TACC
TEXAS ADVANCED COMPUTING CENTER
WWW.TACC.UTEXAS.EDU

TEXAS
The University of Texas at Austin

Application Performance Analysis: a Report on the Impact of Memory Bandwidth

ISC 2023 IXPUG Workshop

Yinzhi Wang, John D. McCalpin, Junjie Li, Matthew Cawood, John Cazes, Hanning Chen, Lars Koesterke, Hang Liu, Chun-Yaung (Albert) Lu, Robert McLay, Kent Milfield, Amit Ruhela, Dave Semeraro, and Wenyang Zhang

Texas Advanced Computing Center, The University of Texas at Austin

NSF

Context

- The Texas Advanced Computing Center at the University of Texas at Austin deploys and runs the largest academic supercomputer systems in the United States.
- The results presented today were obtained as part of our ongoing studies of the performance characteristics of applications and systems, with particular focus on applications expected to be important in our next-generation capability system. *("Characteristic Science Applications")*
- The results today were from single-node testing on 2-socket nodes with Xeon Max 9480 processors. One node was configured with DDR5 memory (HBM disabled), and the other was configured with HBM memory only (DDR5 DIMMs removed)

STREAM Benchmark Results

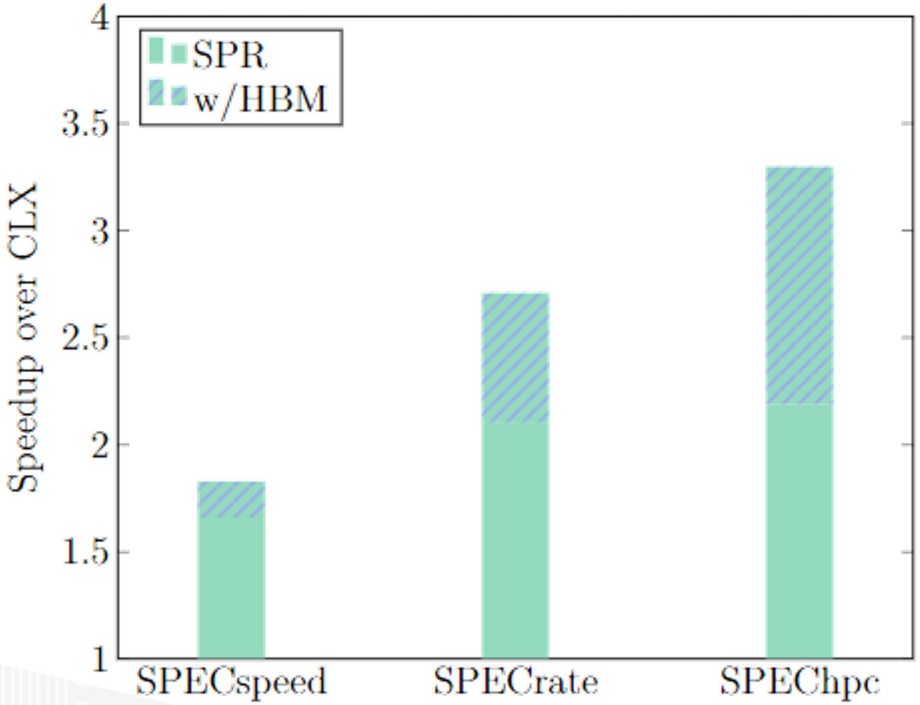
Platform	Sockets	Cores	Copy & Scale	Add & Triad	Peak BW	%Peak
CLX	2	56	204.4	220.4	281.6	73% - 78%
SPR w/DDR5	2	112	377.4	398.3	614.4	61% - 65%
SPR w/HBM	2	112	1371.4	1371.8	3276.8	~42%

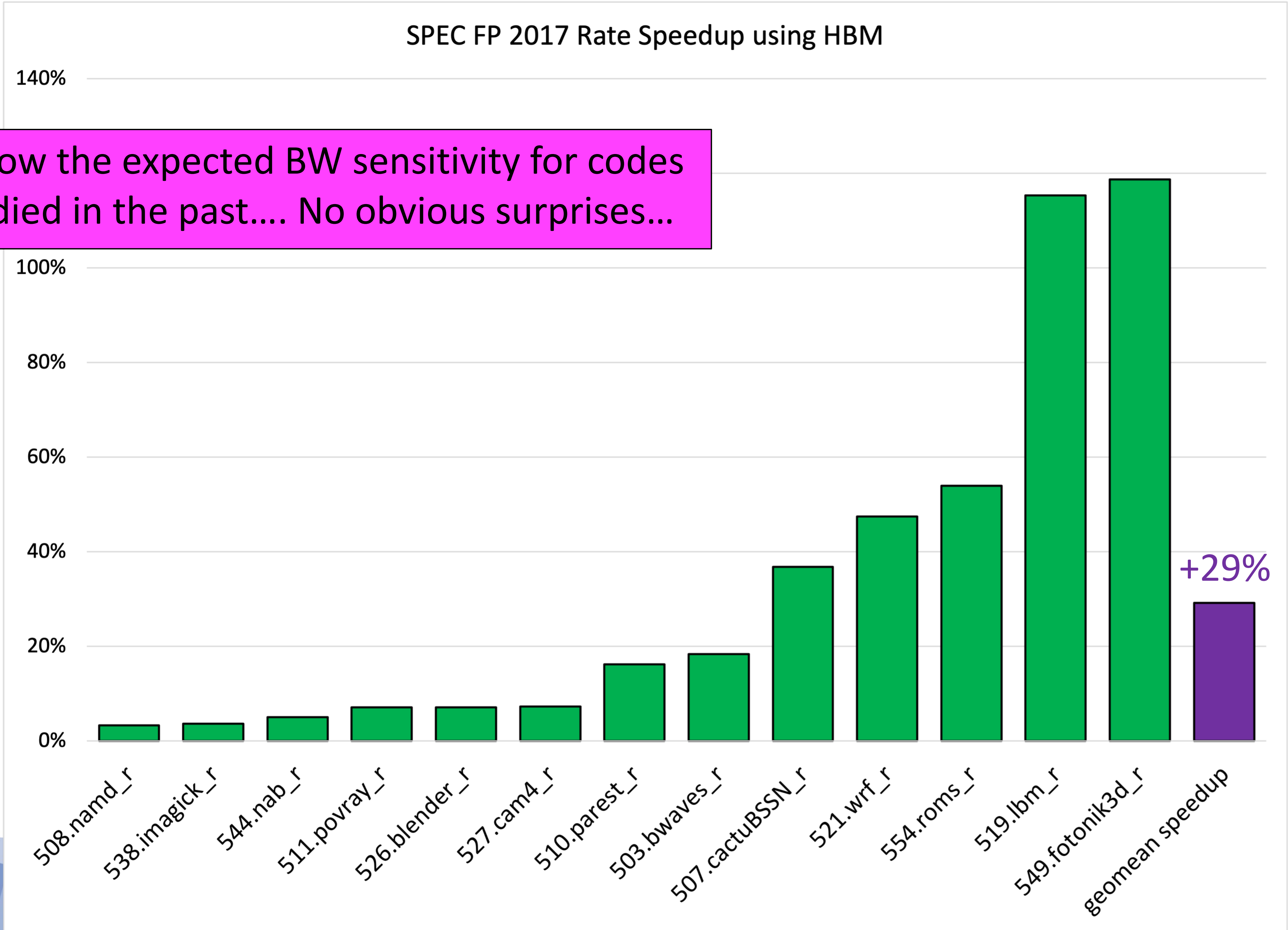
Notes:

1. CLX system has one dual-rank DDR4/2933 DIMM per channel
2. SPR/DDR5 system has one single-rank DDR5/4800 DIMM per channel
3. STREAM performance on SPR/HBM is more variable than on most recent systems – causes are still under investigation
4. Results for CLX require compilation with streaming stores.
5. These SPR results also used streaming stores, but the performance impact is only ~3%.

SPEC Benchmark Scores

Benchmark	CLX	SPR w/DDR5	SPR w/HBM	HBM uplift
SPECspeed2017_fp_base	169	280	308	+10%
SPECrate2017_fp_base	350	736	950	+29%
SPEChpc 2021_tny_base	3.15	6.89	10.4	+52%

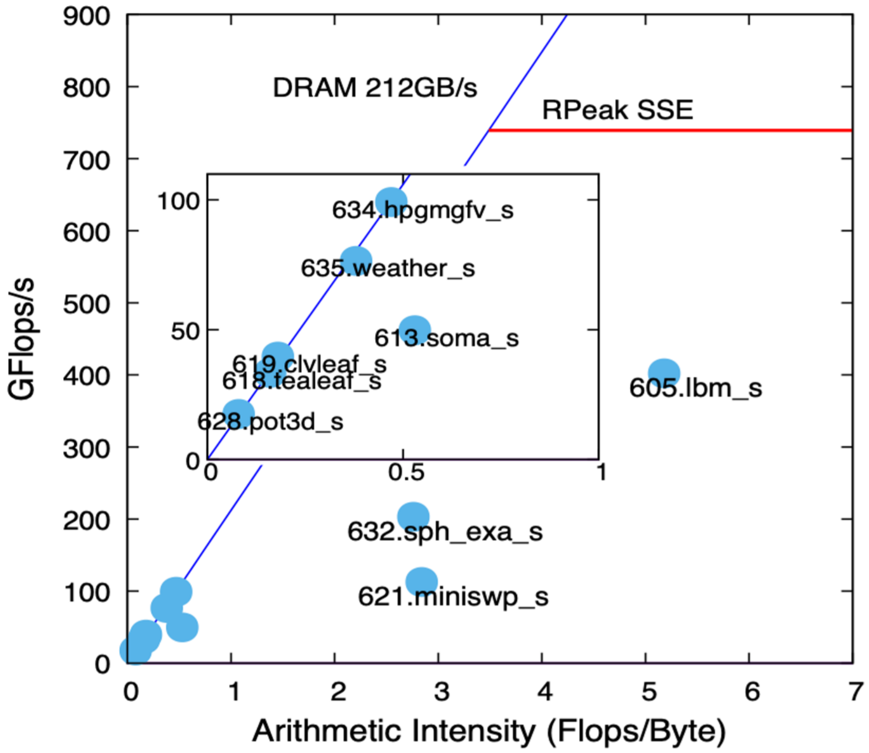




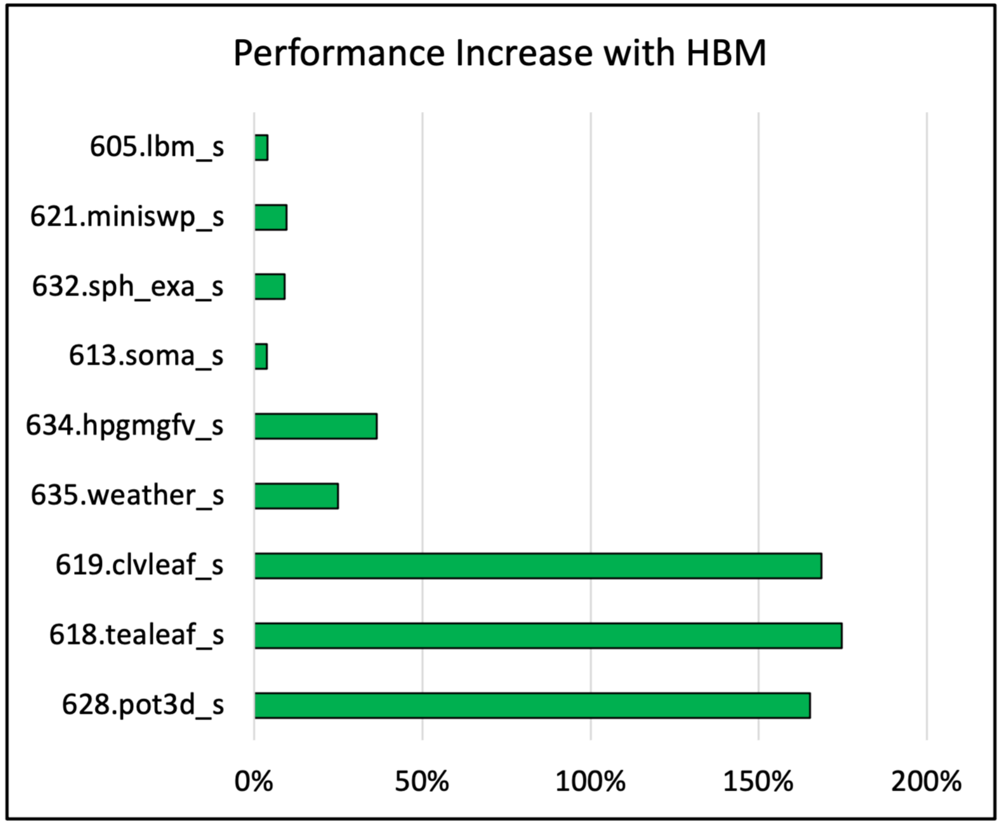
Results show the expected BW sensitivity for codes I have studied in the past…. No obvious surprises…

Tests run by TACC & do not precisely match any vendor submissions…

SPEC HPC 2020 benchmark speedup on HBM matches expectations from roofline analysis

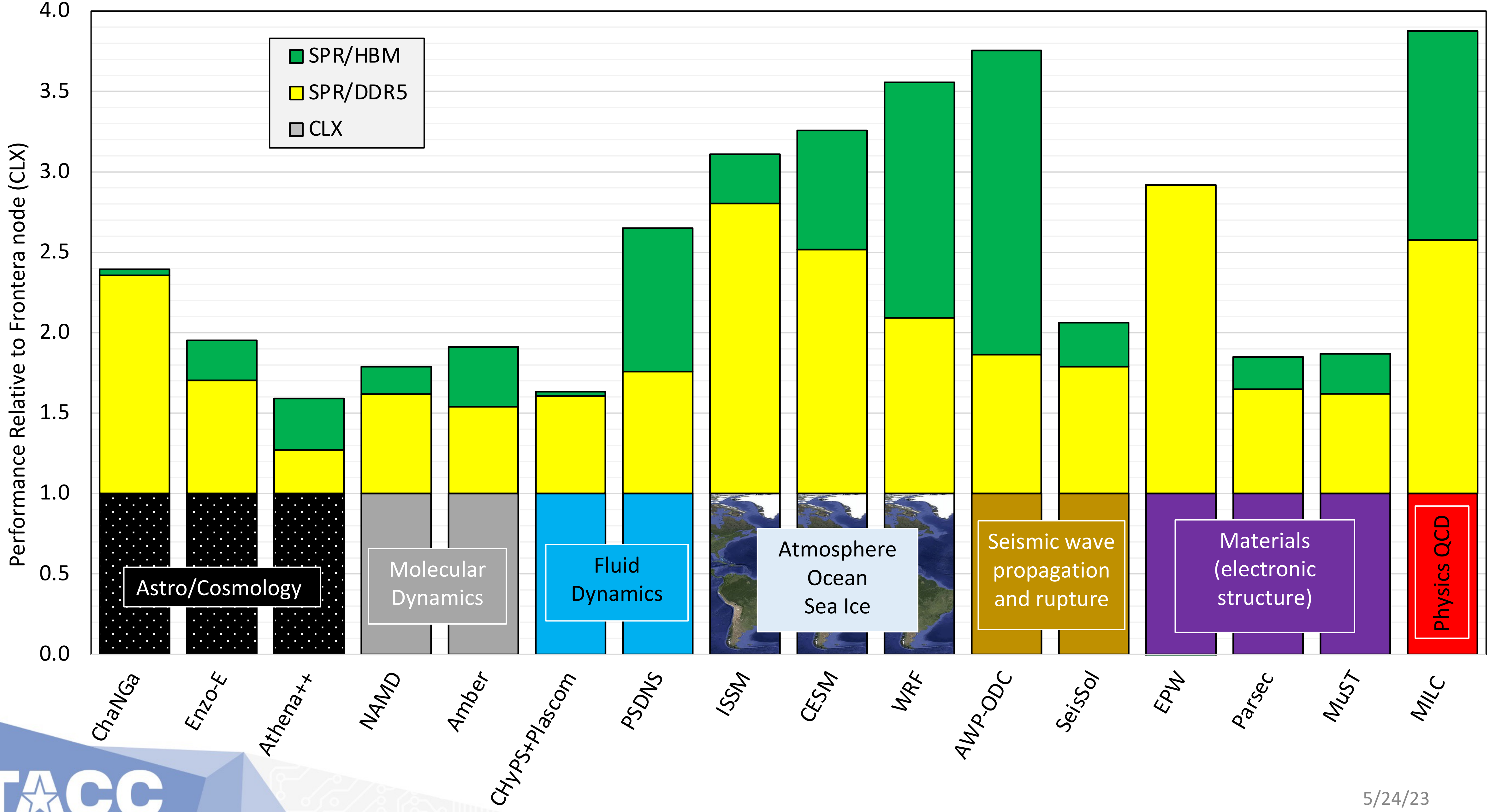


(From: https://arxiv.org/pdf/2203.06751.pdf)

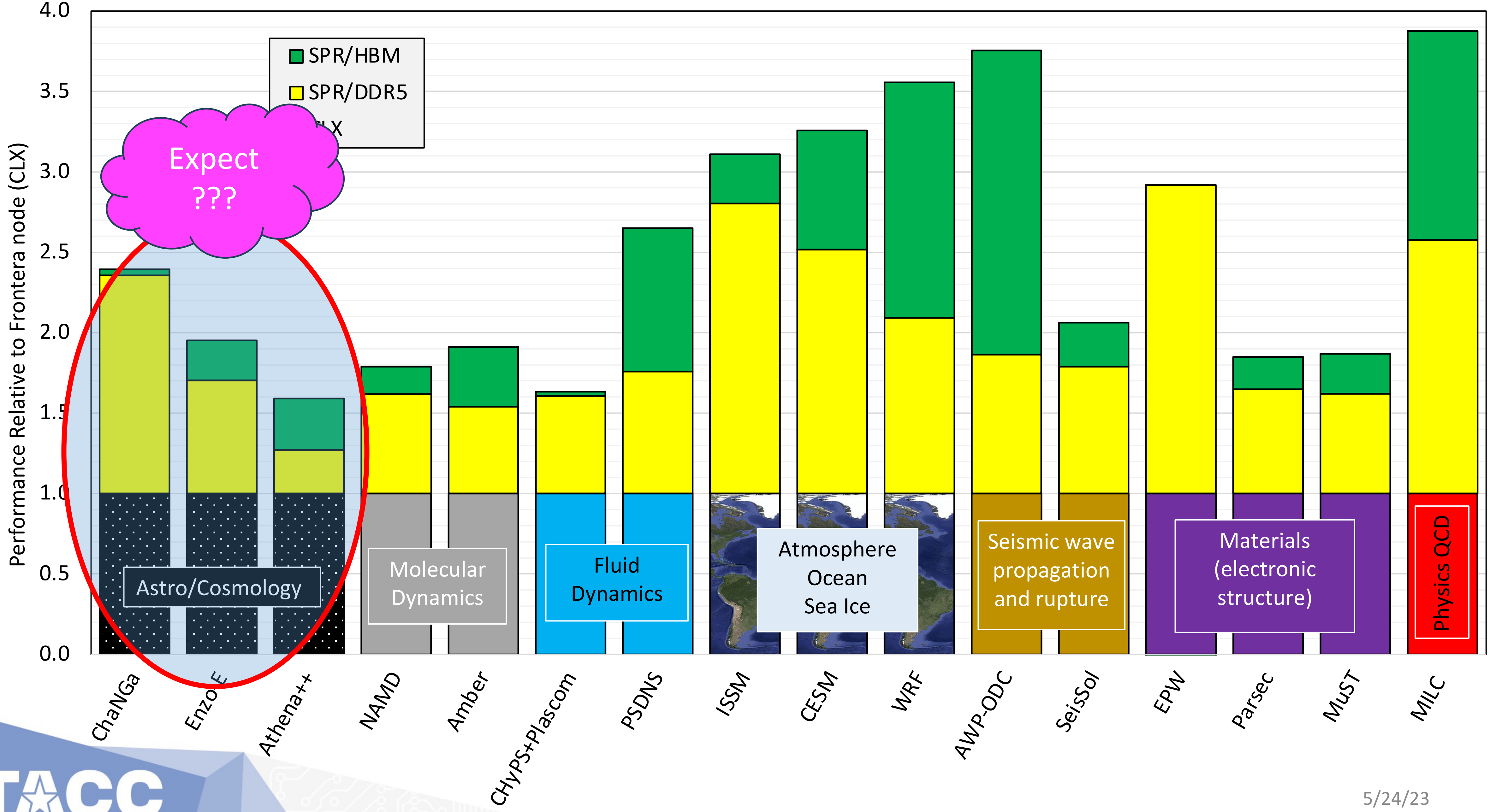


- Large speedups on the three codes in the lower left corner
- Modest speedups on two codes further up the BW-limited diagonal
- Negligible speedups for codes away from the diagonal.

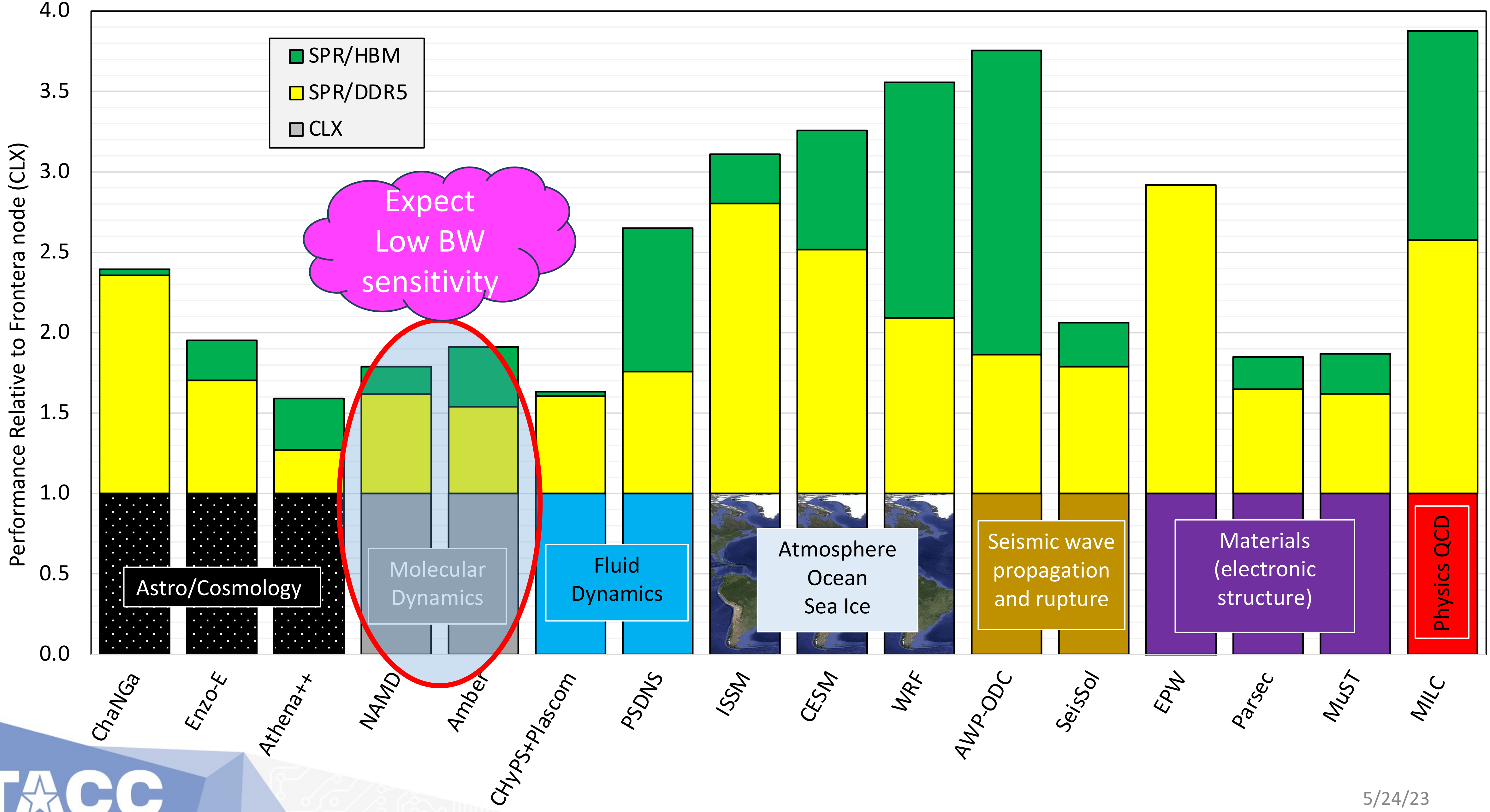
Single-Node Application Performance Ratios



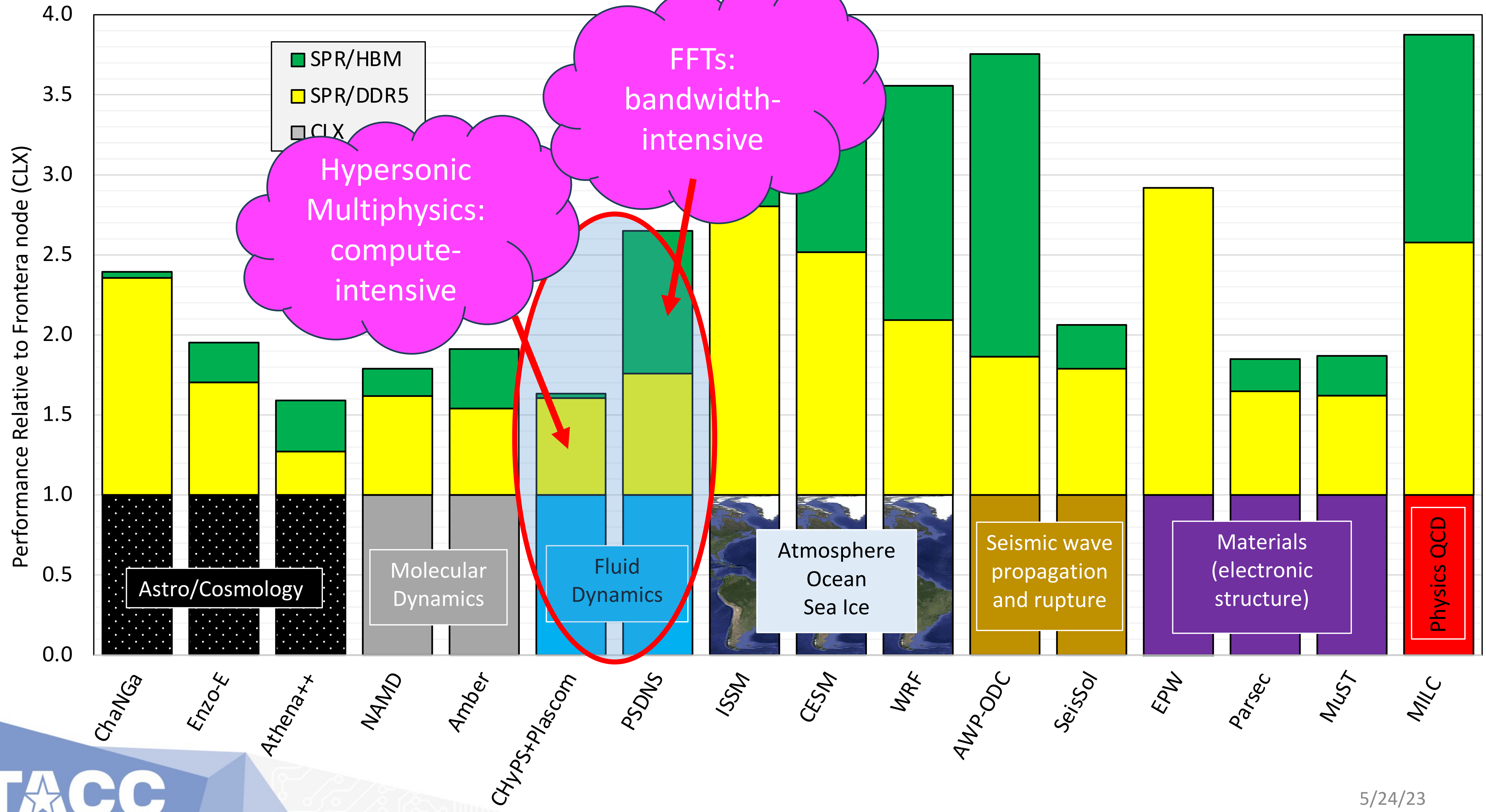
Single-Node Application Performance Ratios



Single-Node Application Performance Ratios



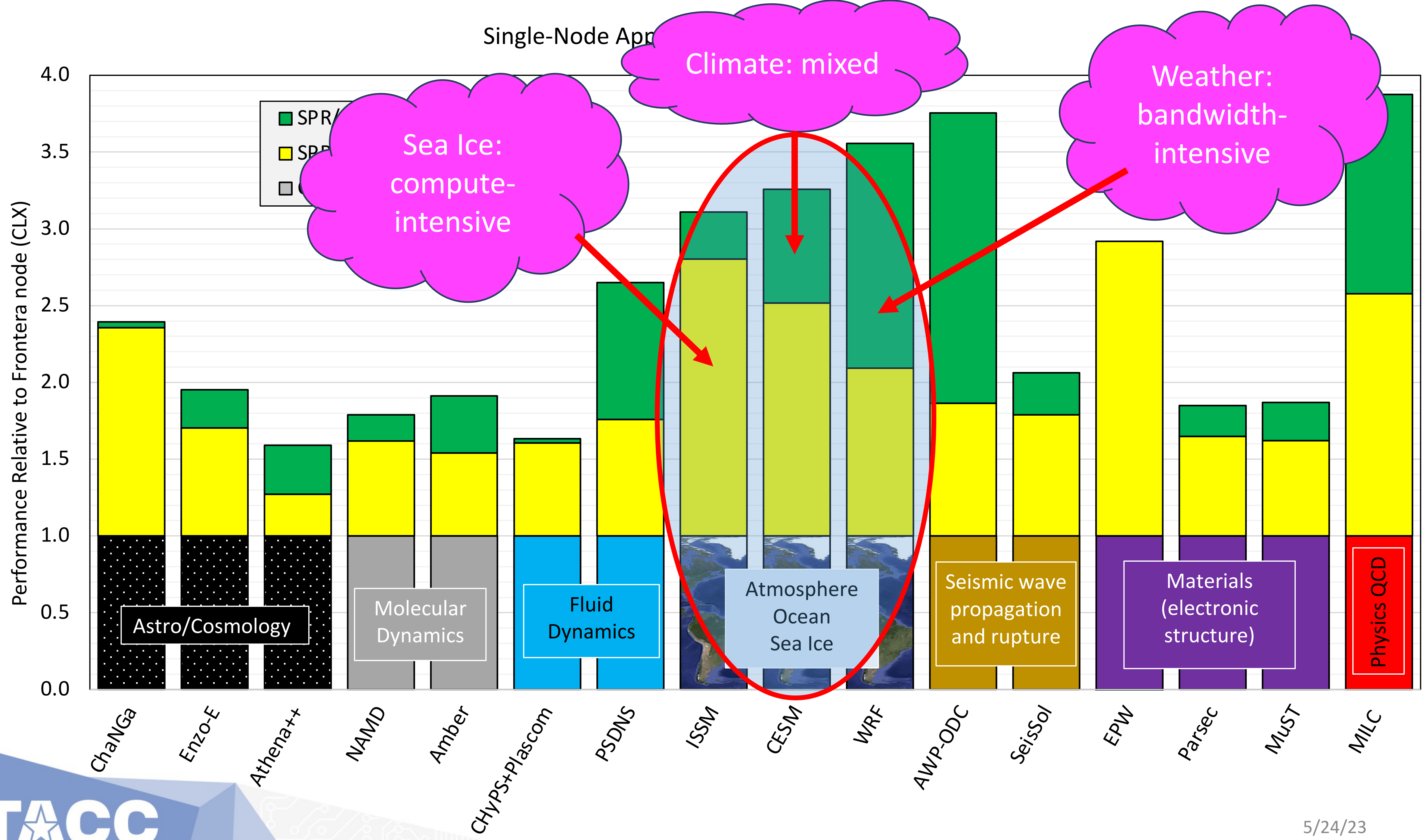
Single-Node Application Performance Ratios



Hypersonic Multiphysics: compute-intensive

FFTs: bandwidth-intensive

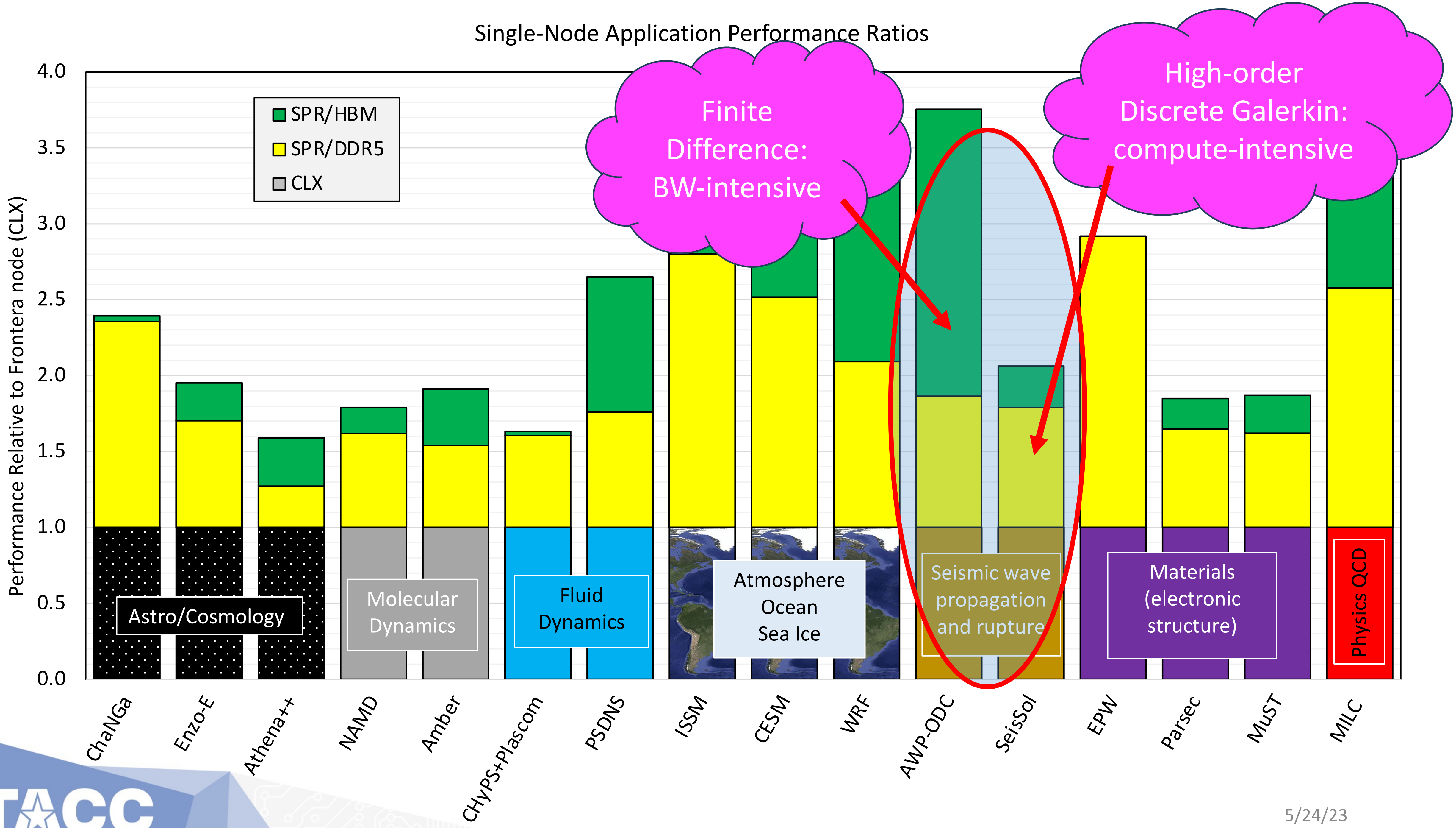
Single-Node App

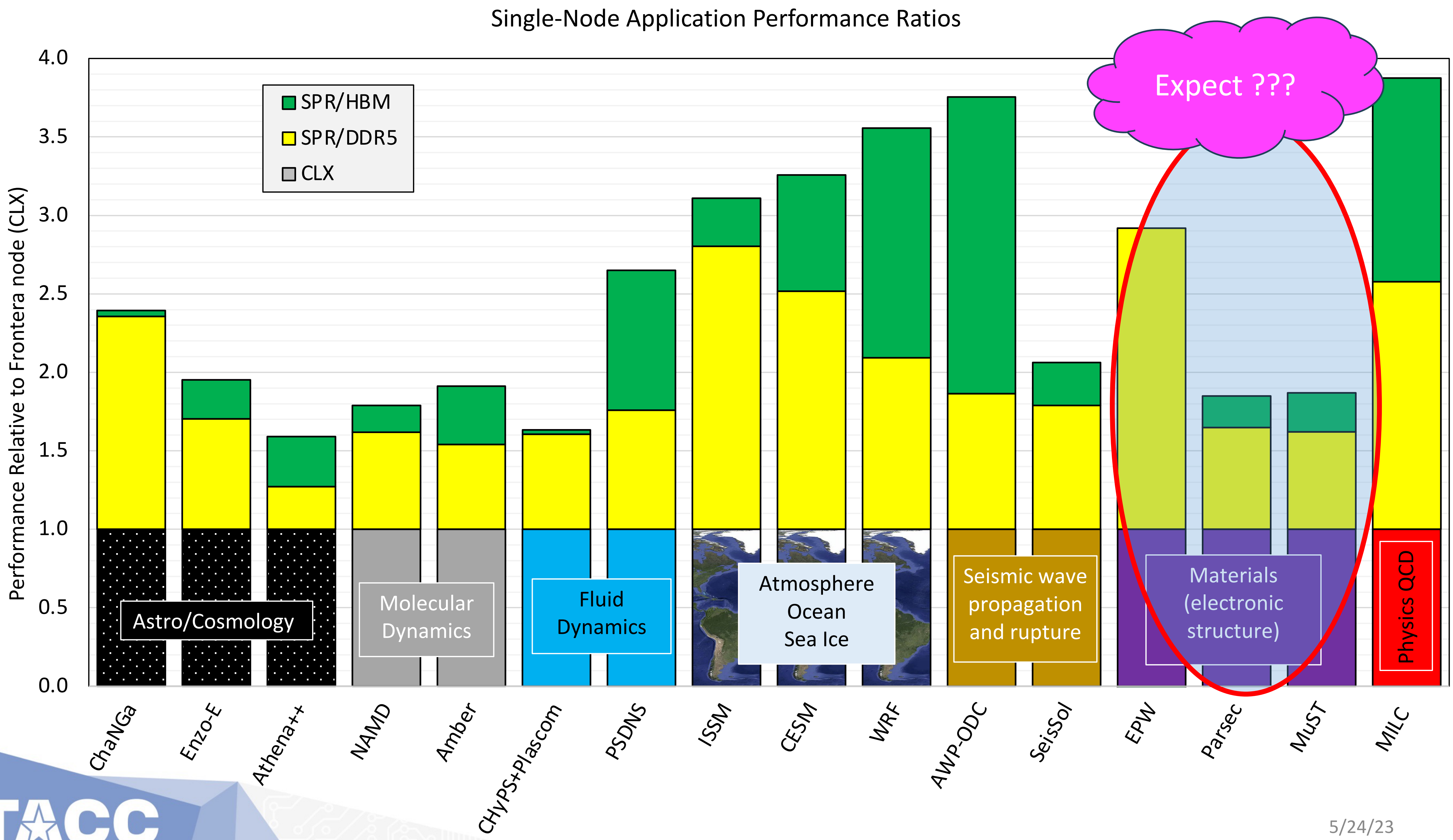


Sea Ice: compute-intensive

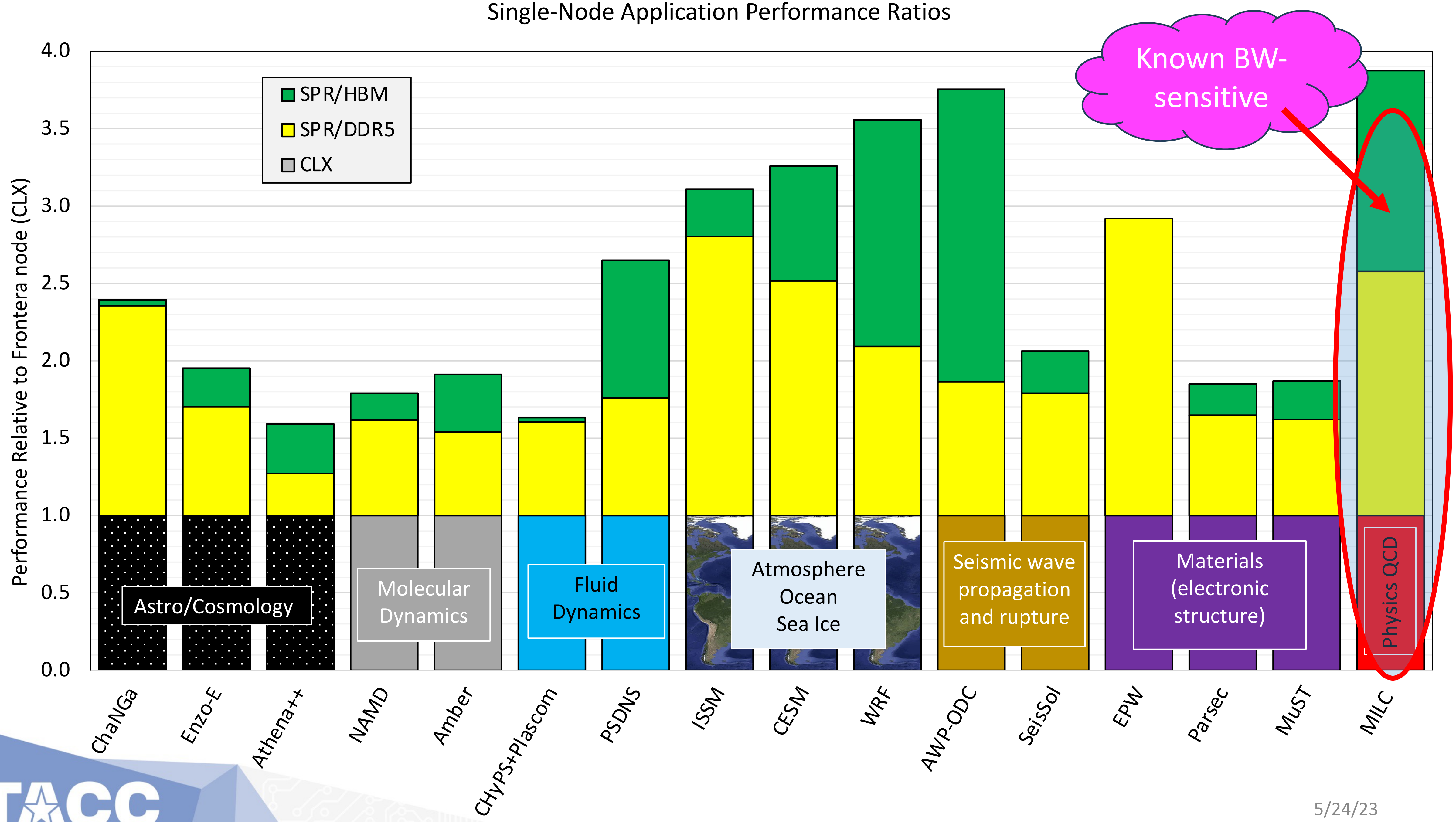
Climate: mixed

Weather: bandwidth-intensive


Single-Node Application Performance Ratios
Performance Relative to Frontera node (CLX)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
🟩 SPR/HBM
🟨 SPR/DDR5
▨ CLX
Finite Difference: BW-intensive
High-order Discrete Galerkin: compute-intensive
Astro/Cosmology
Molecular Dynamics
Fluid Dynamics
Atmosphere Ocean Sea Ice
Seismic wave propagation and rupture
Materials (electronic structure)
Physics QCD
ChaNGa
Enzo-E
Athena++
NAMD
Amber
CHyPS+Plascom
PSDNS
ISSM
CESM
WRF
AWP-ODC
SeisSol
EPW
Parsec
MuST
MILC


Single-Node Application Performance Ratios
Performance Relative to Frontera node (CLX)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
🟩 SPR/HBM
🟨 SPR/DDR5
□ CLX
Expect ???
Astro/Cosmology
Molecular Dynamics
Fluid Dynamics
Atmosphere Ocean Sea Ice
Seismic wave propagation and rupture
Materials (electronic structure)
Physics QCD
ChaNGa
Enzo-E
Athena++
NAMD
Amber
CHyPS+Plascom
PSDNS
ISSM
CESM
WRF
AWP-ODC
SeisSol
EPW
Parsec
MuST
MILC

Single-Node Application Performance Ratios



Summary

- The Xeon Max processor with HBM memory provides a large (>3x) increase in sustained memory bandwidth relative to the DDR5 memory available on the processor (and on other 4th-generation Xeon Scalable Processors)
- Overall speedups on SPEC benchmarks:
  - SPEC fp rate 2017: +3% to +119%, geomean +29%, 7 of 13 > 15%
  - SPEC HPC “tiny”: +3% to +175%, geomean +52%, 5 of 9 > 25%
- Characteristic Science Applications:
  - -2% to +100%, geomean +25%, 7 of 16 > 24%
- These are solid performance improvements for > 1/3 of the workloads tested