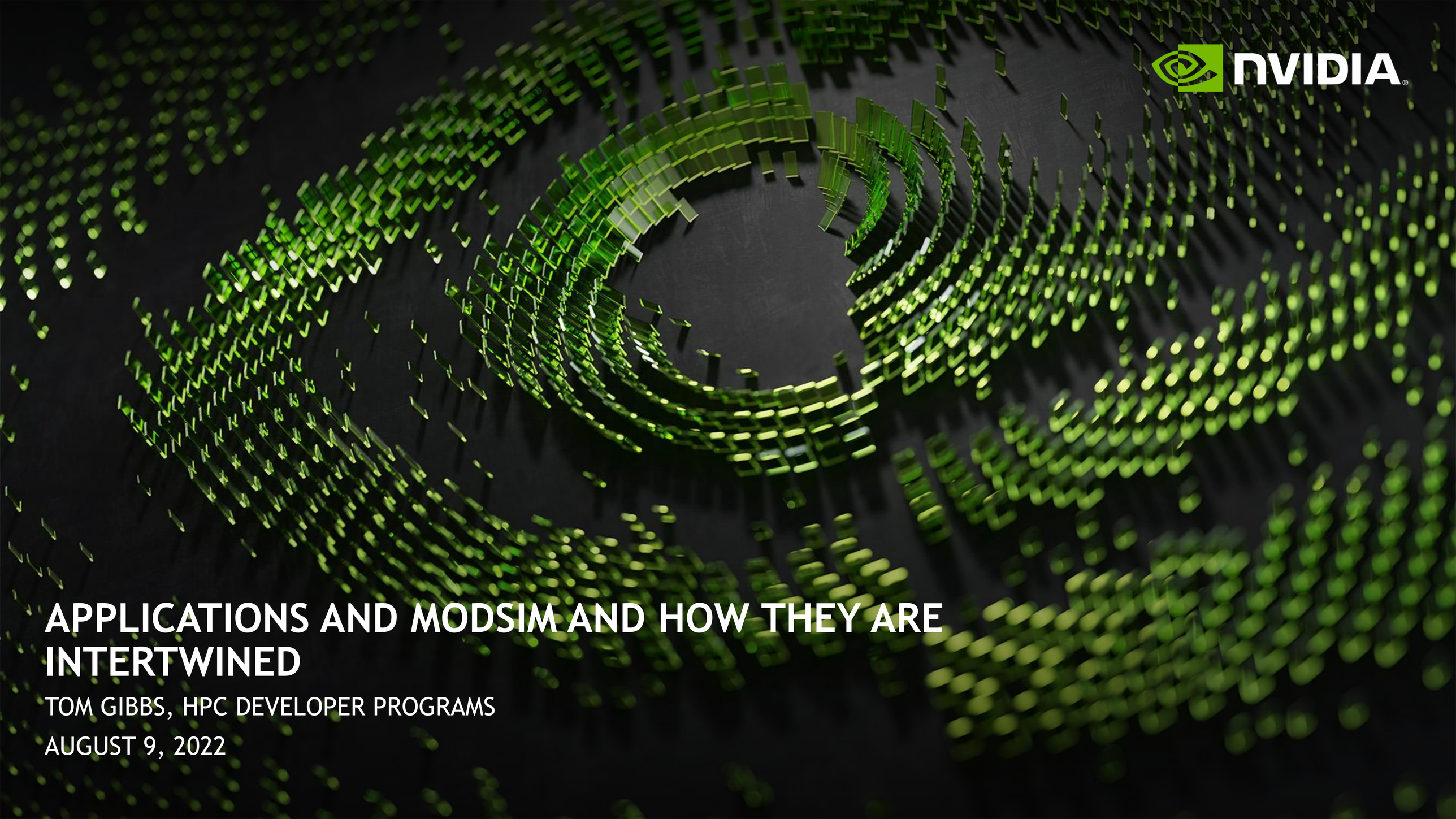
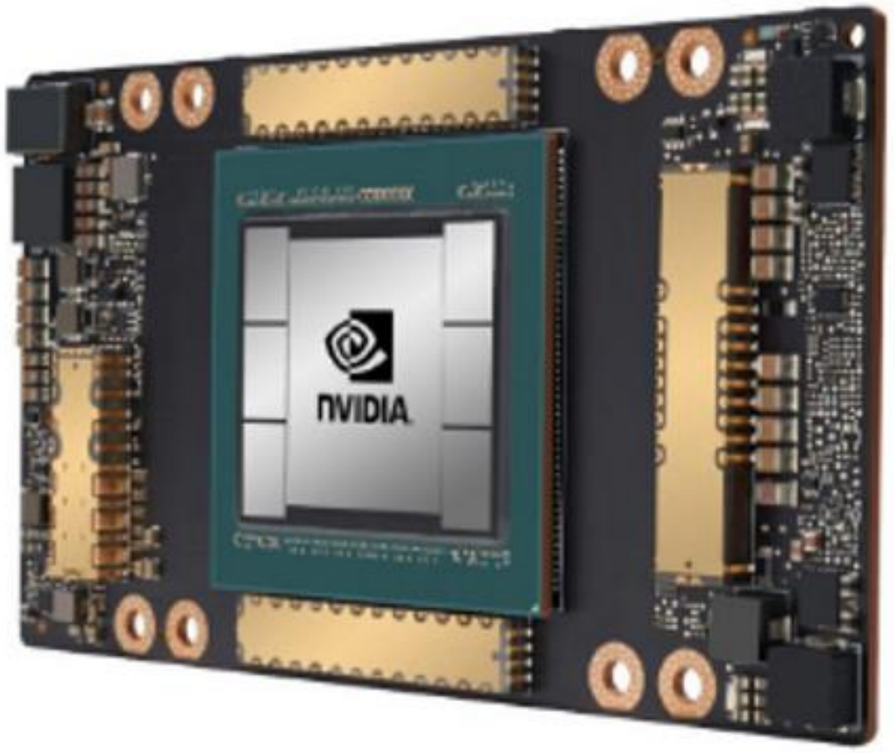


APPLICATIONS AND MODSIM AND HOW THEY ARE INTERTWINED

TOM GIBBS, HPC DEVELOPER PROGRAMS

AUGUST 9, 2022

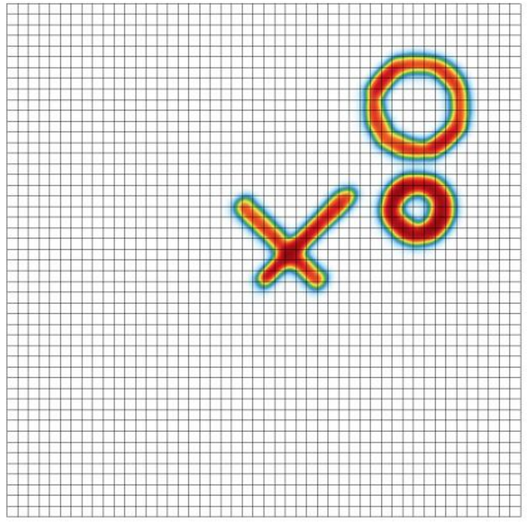
MODSIM WAS KEY TO NVIDIA THEN AND NOW

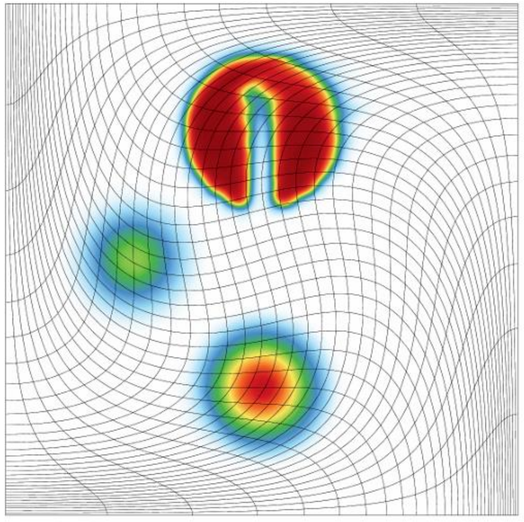


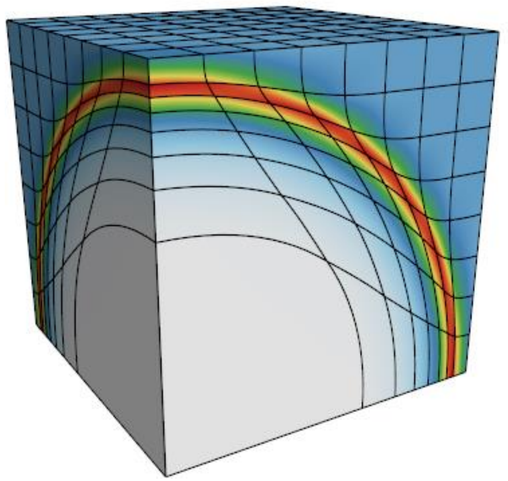
“But the reason I'm talking about it in this episode is that DNA comes from the fact that in order to survive when they had nine months left, the way that they saved themselves was with simulation. It became very clear to the company very early on, the benefits of being able to simulate something rather than having to do it in the real world.”

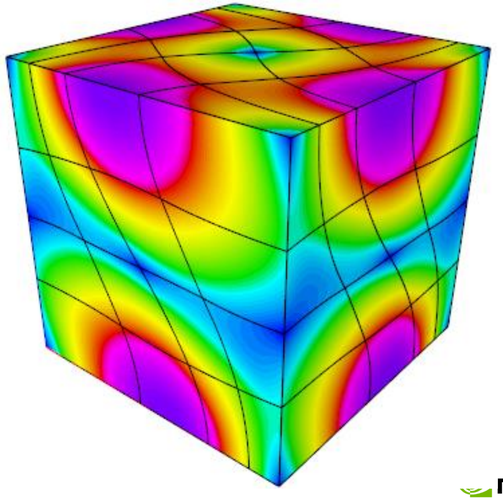
MODSIM HAS COME A LONG WAY

- Today there are more applications to evaluate and they are more complex
- Still key to HW innovation and debugging silicon
- However, simulation of large applications is hard and time consuming so kernels and proxies often used

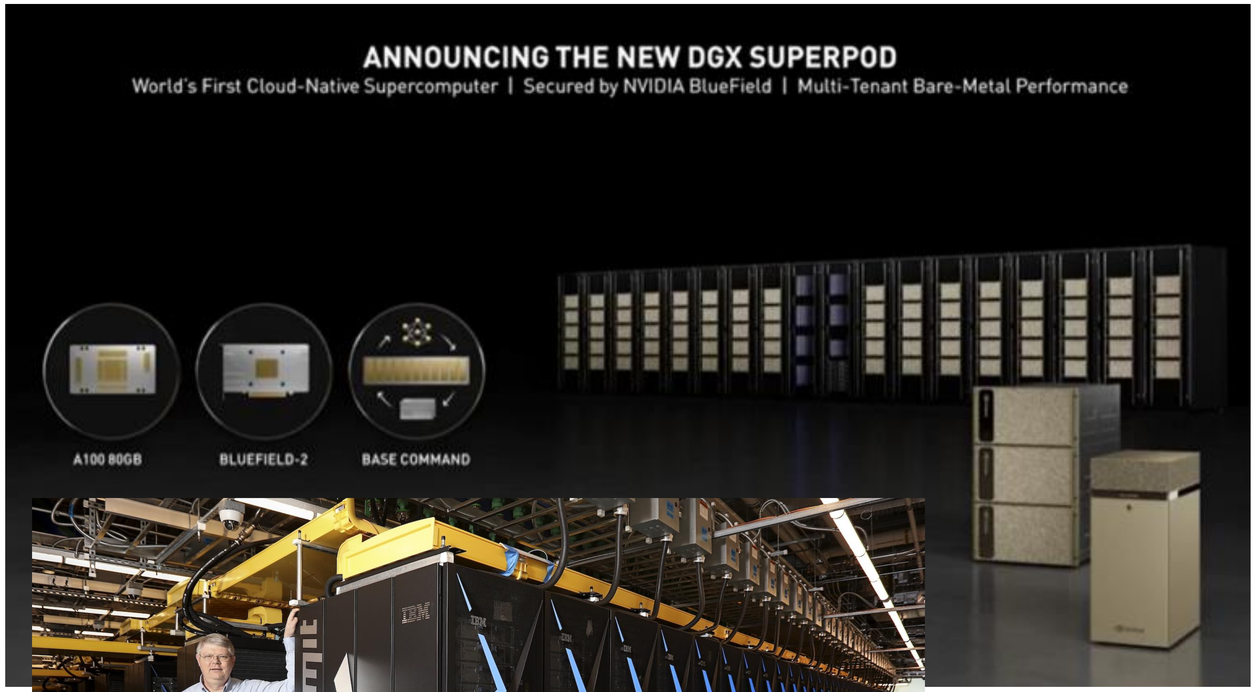


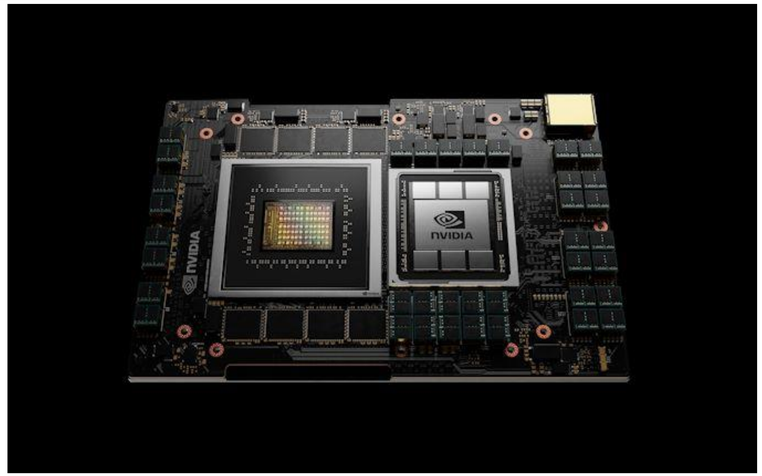


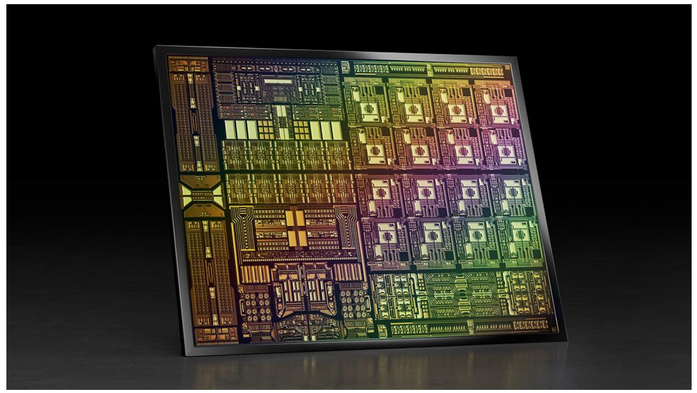




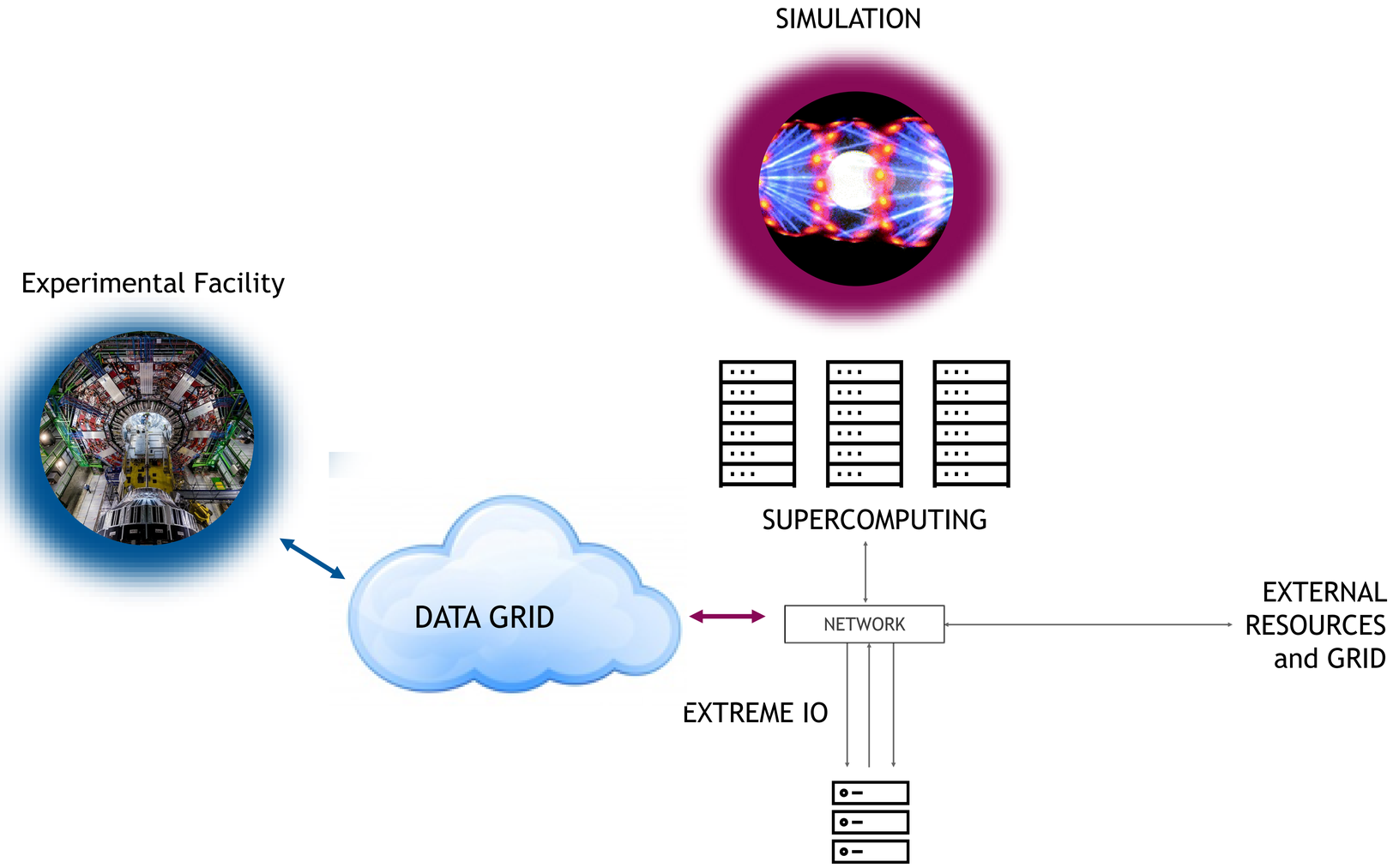
AND WE NEED MODSIM TO CONTINUE TO EVOLVE

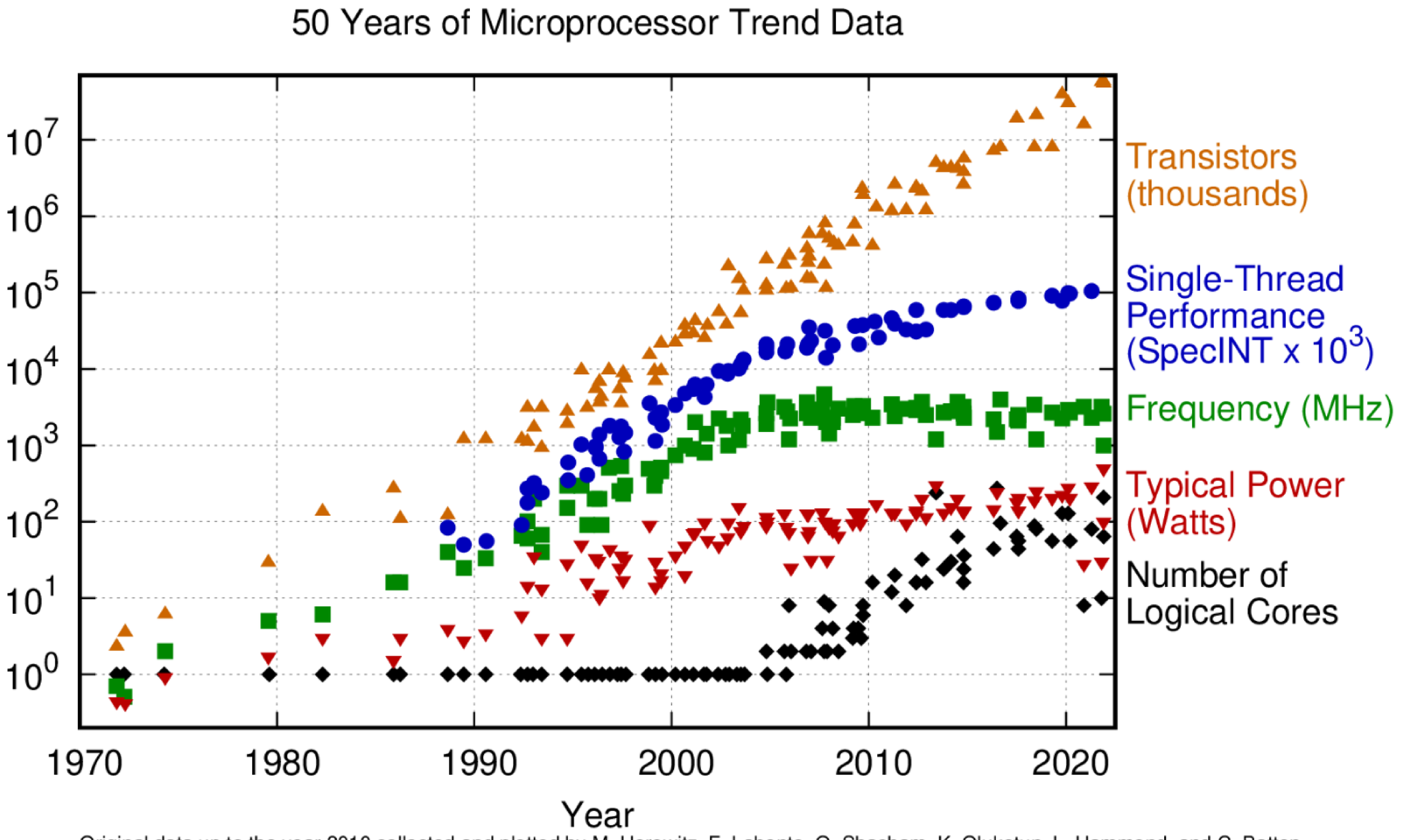




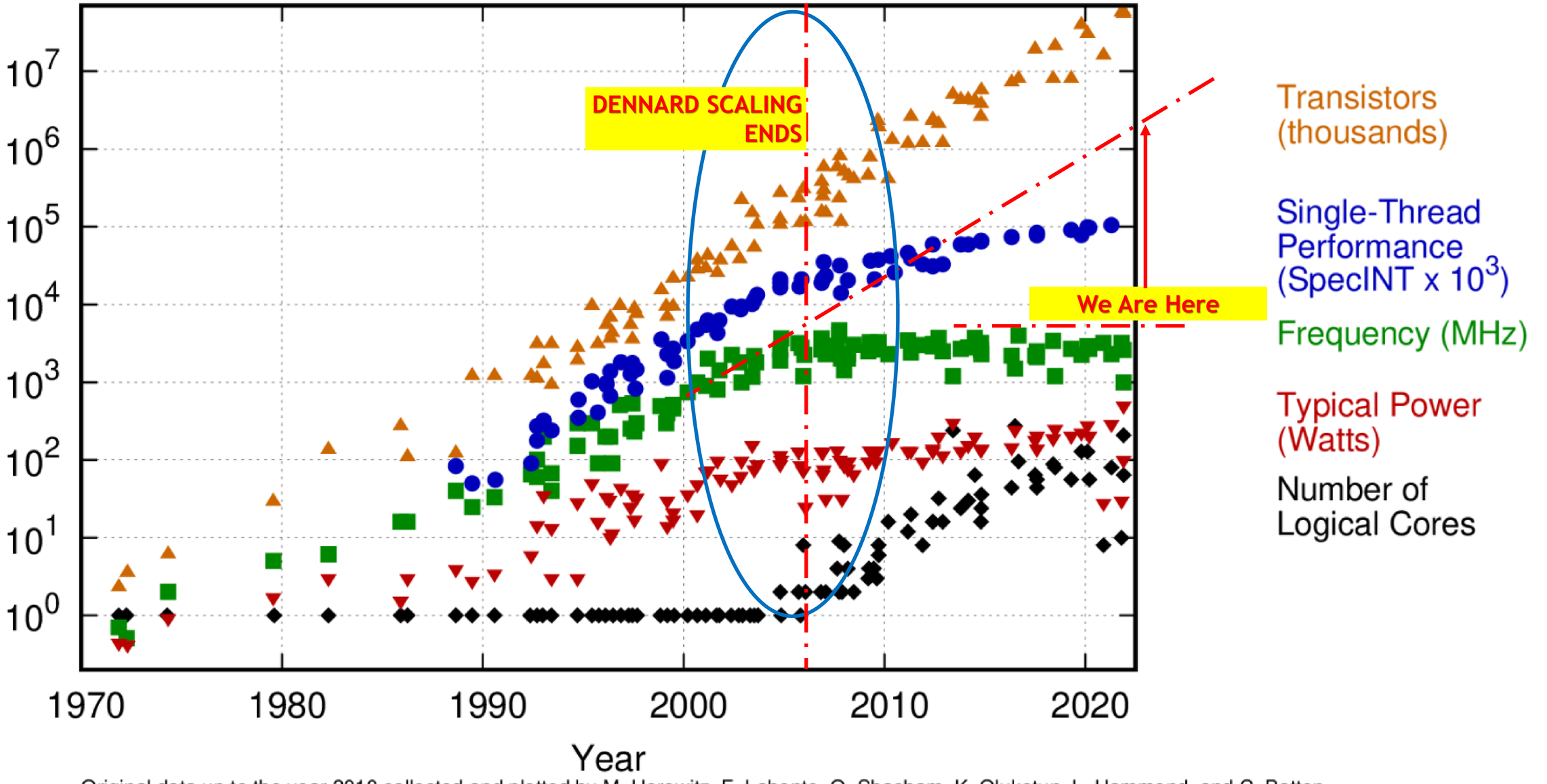


HISTORIC HPC ECOSYSTEM

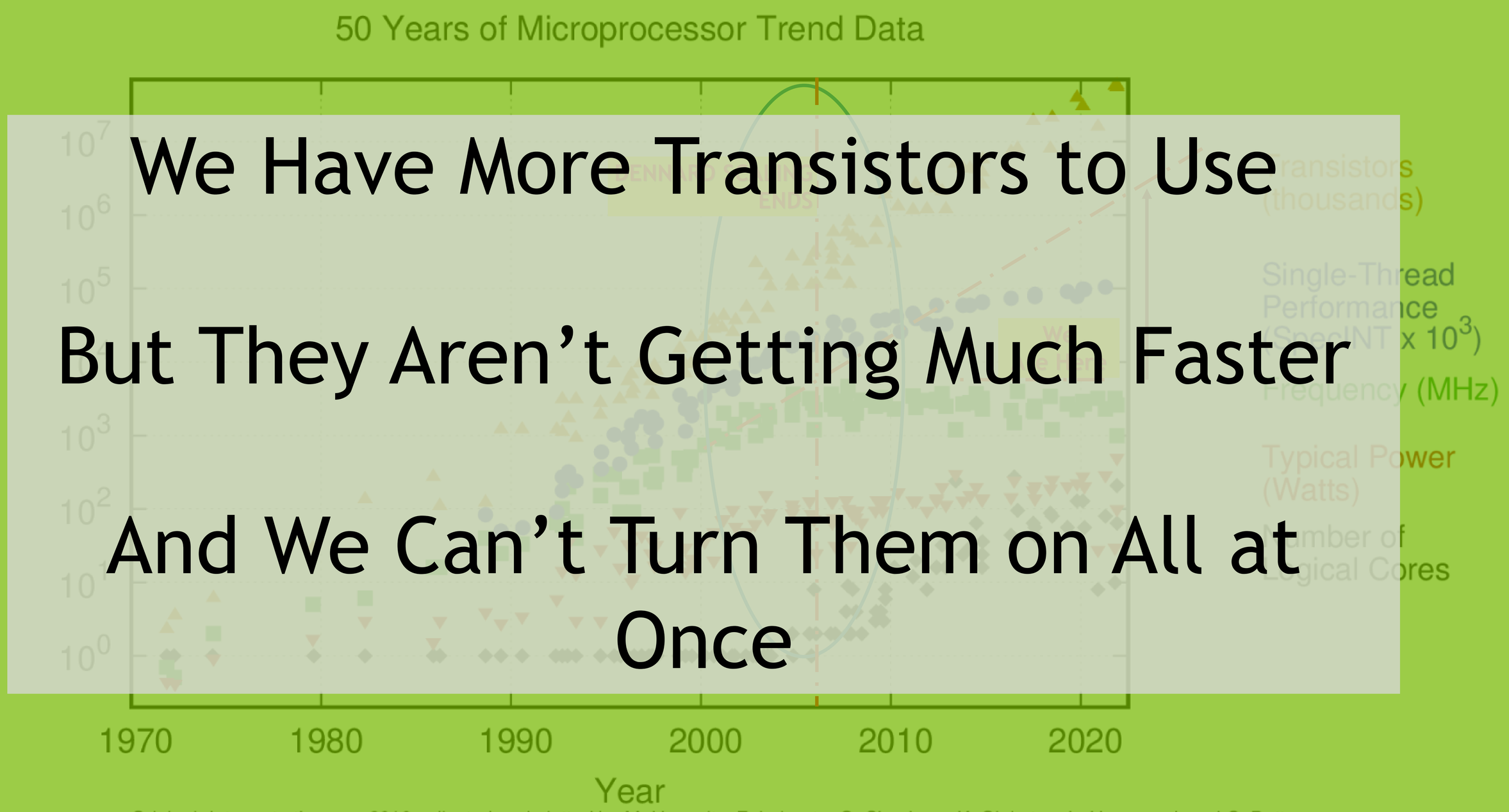



50 Years of Microprocessor Trend Data
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
1970
1980
1990
2000
2010
2020
Year
Transistors (thousands)
Single-Thread Performance (SpecINT x 10^3)
Frequency (MHz)
Typical Power (Watts)
Number of Logical Cores
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

50 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp



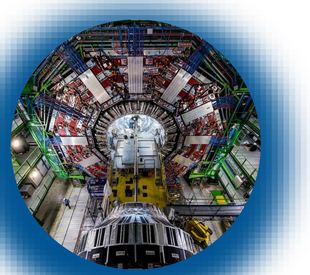
We Have More Transistors to Use

But They Aren't Getting Much Faster

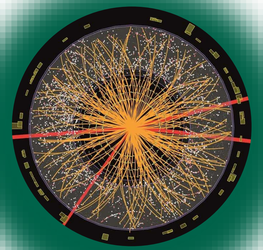
And We Can't Turn Them on All at Once

EXPANDING HPC ECOSYSTEM

EDGE

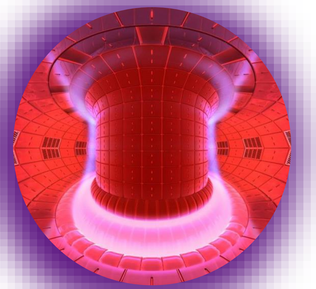


HPC * AI

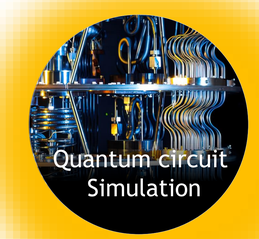


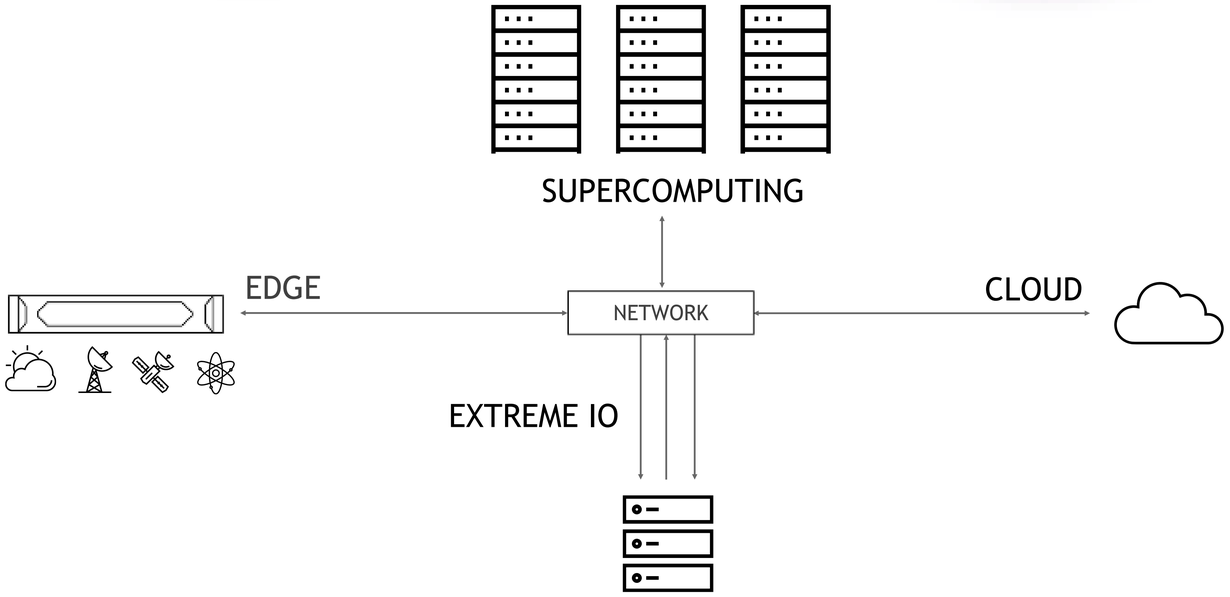
SIMULATION

DIGITAL TWIN



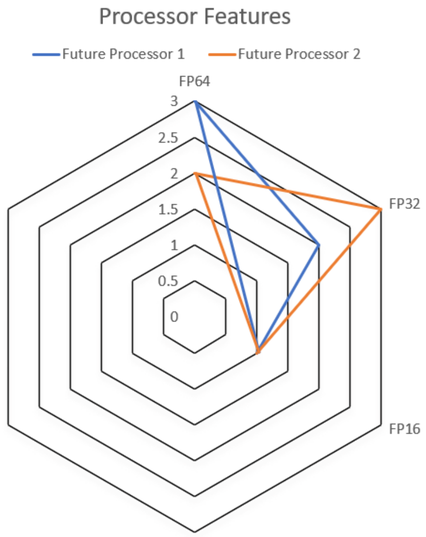
QUANTUM COMPUTING

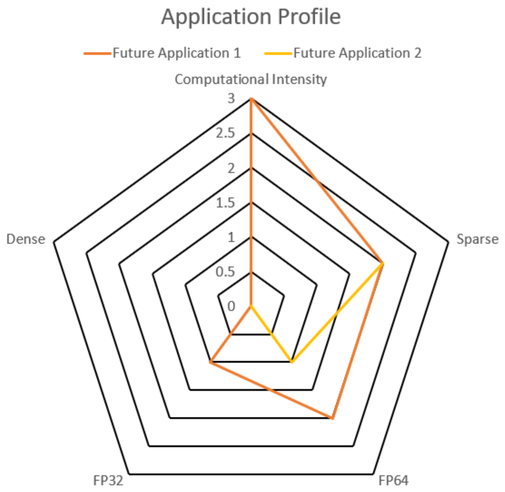




THE CHALLENGES WE ARE ENCOUNTERING

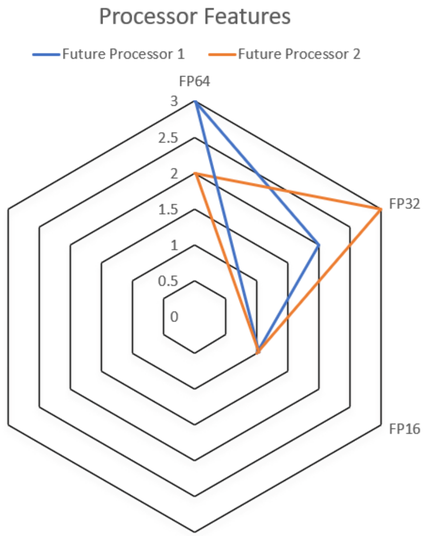
Algorithms, Applications, Workflows, Processor, Node and System Architectures are Evolving at a Rapid Pace

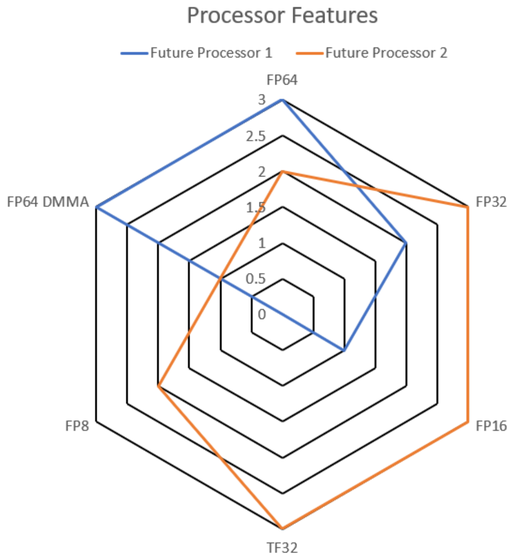


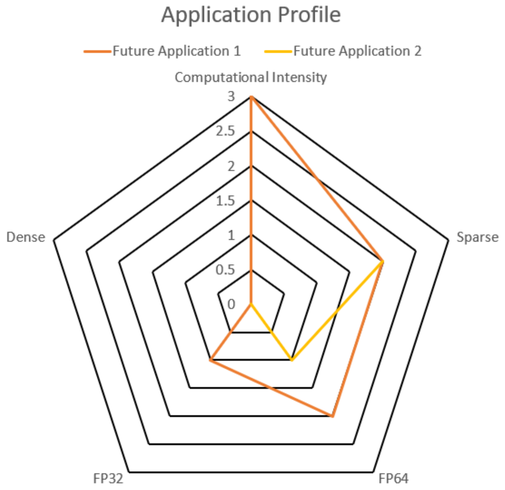


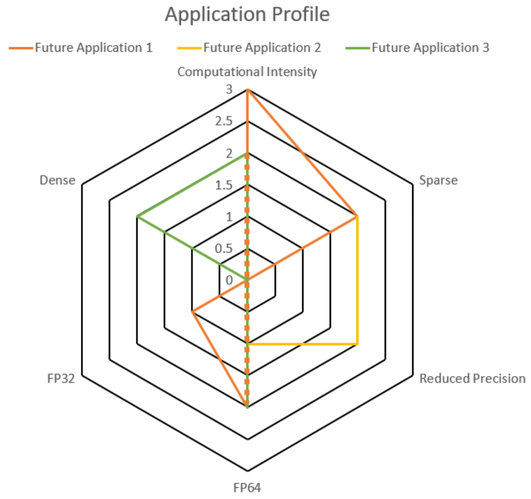
THE CHALLENGES WE ARE ENCOUNTERING

Algorithms, Applications, Workflows, Processor, Node and System Architectures are Evolving at a Rapid Pace



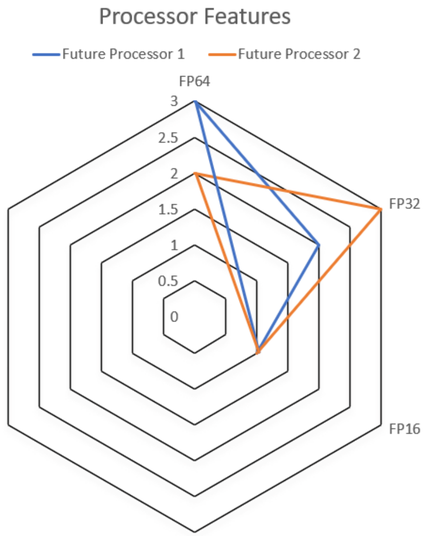


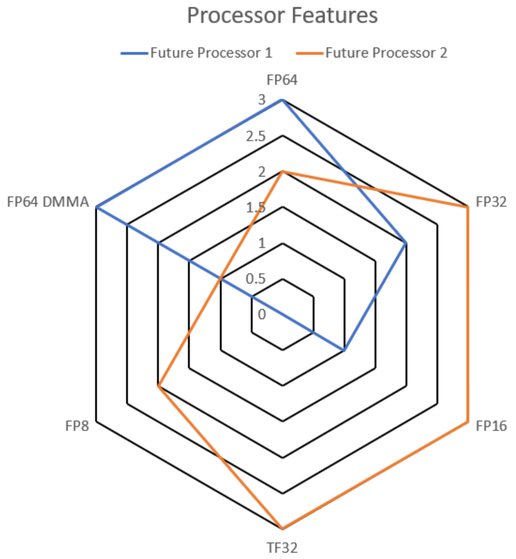


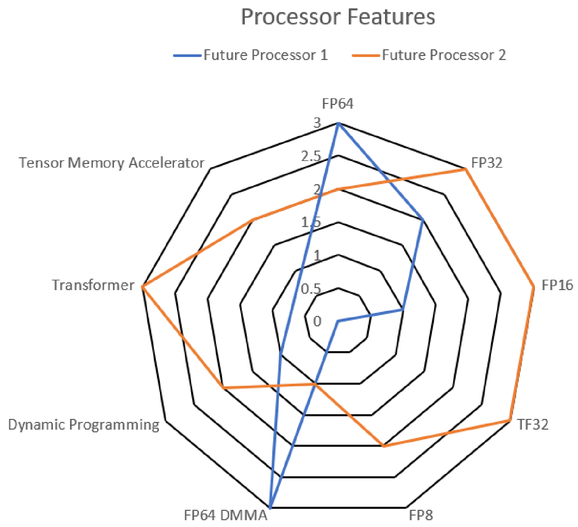


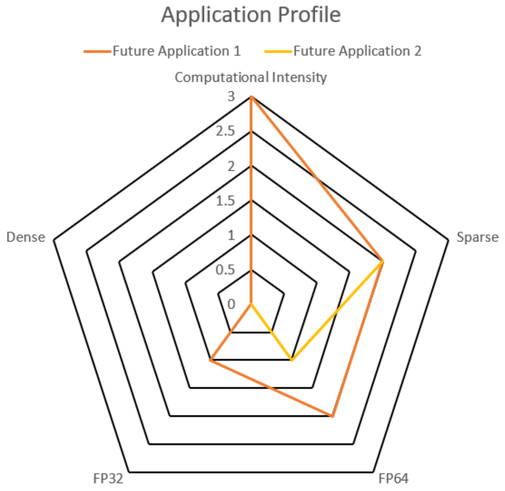
THE CHALLENGES WE ARE ENCOUNTERING

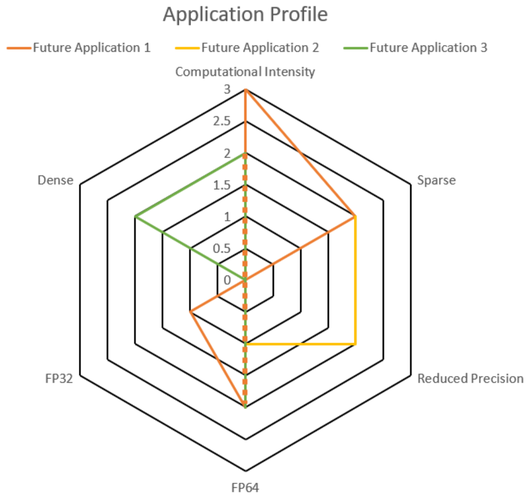
Algorithms, Applications, Workflows, Processor, Node and System Architectures are Evolving at a Rapid Pace

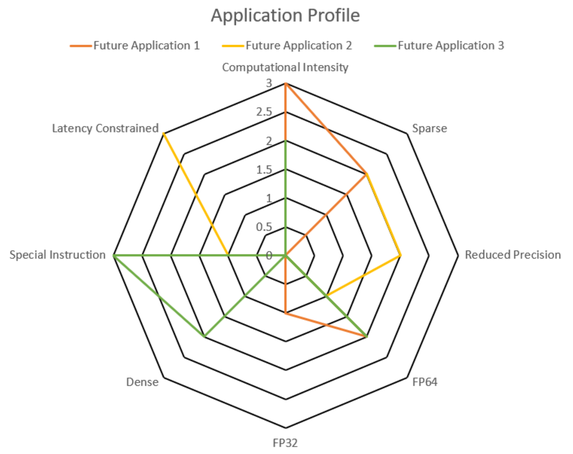






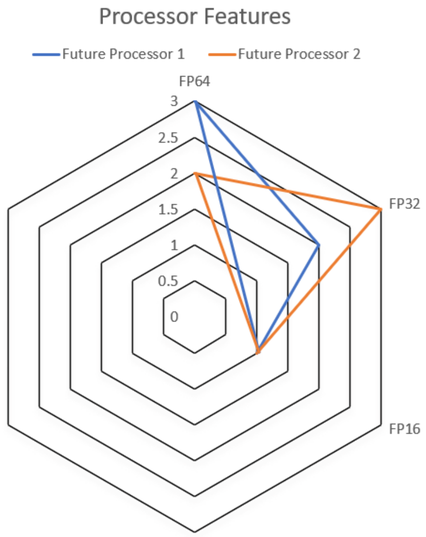


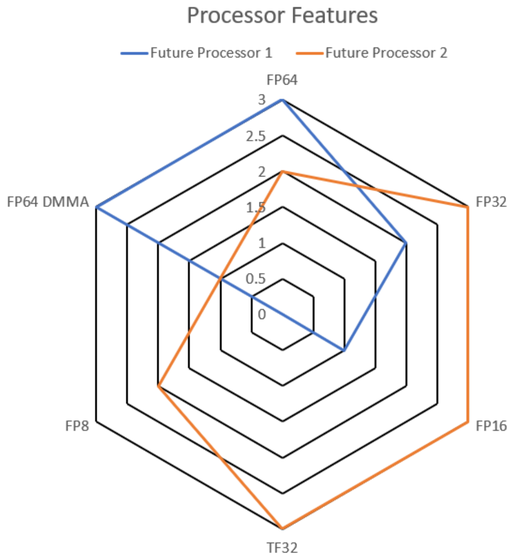


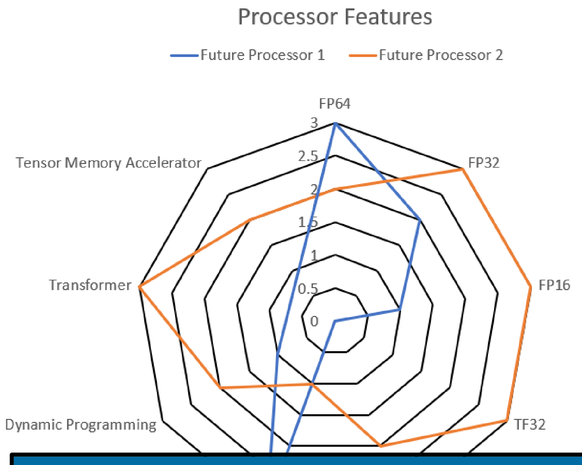


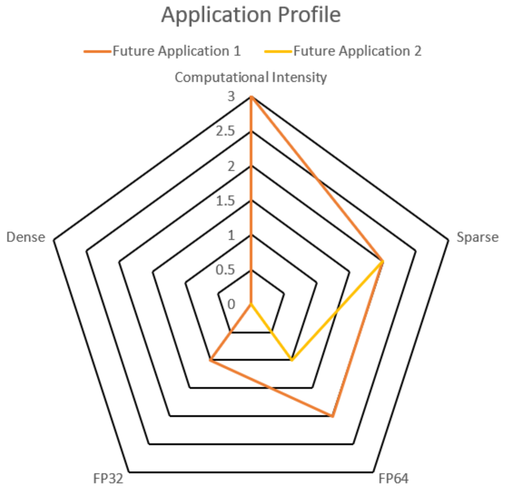
THE CHALLENGES WE ARE ENCOUNTERING

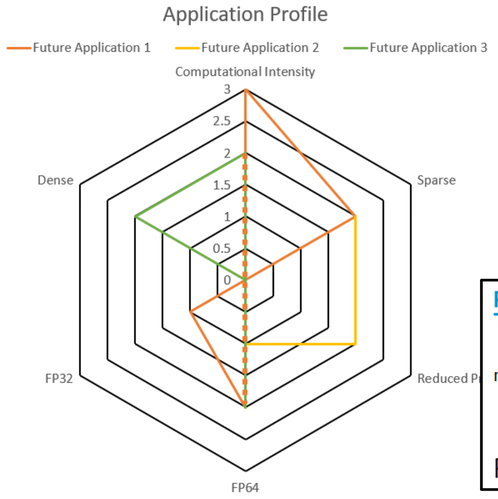
Algorithms, Applications, Workflows, Processor, Node and System Architectures are Evolving at a Rapid Pace











EDITORS: Kathryn Mohror, mohror1@llnl.gov
John M. Shalf, jshalf@lbl.gov

DEPARTMENT: LEADERSHIP COMPUTING

Preparing for the Future—Rethinking Proxy Applications

Satoshi Matsuoka, Jens Domke, Mohamed Wahib, and Aleksandr Drozd, *RIKEN Center for Computational Science, Kobe, 650-0047, Japan*

Andrew A. Chien and Raymond Bair, *Argonne National Laboratory, Lemont, IL, 60439, USA*

Jeffrey S. Vetter, *Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA*

John Shalf, *Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Numerical algorithms for high-performance computational science

Jack Dongarra[1,2,3], Laura Grigori[4] and Nicholas J. Higham[3]

Research

Check for updates

NVIDIA

IMPECCABLE WORKFLOW GORDON BELL 2020

SARS-CoV-2 target identification

Network-based methods for host pathogen interactions

PDB dataset

Candidate targets

AI-based loop modeling and structures

Active learning strategies for virtual screening (Rick Stevens)

Drug Databases
(ENAMIN, DRUGBANK, ZINC, etc.)

Featurization
(MOdred, graph/neural fingerprints)

Physics-based Docking

RL-Dock: AI approach

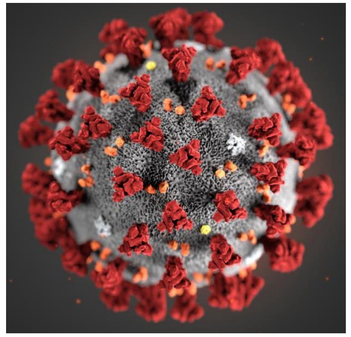
10,000,000,000 compounds screened with AI models
Top 2.5%
250,000,000 poses docked
Top 2.5%
6,250,000 systems build and minimized
Top. 2.5%
156,250 systems simulated
(that's about 12H on 1024 summit nodes)

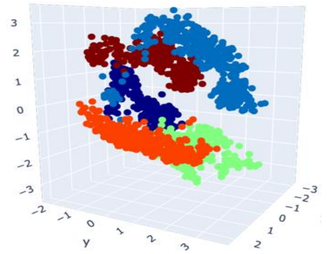
Physics based models

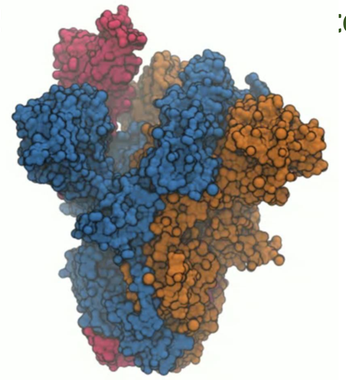
AI-based models

AI-based consensus ranking / scoring

Multi-scale, multi-resolution modeling of the S :eins



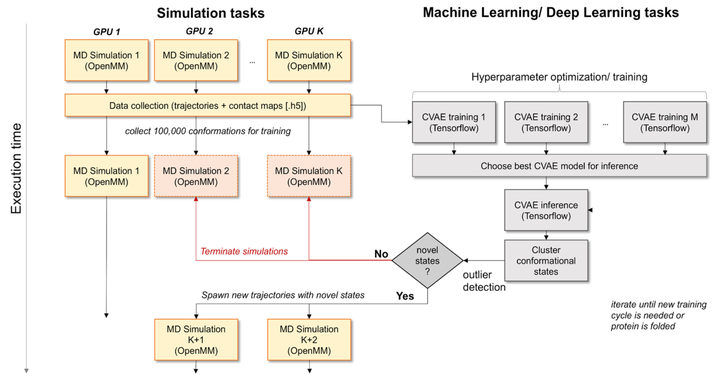




Rommie Amaro

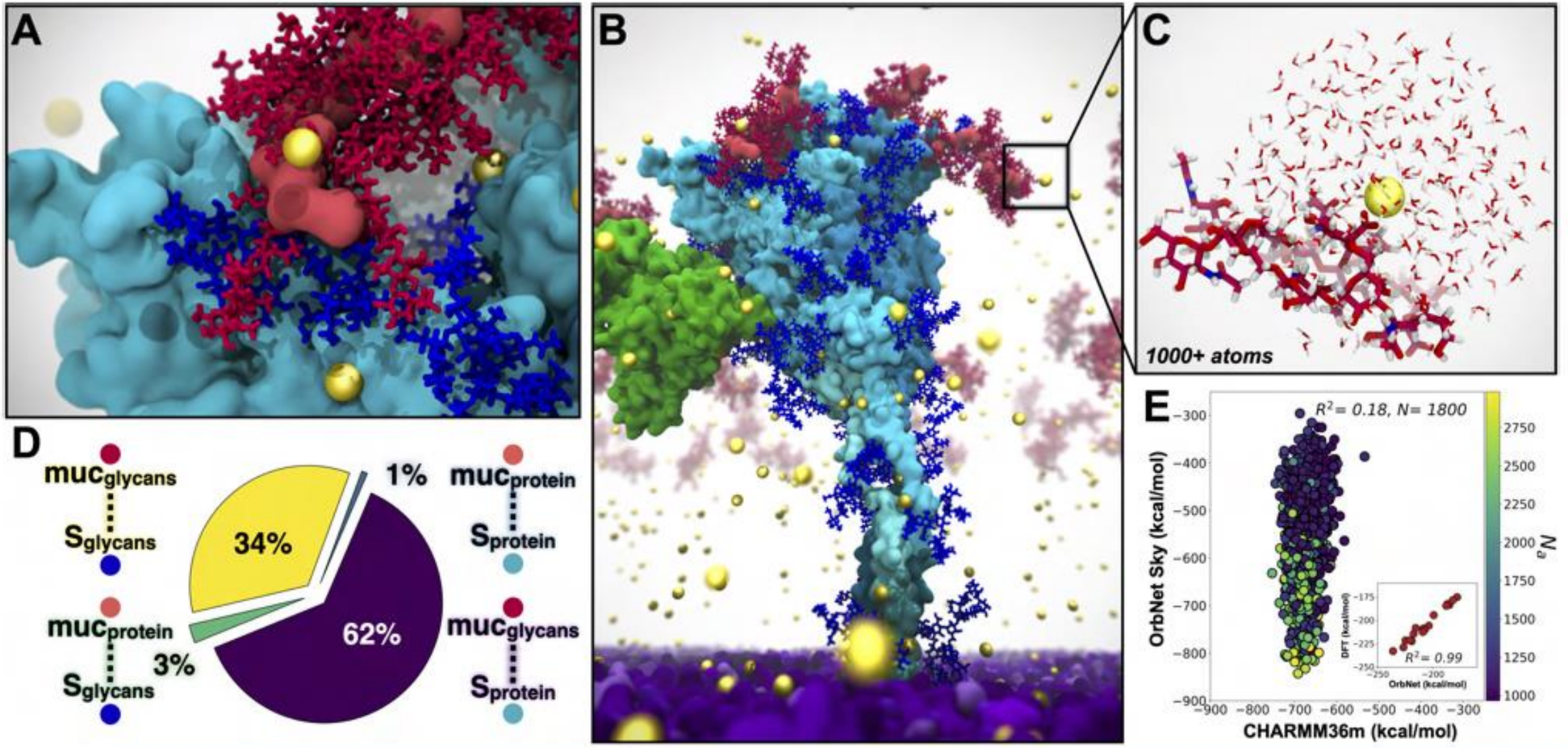
Carlos Simmerling

DeepDriveMD: AI-driven adaptive sampling/ simulations



(Arvind Ramanathan, Shantenu Jha)

2021 WORKFLOW EXTENDED TO MODEL VIRION IN AN AEROSOL WITH DFT ACCURACY



SMA system captured with multiscale modeling from classical MD to AI-enabled quantum mechanics. For all panels: S protein shown in cyan, S glycans in blue, m1/m2 shown in red, ALB in orange, Ca2+ in yellow spheres, viral membrane in purple.

A) Interactions between mucins and S facilitated by glycans and Ca2+.

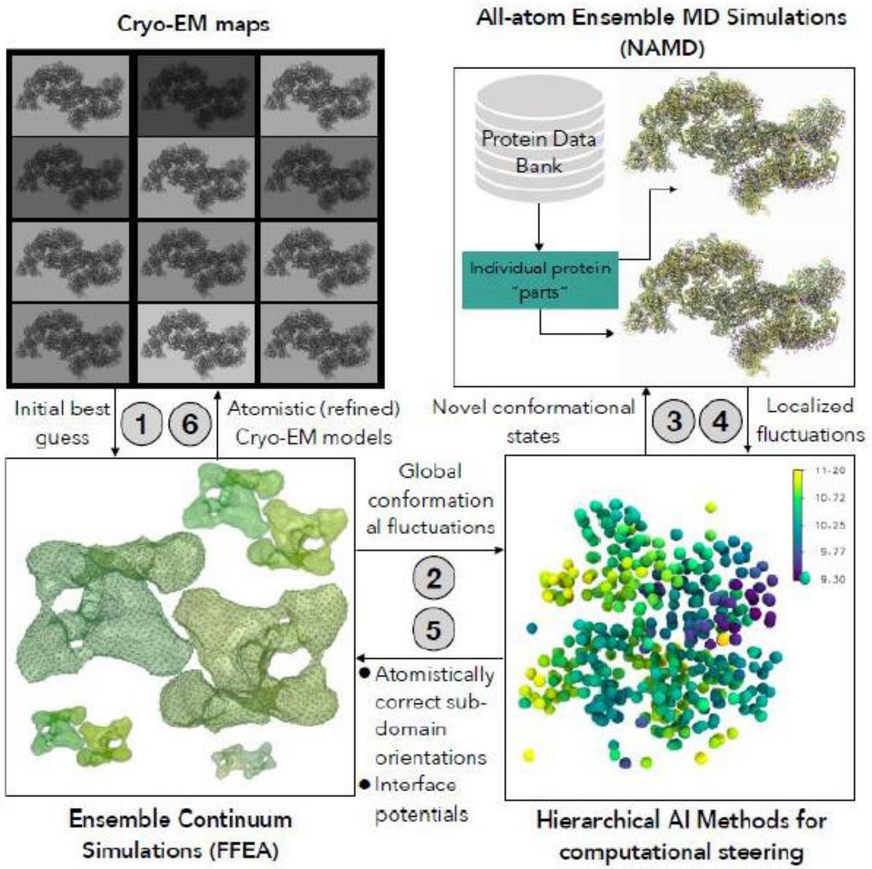
B) Snapshot from SMA simulations.

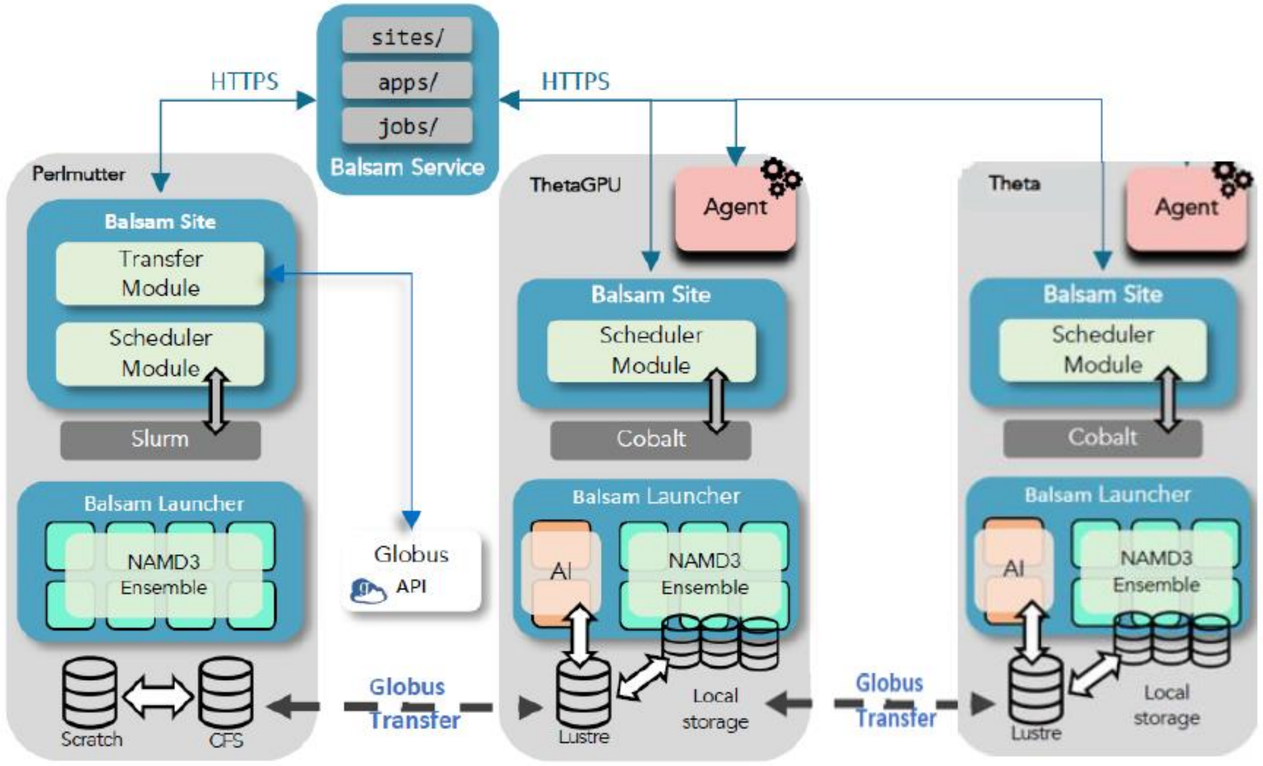
C) Example Ca2+ binding site from SMA simulations (1800 sites, each 1000+ atoms) used for AI-enabled quantum mechanical estimates from OrbNet Sky.

D) Quantification of contacts between S and mucin from SMA simulations.

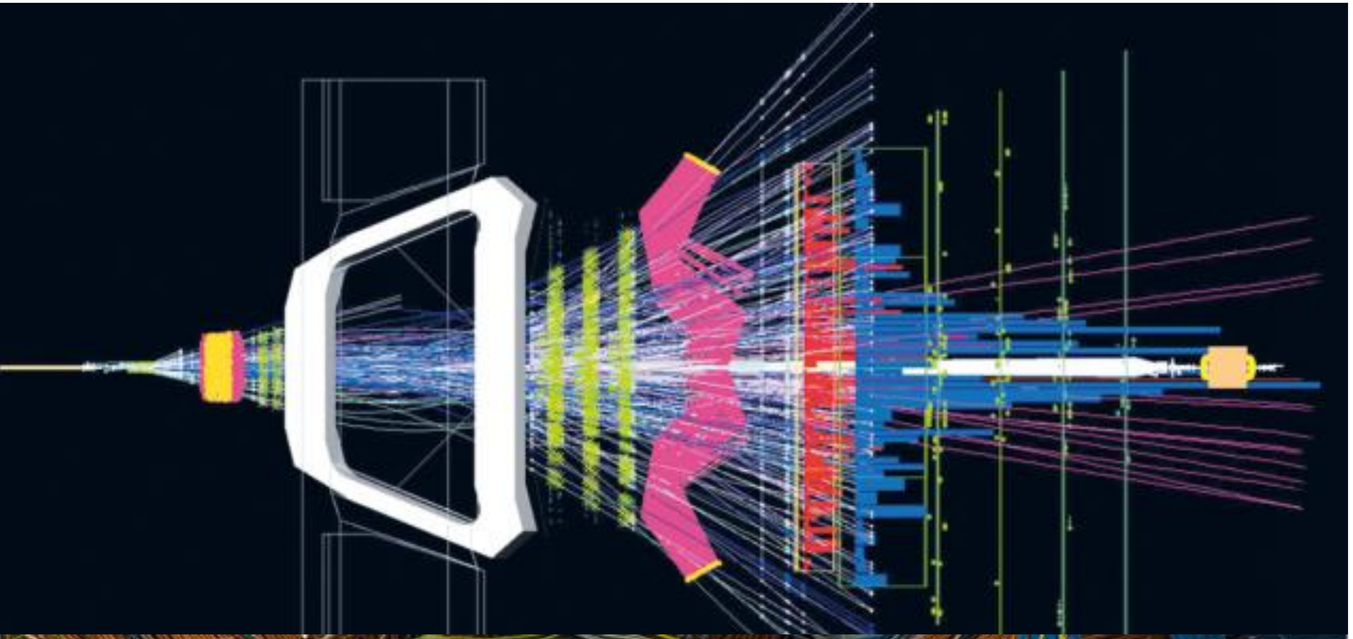
E) OrbNet Sky energies vs CHARMM36m energies for each sub-selected system, colored by total number of atoms. Performance of OrbNet Sky vs. DFT in subplot (ωB97x-D3/def-TZVP, R2=0.99, for 17 systems of peptides chelating Ca2+ (Hu et al., 2021)). Visualized with VMD.

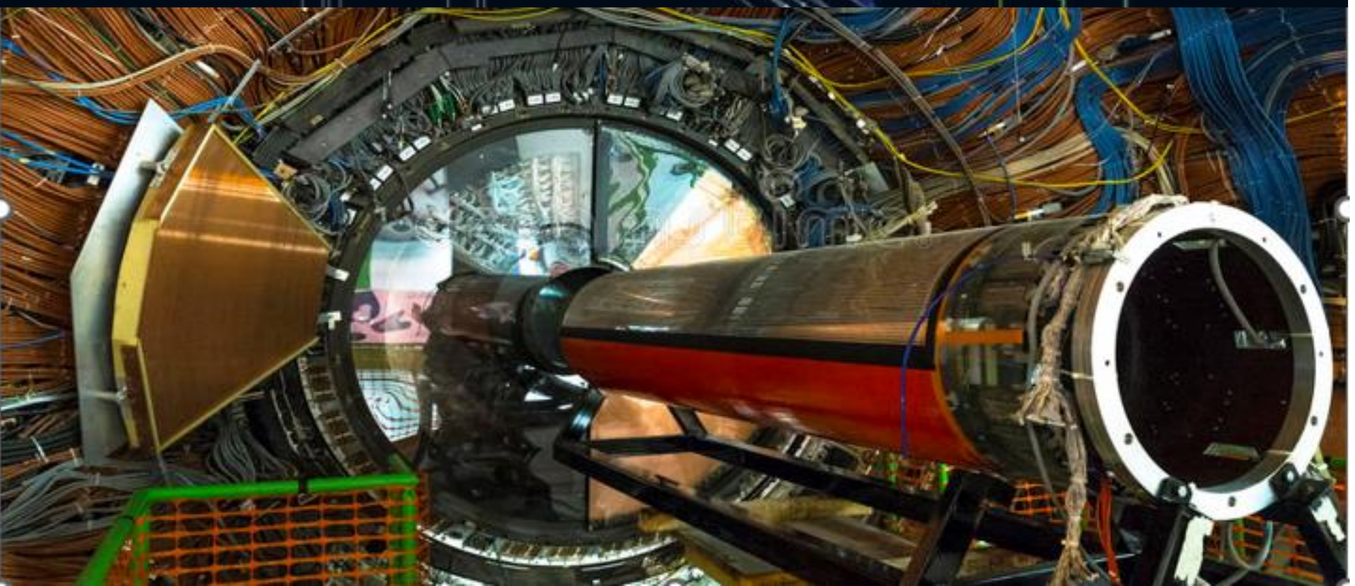
2021 WORKFLOW EXTENDED TO MODEL VIRION DETAIL FROM CRYOEM IMAGES





LHCB INTRODUCES ALLEN FOR REAL TIME TRIGGER AND TRACKING AT THE EDGE





Challenge
Apply conventional FFT and KALMAN Filter methods to perform trigger and tracking in the same process step for the LHCb Upgrade

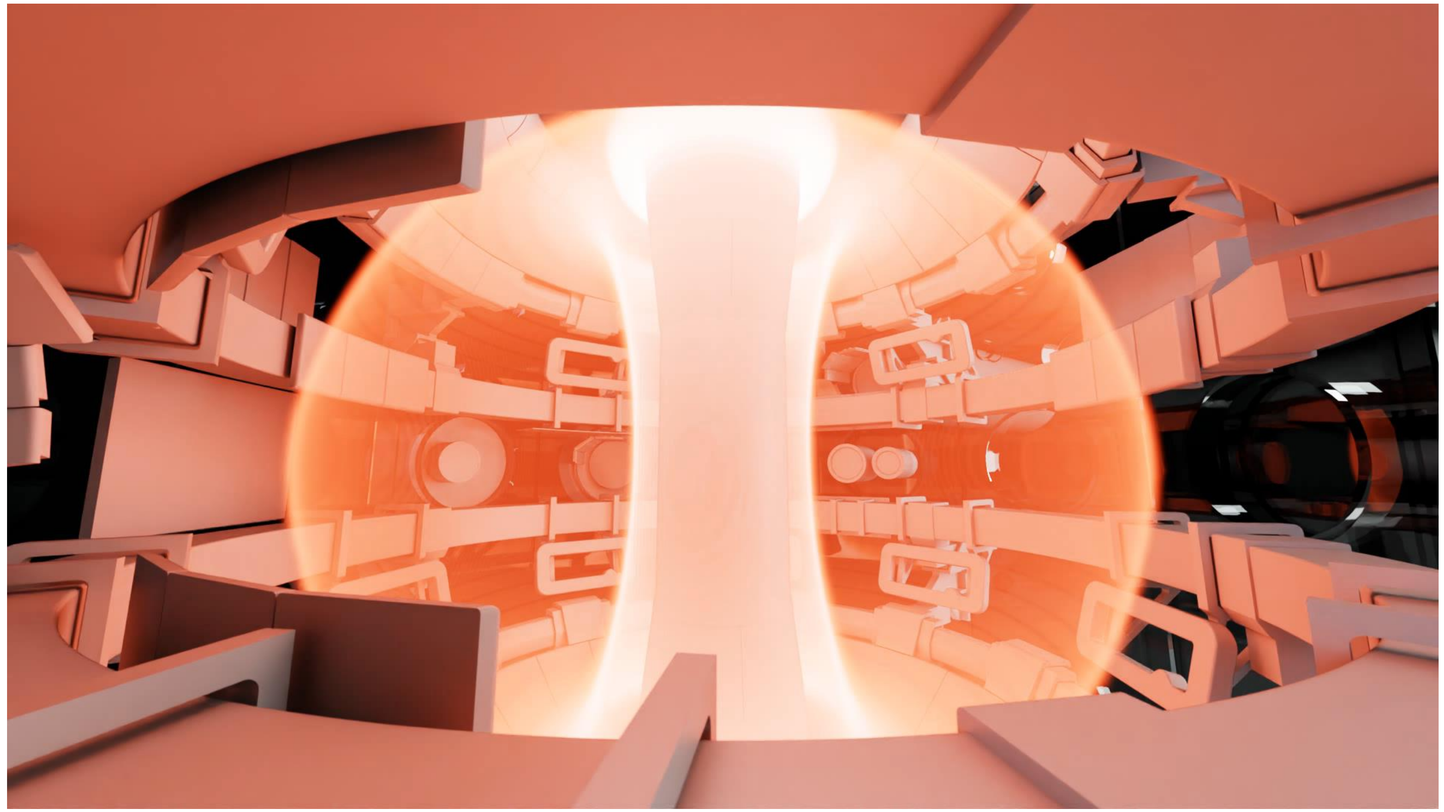
Solution
An optimized suite of algorithms was developed by CERN with support from NV to

Impact
Throughput with trigger and tracking of >60kHz was demonstrated

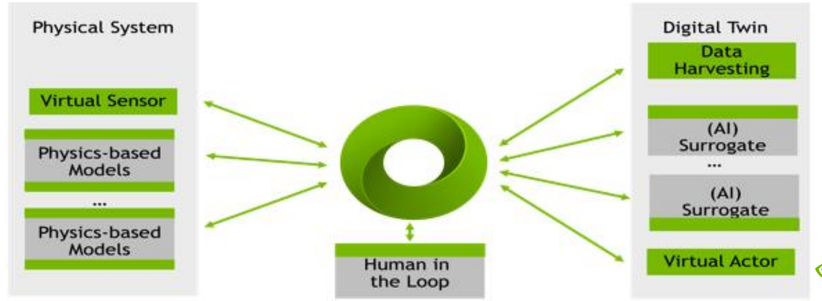
The full experiment can be supported with 500 GPUs

Run 3 commenced this quarter with 10x processing throughput

GYROKINETIC FUSION REQUIRES THE EXPANDED ECOSYSTEM AT THE EDGE



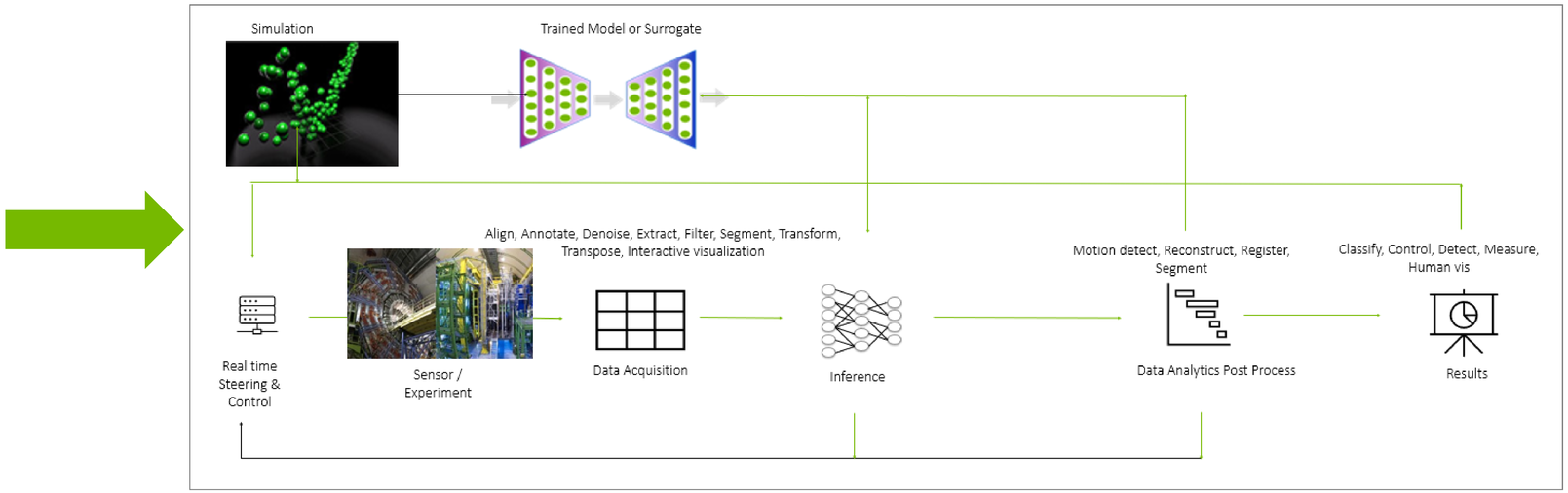
- Accelerated Simulations
  - GTC, XGC, GENE, CGYRO … 10x + unaccelerated
- Surrogate Models
  - SGTC six orders of magnitude faster than GTC
  - QualikNN 4 orders of magnitude faster than Qualikiz
- Control System Prediction for Disruption at the Edge
  - At DIII-D
  - FRNN 86% accuracy based on diagnostics from JET experiment with live testing at DIII-D underway
  - GatedRNN and Random Forest ML in control system now
  - At TCV
  - Reinforcement Learning applied with DeepMind and DIFFER
- Digital Twin
  - Early Demonstration with MAST Experiment at UK AEA

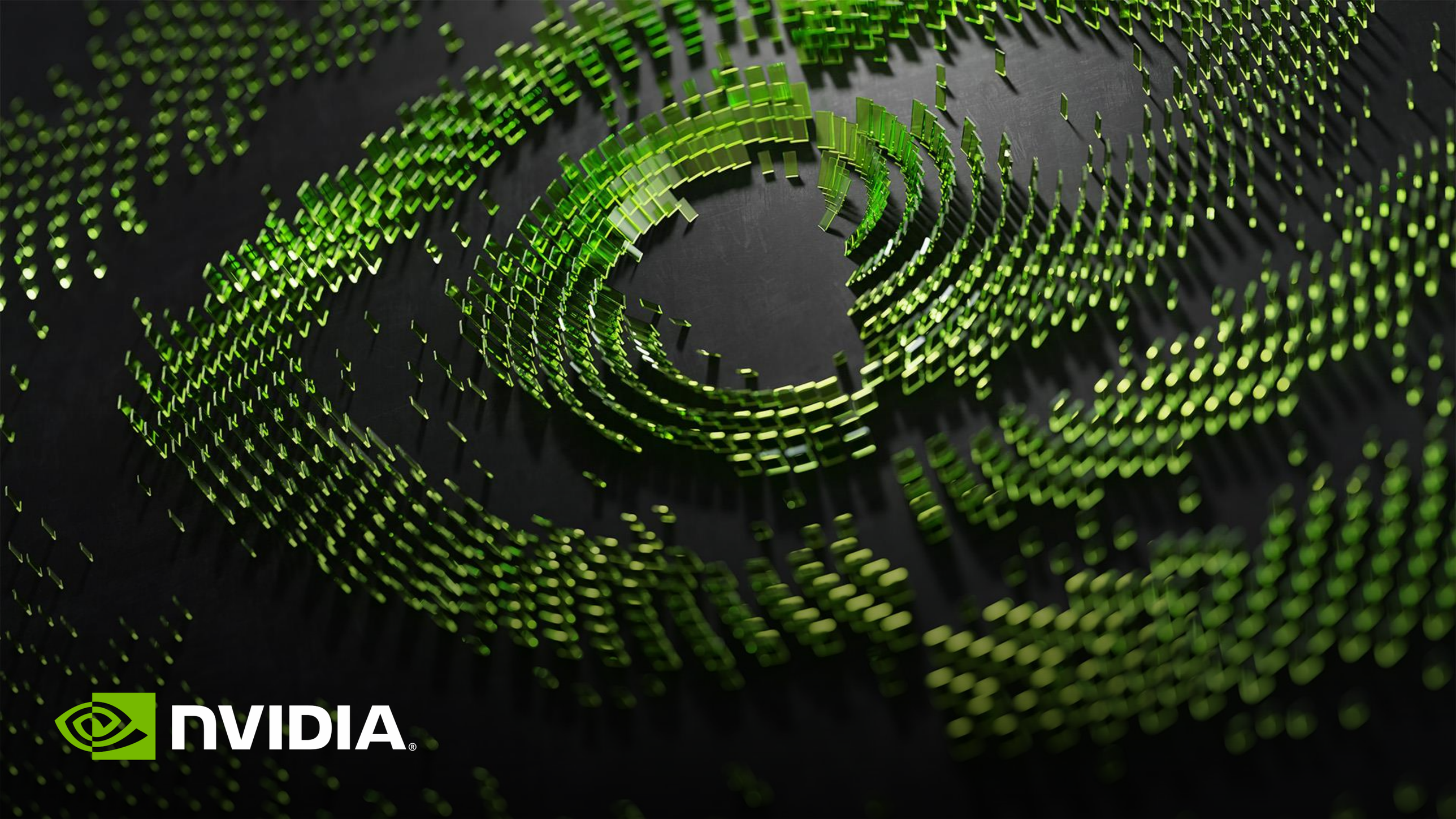


NVIDIA

FINAL THOUGHTS

- ModSim is needed now more than ever as transistor count continues to grow while frequency remains flat and power management becomes critical
- Algorithm diversity is increasing at a rate well beyond historical norms
- The workload evolving to workflows is introducing new “opportunities” for science discovery but also new bottlenecks as Amdahl’s Law is extended across all the workflow components
- Communication is becoming a more critical factor than ever within a processor, across processor components on a node, within a system, across subsystems and between facilities




NVIDIA