

TACC: SITE UPDATE

The background features a stylized, low-poly illustration of a figure in a blue and white suit, possibly a scientist or researcher, standing on a dark, rocky surface. The figure is holding a glowing blue object. The background is a deep blue with a network of white lines and nodes, resembling a circuit or data flow. The overall aesthetic is futuristic and technological.

Lars Koesterke

Researcher/Manager, High Performance Computing

The University of Texas at Austin

April 15, 2025

A VERY WARM WELCOME TO TACC AND TO IXPUG

- ▶ TACC is a big x86 'shop'
- ▶ I started at TACC in 2007; Ranger made the top-10 in 2008
- ▶ Stampede 1 and 2
- ▶ Lonestar 4, 5, and 6
- ▶ Frontera
- ▶ CPU performance is very important to our broad user base
- ▶ Horizon will have a large CPU partition
 - ▶ Other centers may have 98-99% of performance in GPU

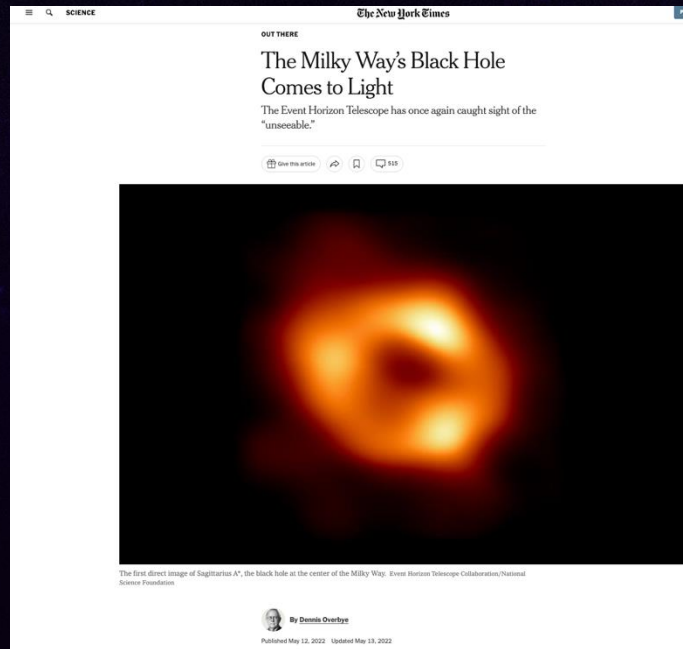


TACC IN A NUTSHELL

- ▶ We operate the Frontera, Stampede3, Vista, and other systems for the National Science Foundation
- ▶ Lonestar6 for our Texas academic and industry users.
- ▶ Altogether, ~20k servers, >1M CPU cores, 1k GPUs
- ▶ About seven billion core hours over several million jobs per year – for 3,000 projects and ~40,000 users per year.
- ▶ 190 Staff (~70 PhD)



ALL FOR UNCLASSIFIED, OPEN SCIENCE



WIDEST VIEW OF EARLY UNIVERSE HINTS AT GALAXY AMONG THE EARLIEST EVER DETECTED

TACC's Frontera, Stampede2 supercomputers enable scientists to combine myriad images, hold frames in memory resulting in single image

Published on August 4, 2022 by Marc Airhart, CNS / Faith Singer, TACC



Members of the CEERS collaboration explore the first wide, deep field image from the James Webb Space Telescope at the Texas Advanced Computing Center's Visualization Lab on the UT Austin campus on July 21, 2022. Credit: Nolan Zink/UT Austin

Two new images from NASA's James Webb Space Telescope show what may be among the earliest galaxies ever observed. Both images include objects from more than 13 billion years ago, and one offers a much wider field of view than Webb's First Deep Field image, which was released amid great fanfare July 12, 2022.

16 NVDIMM Nodes

Each node contains:

- 4 Intel Xeon Platinum 8280M chips
- 2x 28 core 2.2 Ghz Xeon cores
- 384 GB DRAM
- 2 TB NVMe RAM
- 4 TB NVMe disk



8368 Cascade Lake Nodes

Each node contains:

- 2 Intel Xeon Platinum 8280 chips
- 2x 28 core 2.2 Ghz Xeon cores
- 192 GB DRAM
- Mellanox HDR Infiniband



90 GPU Nodes

Each node contains

- 4 NVIDIA QUADRO RTX 5000 GPUs
- 2 Intel Xeon E5-2620 v4
- 192 GB DRAM

Frontera

Dell 8000+ node cluster
40 Pflops
43 PB Lustre filesystem



SIXTH YEAR OF PRODUCTION ON FRONTERA

- ▶ Frontera is getting old, but is still going strong
- ▶ In the last 12 months:
 - ▶ Uptime of 98%
 - ▶ Average Utilization of 95%
 - ▶ 1M jobs completed
 - ▶ 72M SUs delivered
 - ▶ Zero security incidents
- ▶ Performance of a single node is 'ok'
- ▶ But still 8000+ nodes available; large aggregate memory



USAGE

- ▶ **>2,200 jobs were > 512 nodes**
- ▶ >300 jobs at half or full system scale (Consider if all jobs were full scale, and averages 24 hours, we'd only run 365 jobs a year, as opposed to 1M jobs).
- ▶ Flex jobs, used for backfill, represent 15% of the jobs run, but represent less than 0.5% of SUs delivered (263K out of 70M).
- ▶ Small jobs represent ~30% of jobs, but less than 2% of cycles delivered.
 - ▶ So **97% of time goes to jobs >2 nodes.**
 - ▶ Average jobs size about **6x that of Stampede2** – this machine **is** used differently.
- ▶ We tune the scheduling policy multiple times a year... essentially adjusting to demand.

LONESTAR 6

- ▶ Research computing for Univ of Texas System and other contributing partners
- ▶ Lonestar 6 configuration
 - ▶ 560 node, AMD EPYC 64-core, Dual-socket
 - ▶ (Most) nodes immersed in oil
 - ▶ GPU partition with 240 A100 GPUs



224 Icelake Nodes

(from Stampede 2)

- **2 Intel Xeon Platinum 8380 chips**
- **2x 40 core 2.3 GHz Xeon cores**
- **256 GB DRAM**

1060 Skylake Nodes

(from Stampede 2)

- **2 Intel Xeon Platinum 8160 chips**
- **2x 24 core 2.2 GHz Xeon cores**
- **192 GB DRAM**

Stampede3

- **10 PB VAST Filesystem**
- **400 Gb/sec Omni-Path Express**

560 Sapphire Rapids HBM Nodes

Each node contains:

- **2 Intel Max 9480 chips**
- **2x 56 1.9 GHz cores**
- **2x 64 GB HBM2e**

20 Ponte Vecchio Nodes

Each node contains

- **4 Intel Data Center GPU Max GPUs**
- **128 X^e cores**
- **128 GB HBM2e**

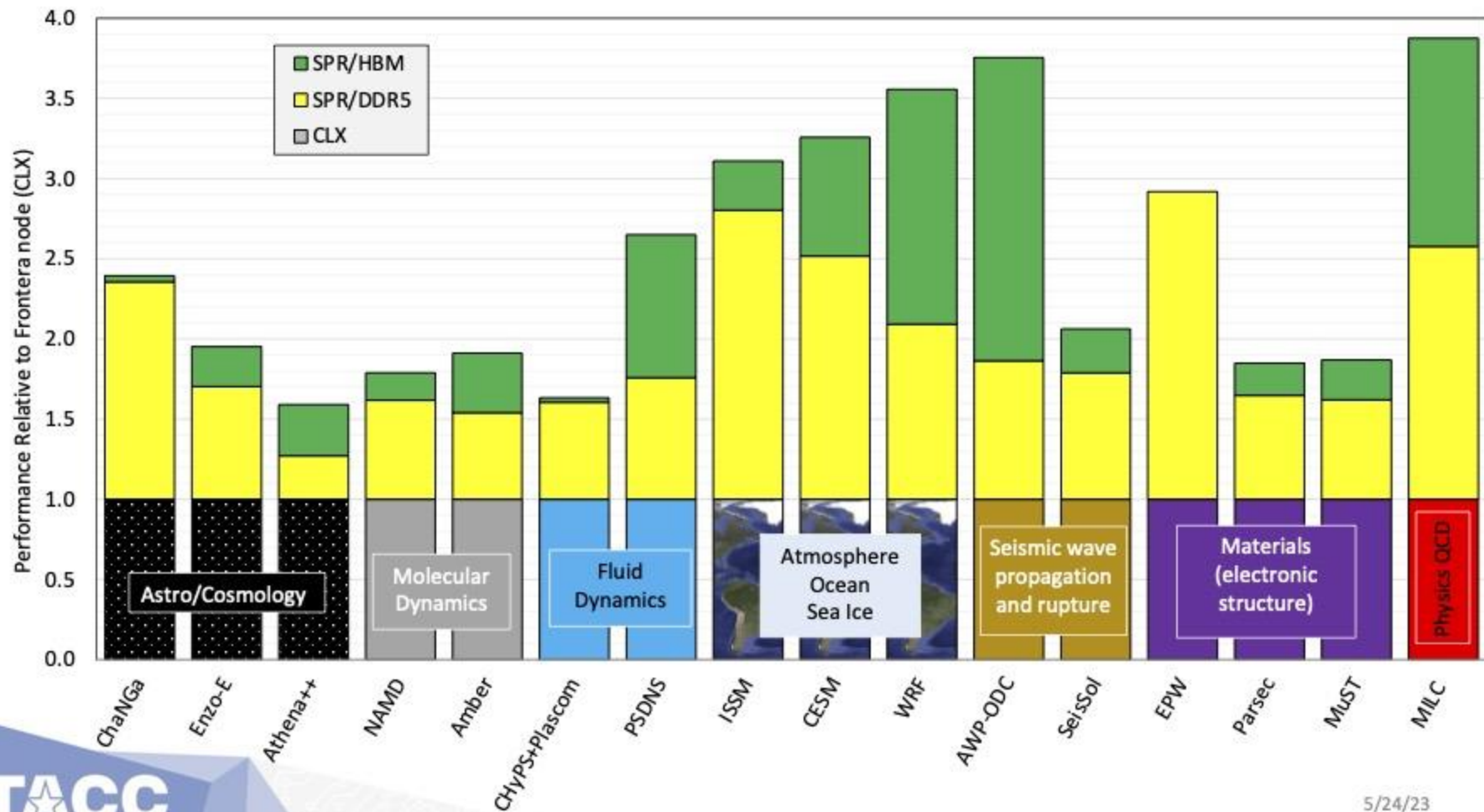
100Gb/sec Intel Omni-Path

100Gb/sec
Intel Omni-
Path

100Gb/sec Intel
Omni-Path

400Gb/sec
Cornelis

Single-Node Application Performance Ratios



COMPARISON

Sapphire rapids with HBM vs Granite Rapids with MRDDR

- ▶ Massive increase of core count
- ▶ MRDDR provides enough bandwidth for increased core count
- ▶ HBM bandwidth likely 'overprovisioned' for Sapphire Rapids

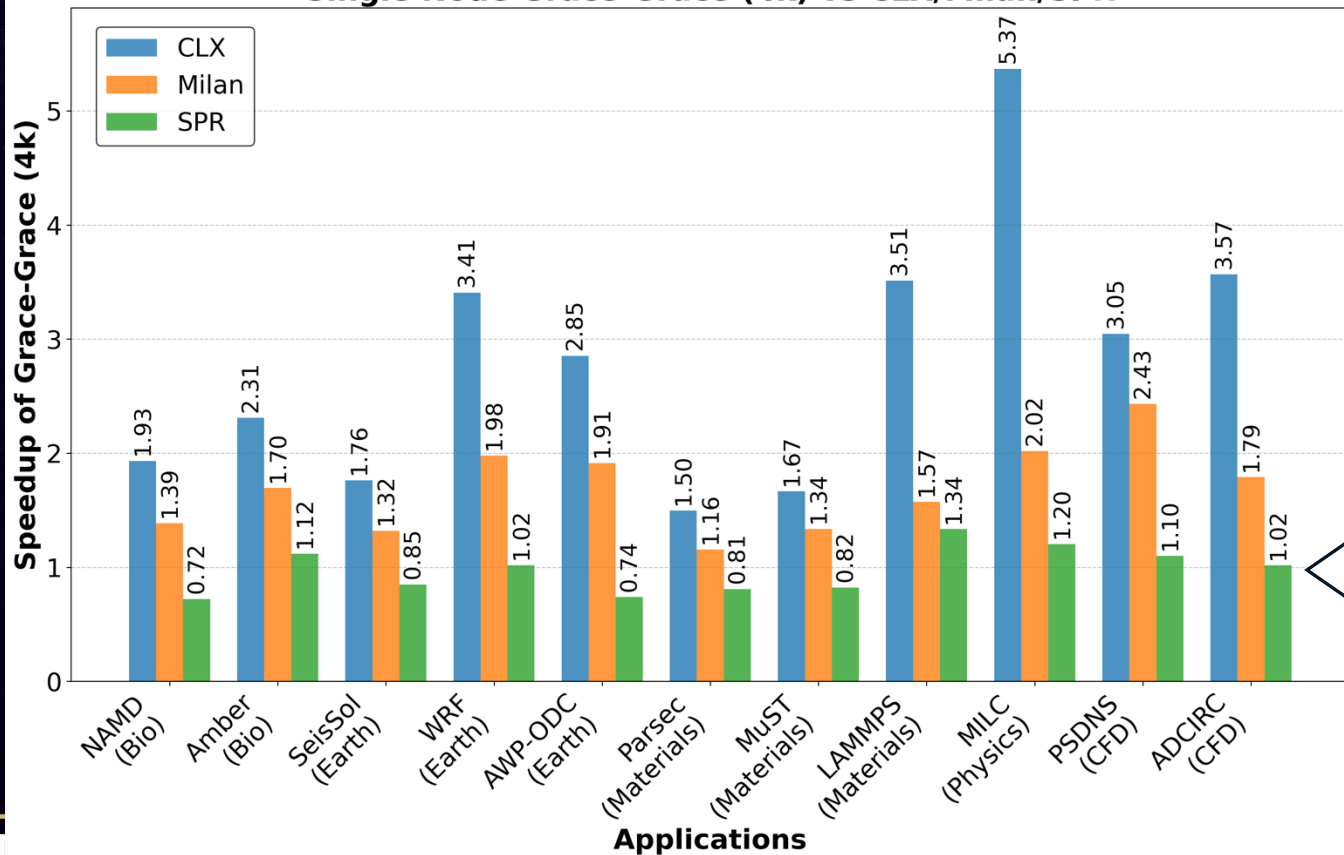
Single node speedup

WRF	1.89
Amber	1.28
MILC	2.13
NAMD	1.90
PSDNS	1.20

x86 is going strong!

COMPARISON x86 vs Grace (single node, dual-socket)

Single Node Grace-Grace (4k) vs CLX/Milan/SPR



Grace performance somewhere between Icelake and Sapphire Rapids

Performance normalized to Grace = 1
Below 1: x86 performs better



FRONTERA

TACC



TEXAS

lars@tacc.utexas.edu