# HPC Strategy & US Exascale Program

James Amundson, Scientific Computing Division Head

Inaugural Meeting of the International Computing Advisory Committee

March 14, 2019

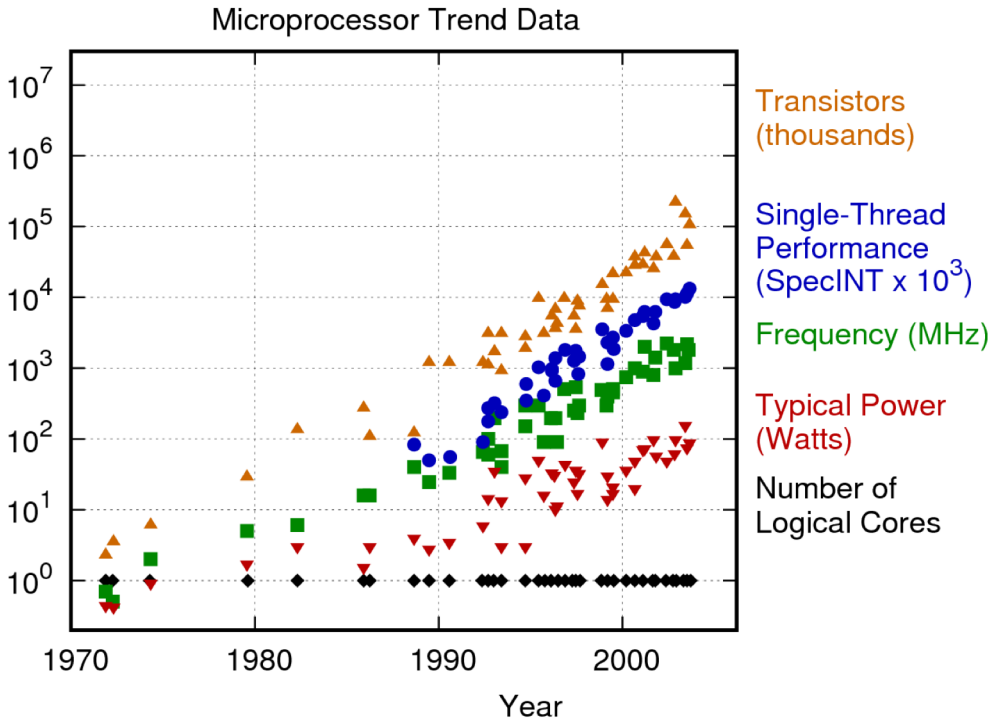
# My Perspective

- The vast majority of HEP computing to date has been high throughput computing (HTC).
- My personal *technical* experience includes:
  - particle theory (ancient history)
  - Tevatron Run-II computing
  - CMS grid computing (through 2001)
  - particle accelerator simulation (after 2001)
    - primarily for HPC
- I have observed the gap between the HEP and HPC communities from a unique vantage point

# “Moore’s Law” – the good old days

https://www.karlrupp.net/2018/02/42-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 2010-2017 by K. Rupp

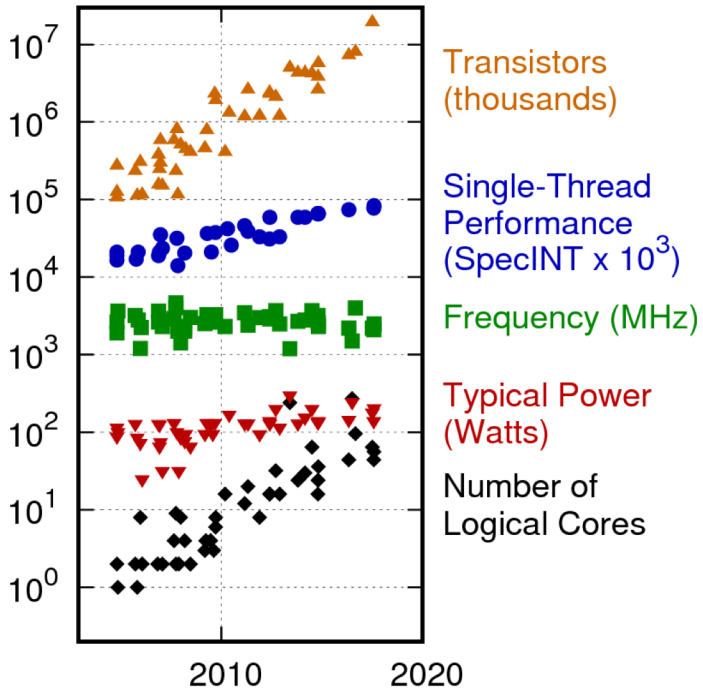





Fermilab

# “Moore’s Law” – recent times

Trends have changed






Fermilab

# Computing is changing

- Architectures are changing
  - Driven by solid state physics of CPUs
    - Multi-core
    - Limited power/core
    - Limited memory/core
    - *Memory bandwidth increasingly limiting*
    - *GPUs are winning the day when they can be used*
- High Performance Computing (HPC, aka Supercomputers) are becoming increasingly important for HEP
  - 2000s: HPC meant Linux boxes + low-latency networking
    - No advantage for experimental HEP
      - Really? Low-latency not very important, but high-bandwidth is
  - Now: HPC means power efficiency
    - Rapidly becoming important for HEP, everyone else
- The old times are never coming back
  - Today's HPC technologies are tomorrow's commodity technologies

# Exascale computing is coming

*President Obama, July 29, 2015:*

EXECUTIVE ORDER

CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and deployment, it is hereby ordered as follows:

…

Sec. 2. Objectives. Executive departments, agencies, and offices (agencies) participating in the NSCI shall pursue five strategic objectives:

1. **Accelerating delivery of a capable exascale computing system that integrates hardware and software capability to deliver approximately 100 times the performance of current 10 petaflop systems across a range of applications representing government needs.**

…

DOE is spending $2B on the Exascale Project

# Exascale

http://science.energy.gov/ascr/research/scidac/exascale-challenges/

- Power. Power, power, power.
  - Naively scaling current supercomputers to exascale would require a dedicated nuclear power plant to operate.
    - ALCF's Mira: 4 MW, 0.01 exaflop
    - "The target is 20-40 MW in 2020 for 1 exaflop"
- Exascale computing is the leading edge of advances of computing architecture
  - The same changes are happening outside of HPC, just not as quickly
    - Optimizing for Exascale is really optimizing for the future
- Storage of large-scale physics data sets will remain our job
- The Exascale machines will be a large fraction of the U.S. computing resources in the HL-LHC/DUNE era

# US DOE Supercomputer Centers

- NERSC (LBL)
  - Most “open” facility
  - Allocations are awarded by science offices
  - Current: Cori
    - Phase I: 2,388 Intel Xeon Haswell processor nodes
    - Phase 2: 9,688 Intel Xeon Phi Knight's Landing nodes
      - AVX-512
    - Top 500 Rank: 12
  - Next: Perlmutter
    - AMD + GPU
      - No AVX-512
    - to be delivered in 2020

# US DOE Supercomputer Centers

- Argonne Leadership Computing Facility (ALCF)
  - Time allocated primarily through competitive awards (INCITE, ALCC)
  - Current: Theta
    - 4,392 Intel Xeon Phi Knight's Landing nodes
      - Very similar to Cori Phase 2
    - Top 500 Rank: 24
  - Next: Aurora
    - Architecture: unknown*
      - Intel
      - definitely not Knight's anything
    - Coming in 2021
    - Scheduled to be first exascale machine


Aurora

# US DOE Supercomputer Centers

- Oak Ridge Leadership Computing Facility (OLCF)
  - Time allocated primarily through competitive awards (INCITE, ALCC)
  - Current: Summit
    - IBM POWER9™ 9,216 CPUs
    - NVIDIA Volta™ 27,648 GPUs
    - Top 500 Rank: 1
  - Next: Frontier
    - IBM
    - expected 2021, user availability expected in 2022
    - exascale


# Cultural Observations

- HPC tape storage tends to be write only
- Storage/CPU ratio much smaller than in HEP
- In HEP, jobs are typically allocated in tiny units, down to single cores
  - On Theta, the smallest allocatable unit is 8192 cores
- Cutting-edge C++ is popular in HEP
  - Viewed skeptically in HPC
- HPC people expect to use a variety of vendor-supplied compilers
  - Clang is rapidly taking over, sometimes as a frontend
- HPC users have no control over their surroundings
  - However, containers are rapidly gaining traction in HPC
- There are many HPC paradigms that are different than HEP paradigms
  - Not more or less complicated, just different

# Technology Disconnect

- Many pieces of software infrastructure are under active development to meet the needs of the Exascale project
  - Lower level:
    - OpenMP
    - HDF5
    - MPI
  - Higher level:
    - Kokkos & Raja
- Many things ubiquitous in HEP are unfamiliar (or worse) in the HPC community
  - TBB (uncommon, not unknown)
  - C++ after 14 (will happen eventually)
  - Root (unheard of)

# Fermilab Exascale Strategy

- Assume significant, but not all, CPU resources will come from Exascale
  - Work on Exascale-friendly software
    - Actively engage the HPC community
      - Pursue ASCR partners
      - (Re-)investigate mainstream HPC technologies
  - Work closely with our neighbors at Argonne
    - Also include work with Oak Ridge
  - Utilize HEPCloud to submit to Supercomputer facilities
    - more on HEPCloud later
- Maintain Fermilab as a primary storage site
  - Pursue a Terabit link with Argonne