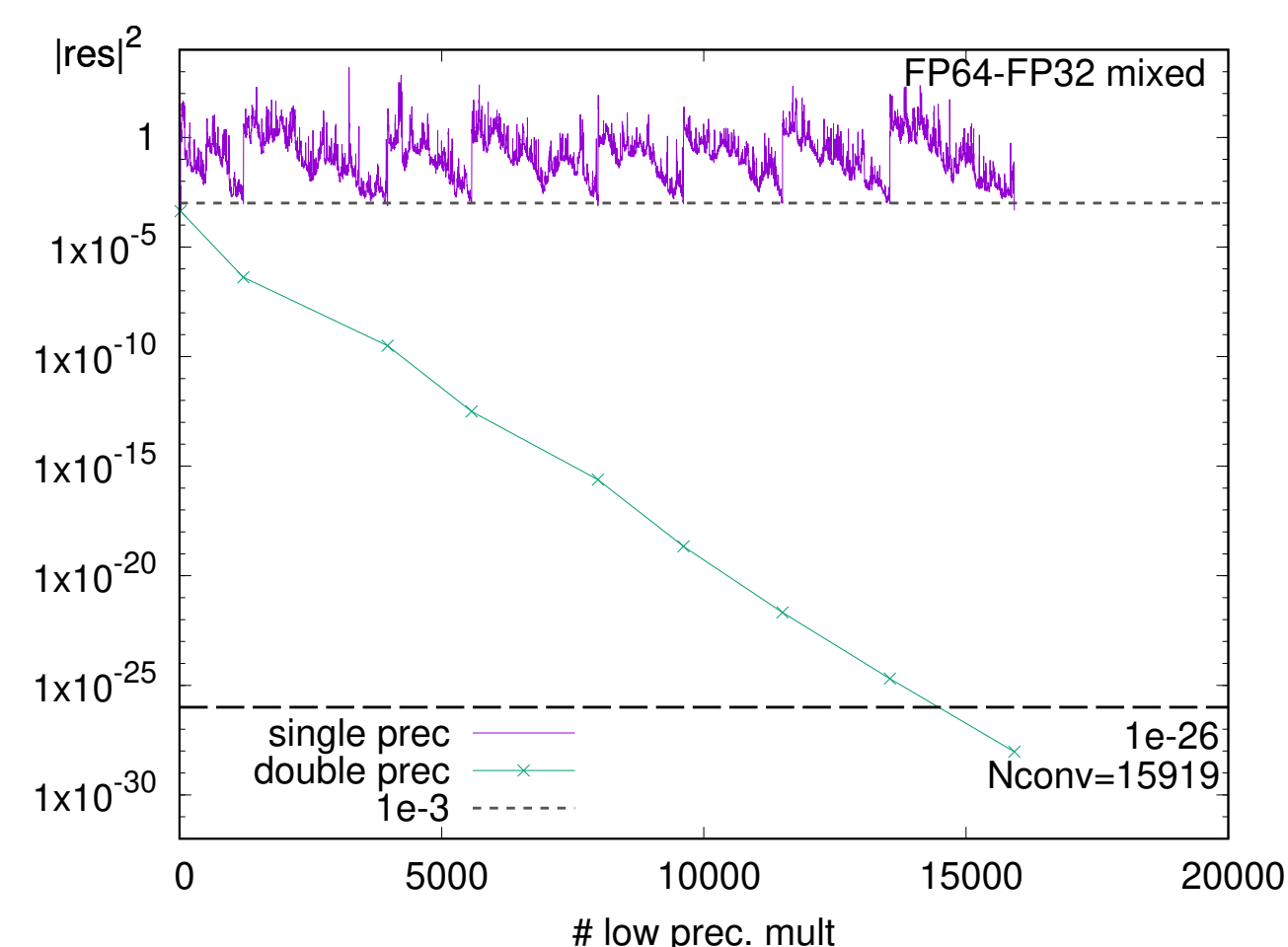
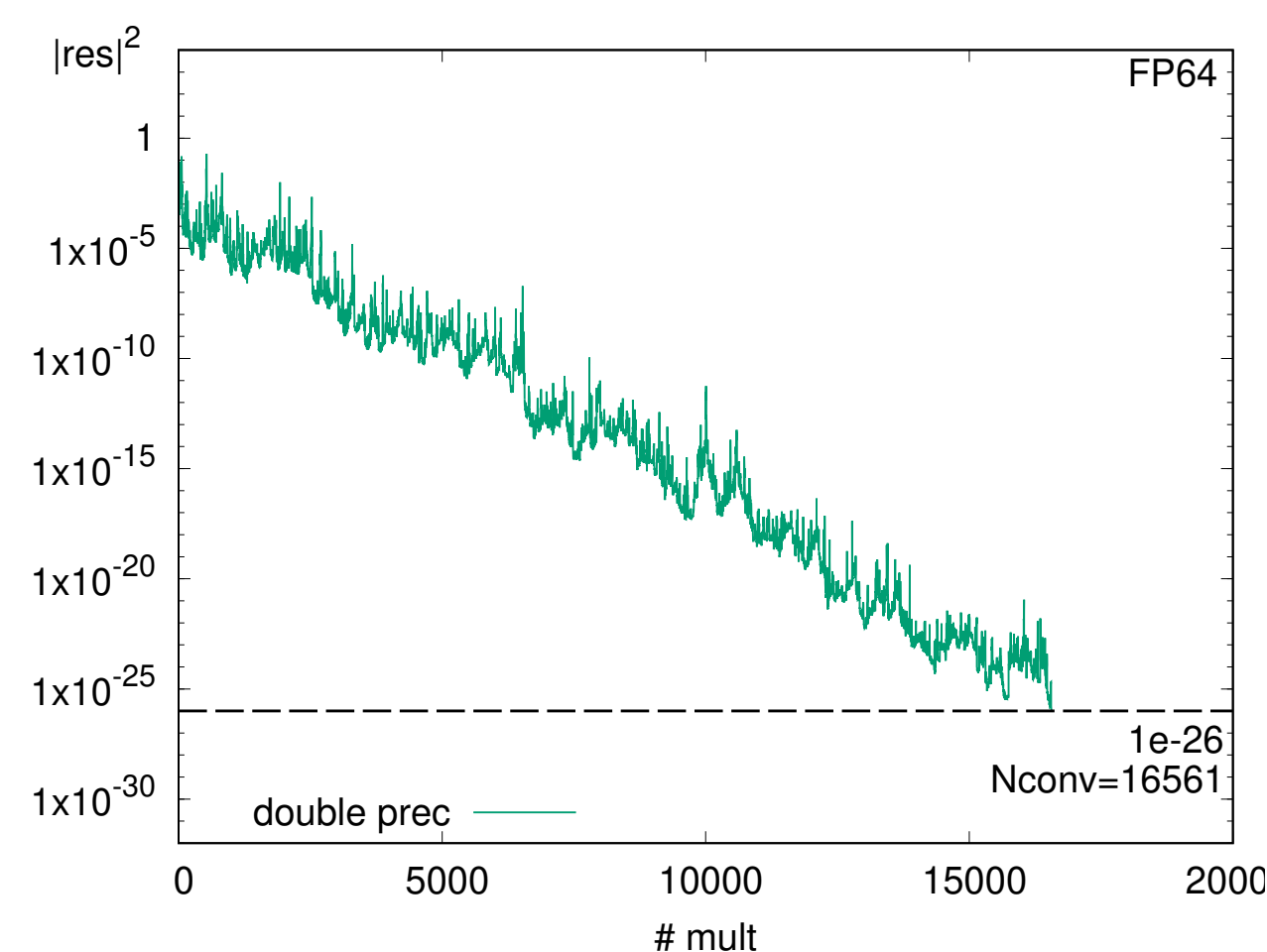
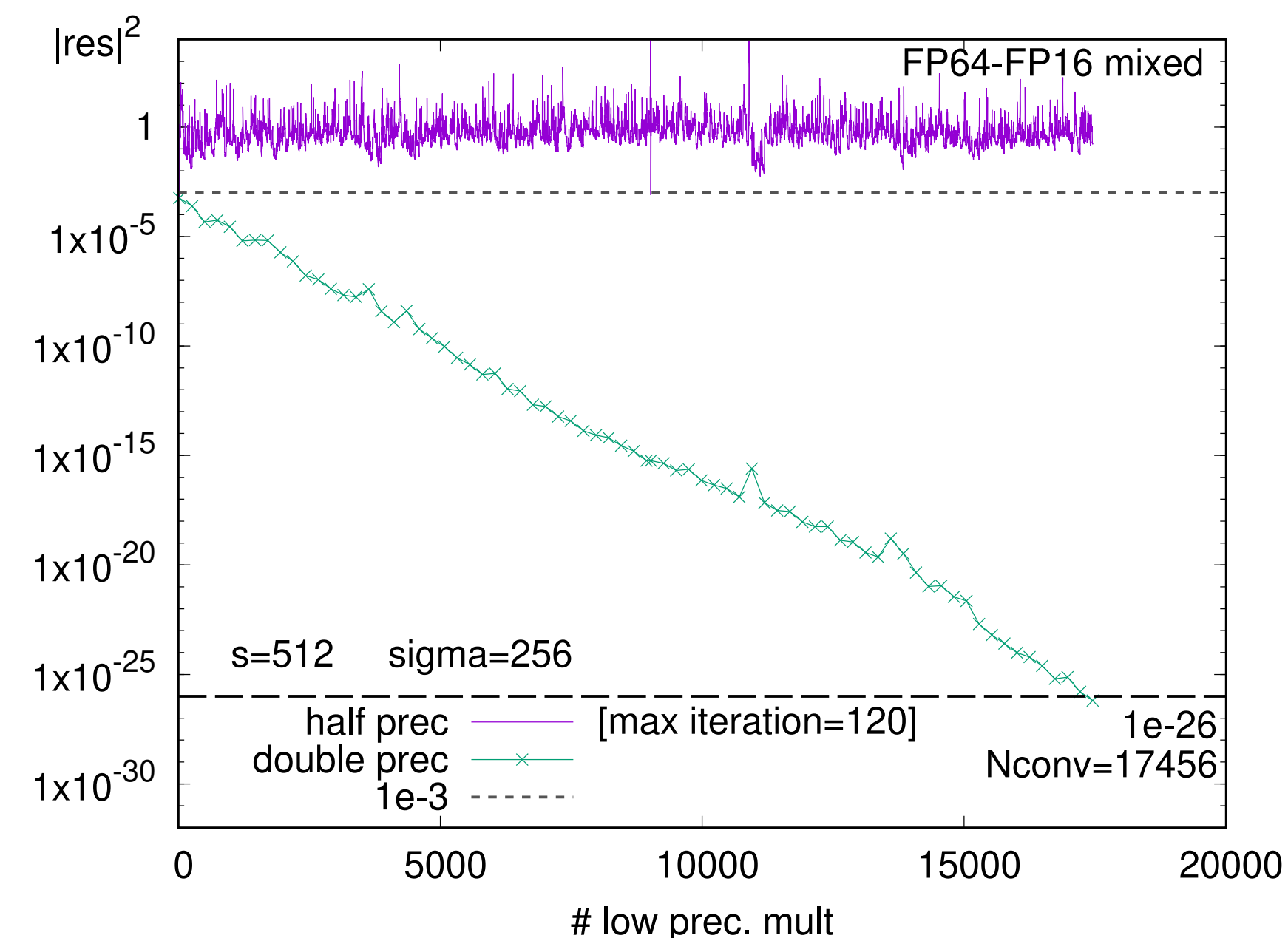
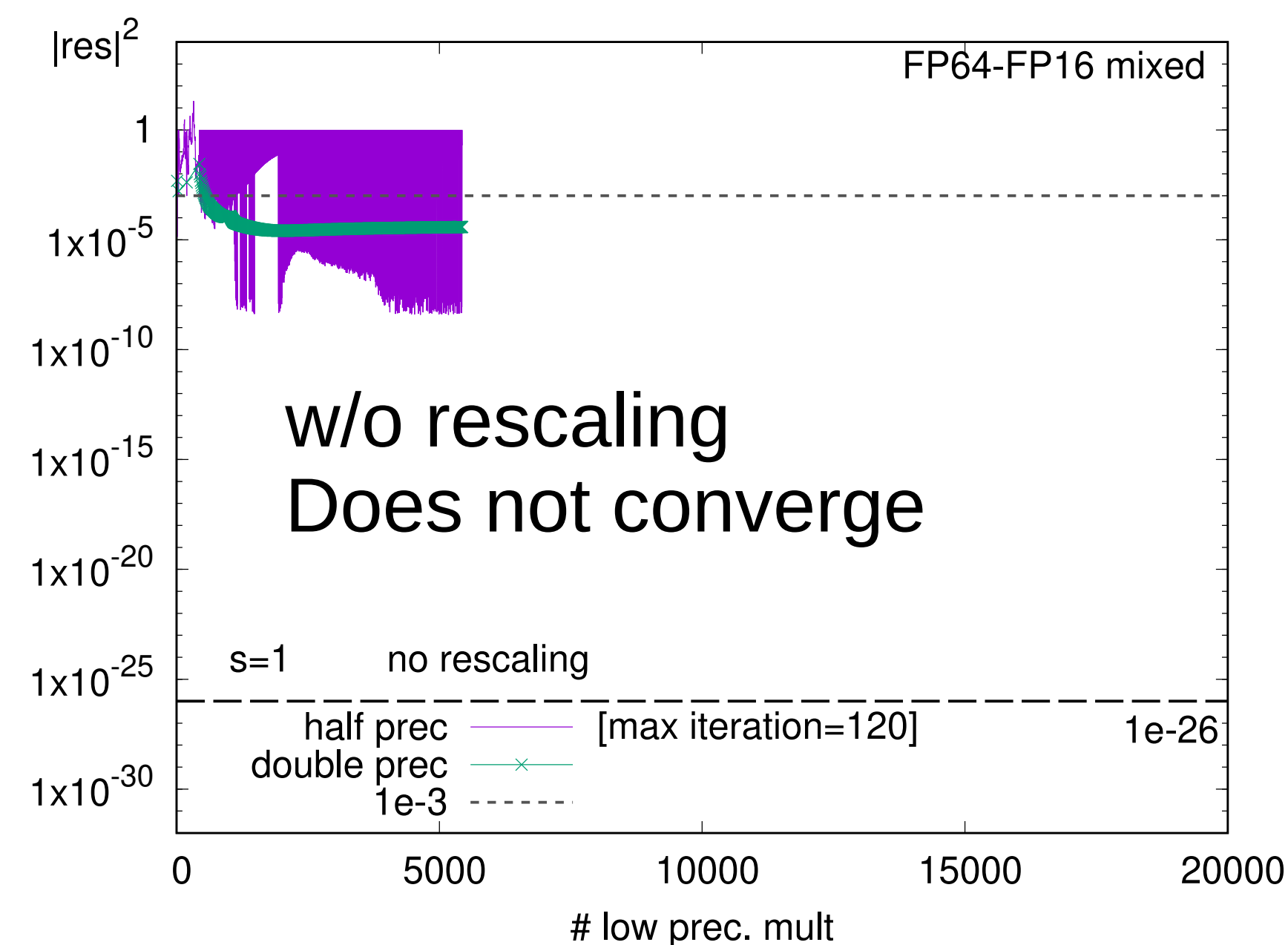
- Generate coarse configurations on $(L/2)^d$
- Upscale to L^d with the conditional flow
- Rethermalize with exact coarse and detail patch updates
- Retain one configuration per chain at a fixed sweep

solvers, Dirac systems

- Issaku Kanamori: A Stabilized FP16 Mixed-Precision Solver, 7/27 2 PM
- Benjamin Luke: Optimized solver algorithms for computing $\text{Tr}(A^{-1})$, 7/29 10:20 AM
- Patrick Oare: Spectral Properties of the non-Hermitian Dirac Operators from the Krylov-Schur Algorithm, 7/29 11:10 AM
- Chulwoo Jung: Topological Modes in Lattice QCD evolution, 7/29 11:30 AM

FP16 solver with rescaled working vectors (even-odd prec + BiCGStab + iterative refinement)

Issaku Kanamori 7/27 2 PM



	time [sec]	speed up
FP64 BiCGStab	142.2 sec	---
FP64-FP32 mixed	67.3 sec	2.1
FP64-FP16 mixed	45.3 sec	3.1

PACS 96^4 lattice, Clover, $M_{\pi} \sim 146$ MeV
[PoS LATTICE2015, 075 (2016)]
Supercomputer Fugaku, 144 nodes
Code: Bridge++

Accelerating iterative linear equation solver using modified domain-wall fermion matrix in lattice QCD simulations

Wei-Lun Chen¹[0000-0002-8542-9126], Issaku Kanamori²[0000-0003-4467-1052],
Hideo Matsufuru³[0000-0003-1056-3969], and Hartmut Neff⁴

¹ Particle and Nuclear Physics, Graduate Institute for Advanced Studies, Graduate University for Advanced Studies (SOKENDAI), Oho 1-1, Tsukuba 305-0801, Japan
wlchen@post.kek.jp

² RIKEN Center for Computational Science 7-1-26 Minatojima-minami-machi, Chuo-ku Kobe, Hyogo 650-0047, Japan kanamori-i@riken.jp

³ Computing Research Center, High Energy Accelerator Research Organization (KEK) and Accelerator Science Program, Graduate University for Advanced Studies (SOKENDAI), Oho 1-1, Tsukuba 305-0801, Japan hideo.matsufuru@kek.jp

⁴ Luzernerstrasse 43, 6330, Switzerland hartmutneff@aol.com

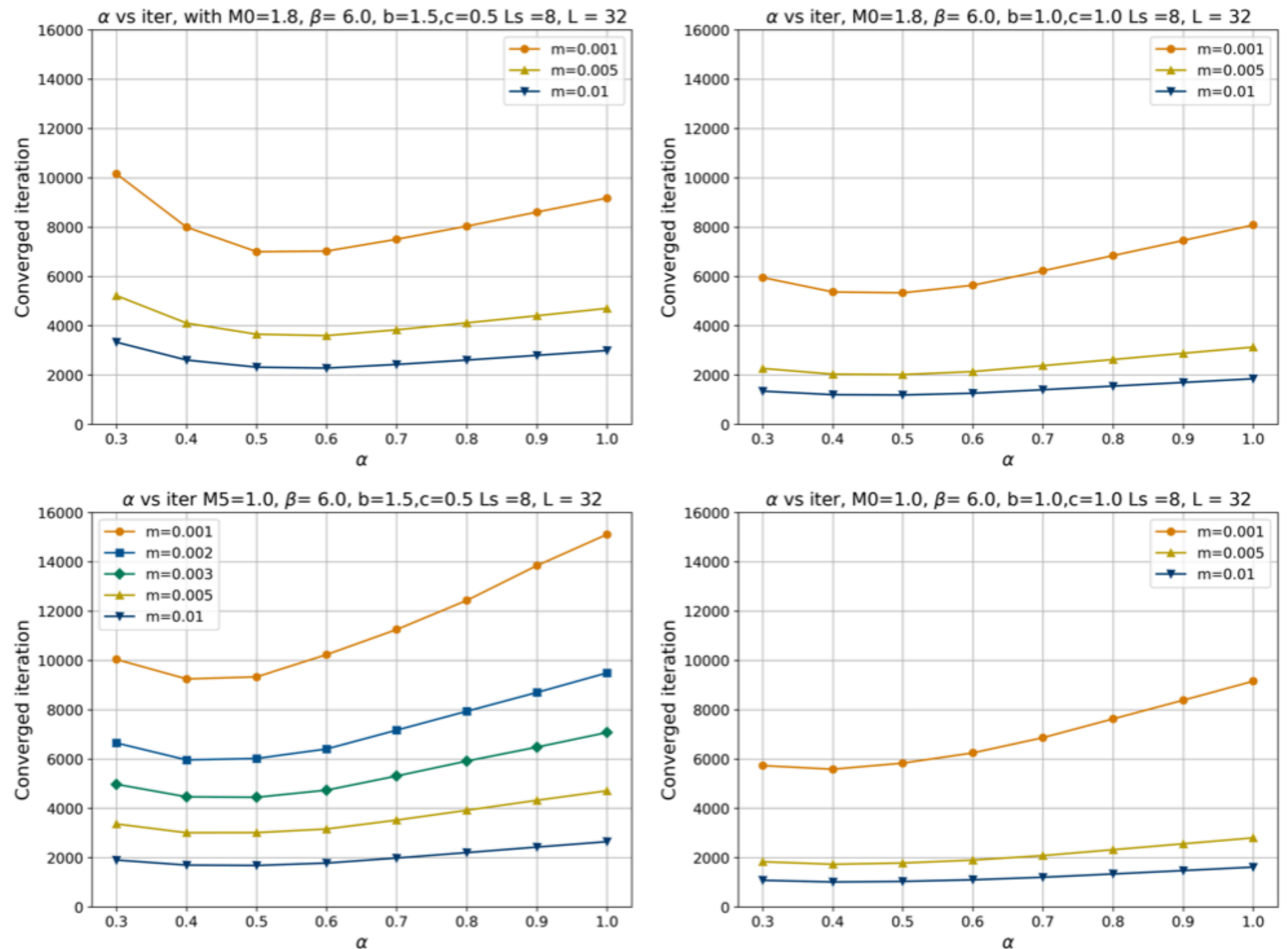


Fig. 5. Converged numbers of CG iterations on 32^4 lattice at $\beta = 6.0$. Top panels show $M_0 = 1.8$ (without smearing), $(b, c) = (1.5, 0.5)$ (left) and $(b, c) = (1.0, 1.0)$ (right). Bottom panels show $M_0 = 1.0$ (with smearing), $(b, c) = (1.5, 0.5)$ (left) and $(b, c) = (1.0, 1.0)$ (right). In both cases, $L_s = 8$ is used.

Figure 5 shows the results of the converged number of CG iterations for unsmeared and smeared cases with $(b, c) = (1.5, 0.5)$ and $(1.0, 1.0)$. The acceleration of CG convergence for $m = 0.001$, the severest one of the measured quark mass, is in the range of 24% (top-left panel)–40% (bottom-left).

non-standard Monte Carlo dynamics

- James Osborn: Generalized Dynamics for HMC, 7/28 2 PM
- Robert Marwhinney: Studying the impact of multiple, localized, gauge-fixed subvolumes with coherent conjugate momenta on the Hybrid Monte Carlo, 7/28 2:20 PM
- Yacob Ozdalkiran: Event-Chain Monte Carlo for pure gauge lattice QCD, 7/28 3:20 PM

Generalized Dynamics for HMC

K. Harrison^{1,2}, X.-Y. Jin¹ and J.C. Osborn¹

¹ Argonne National Laboratory

² North Carolina State University

- A general solution

$$F_i(z) = C_i(z)e^{H_i} + \left(\frac{dH}{dz_j} - \frac{d}{dz_j}\right) G_{ij}(z)$$

$H_i(z)$ is all parts of $H(z)$ that depend on z_i

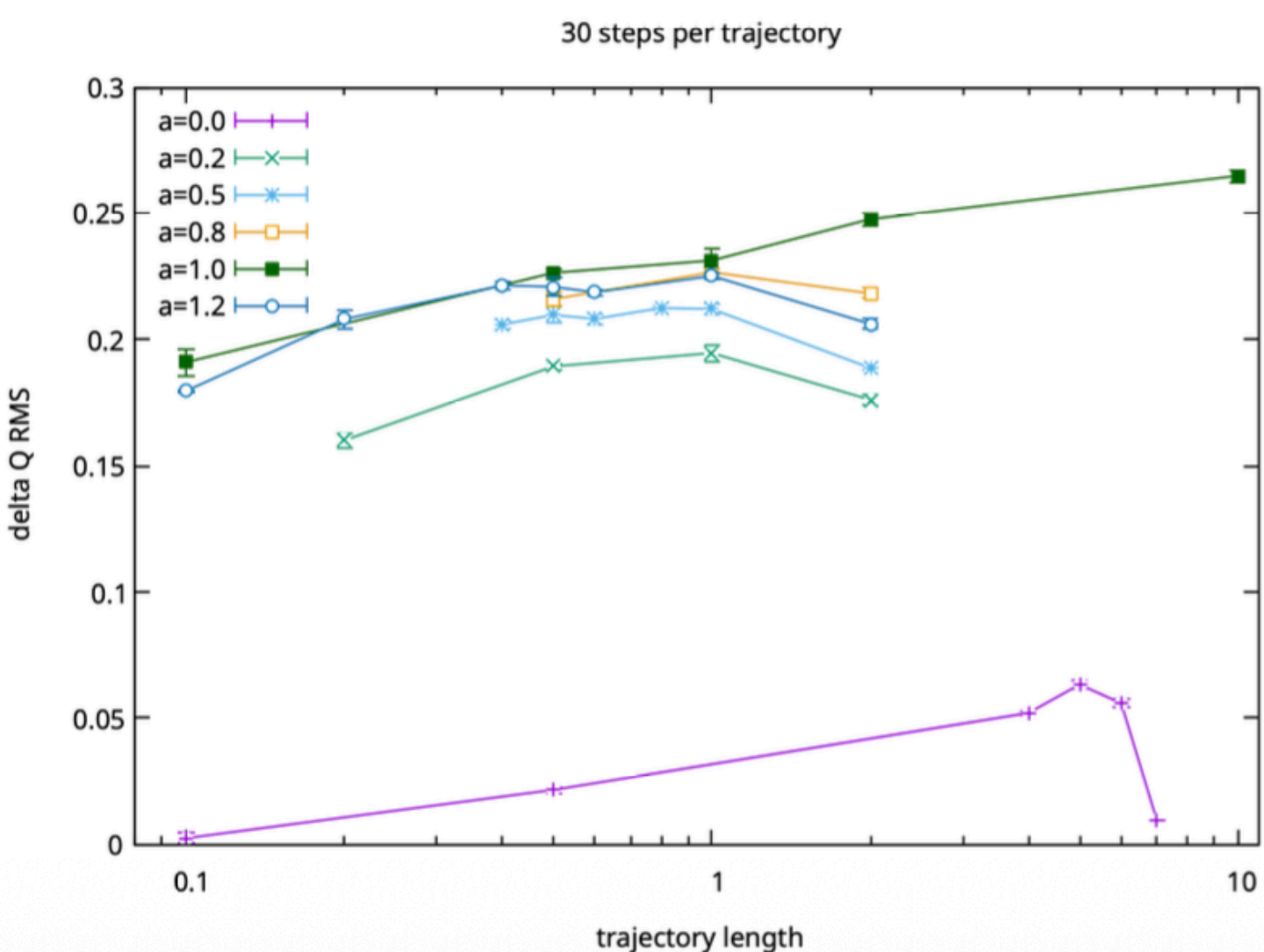
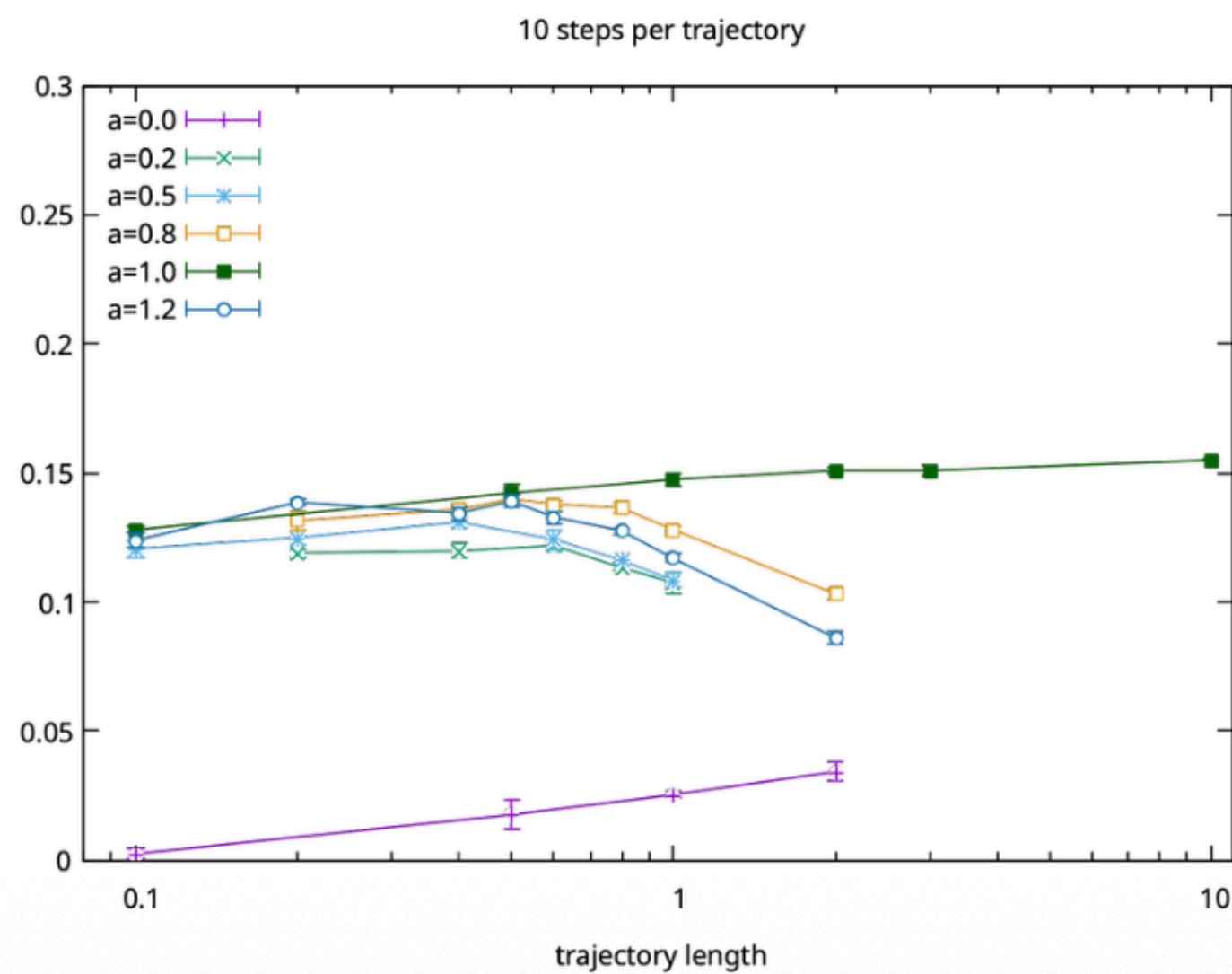
$C_i(z)$ does not depend on z_i

$$G_{ij}(z) = -G_{ji}(z)$$

Reproduces FTHMC and
RMHMC by choosing
the appropriate G_{ij}

Topology
tunneling rate

- Define tunneling rate
 $\text{delta Q RMS} = \sqrt{\sum (Q_{i+1} - Q_i)^2 / (N - 1)}$
- GDHMC has 4-5x tunneling rate of HMC



Generalized Dynamics for HMC, Lattice 2026

$$G_i^{\theta p} = (1 - a) + ae^{V_i}$$
$$\dot{p}_i = (1 - a)F_i$$

- Thank all the authors of the slides/papers that I copy-pasted here
- Thank all of you who send me your materials: Akia Tomiya, Anna Hasenfratz, Claudio Bonanno, Daniel Nogradi, Hartmut Neff, Issaku Kanamori, Nobuyuki Matsumoto, Timo Eichhorn, Yoshinobu Kuramashi
- Used ChatGPT for searching the web