ILLINOIS INSTITUTE OF TECHNOLOGY

High-Performance Computing Research

Zhiling Lan
Professor of Computer Science
lan@iit.edu
Office: 208E

ILLINOIS INSTITUTE OF TECHNOLOGY

Self Introduction

- **Ph.D. in Computer Engineering**, Northwestern University, 2002
- **Professor** of Computer Science
- **Guest Research Faculty** at Argonne National Lab
- **Research Interests**: high-performance computing (HPC), parallel and distributed systems
  - http://www.cs.iit.edu/~lan

2

ILLINOIS INSTITUTE OF TECHNOLOGY

My Research Group

Welcome to the SPEAR (Systems for Performance, Energy, And Resiliency) team's web page!

SPEAR

The team conducts research spanning various areas of parallel and distributed systems including cluster management, interconnection networking, performance modeling and simulation, power and energy efficiency, and fault tolerance. Our mission is to design scalable methods and software for large-scale HPC, AI, and data analysis. The team has a strong collaboration with the ALCF and MCS divisions at Argonne National Lab.

3

ILLINOIS INSTITUTE OF TECHNOLOGY

Students

- **Current**
  - Yuping Fan (PhD)
  - Xin Wang (PhD)
  - Yao Kang (PhD)
  - Boyang Li (PhD)
  - Melanie Cornelius (MS=>PhD)
  - Matt Dearing (PhD)
  - Dustin Favorite (MS)
  - Sahil Sharma (BS/MS)
  - Zhong Zheng (BS/MS)
- **Alumni**
  - Eduardo Berrocal (PhD, Intel)
  - Sean Wallace (PhD, Cray Inc.)
  - Xu Yang (PhD, Amazon)
  - Li Yu (PhD, Google)
  - Zhou Zhou (PhD, Salesforce)
  - Wei Tang (PhD, Google)
  - Ziming Zheng (PhD, Tripadvisor)
  - Jingjin Wu (PhD, Global Energy Research Institute)
  - Yawei Li (PhD, Google -> Uber)
  - Yongen Yu (MS, VMWare)
  - Jiexin Gu (MS, Google)
  - …

4

ILLINOIS INSTITUTE OF TECHNOLOGY

What is High-Performance Computing

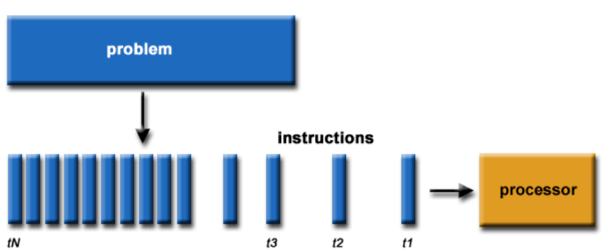
- Aka **supercomputing**
- Use of ***supercomputers*** for running *advanced* ***applications*** efficiently, reliably and quickly
  - **Supercomputer:** high-end computer with extremely fast processing capabilities, usually contains many processors
  - **Application:** typically from the fields of science and engineering, e.g., CFD, weather forecasting, …
    - Emerging AI, data analytics, IoT
  - **Parallel processing:** to divide a large problem into smaller ones to be solved concurrently

5

ILLINOIS INSTITUTE OF TECHNOLOGY

Serial Computing

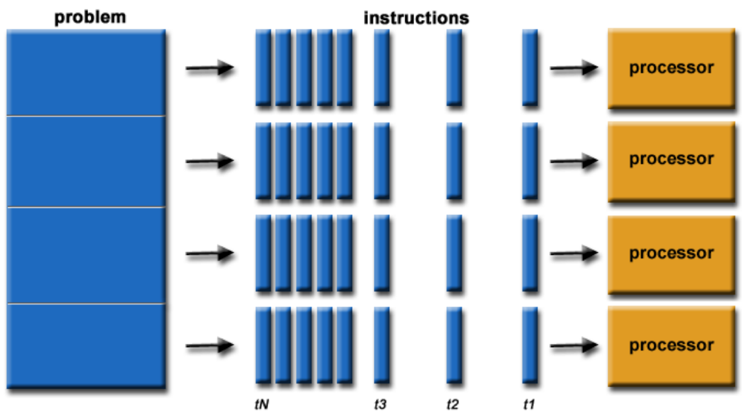
- Traditional, software has been written for serial computation
  - A problem is broken into a discrete series of **instructions**
  - Inst. are executed sequentially (one after another)
  - Executed on a single processor
  - Only one inst. executes at any moment in time



6

ILLINOIS INSTITUTE OF TECHNOLOGY

Parallel Computing



7

ILLINOIS INSTITUTE OF TECHNOLOGY

Parallel Computing

- Addition of 4 numbers: 3+4+5+6

Sequential

	T1	T2	T3	
P+M	3+4	7+5	12+6	After 3 time steps, have the sum of 18

Speedup: Sequential/Parallel = 3/2 = 1.5

Parallel

P+M	3+4	7+11	After 2 time steps, have the sum of 18
P+M	5+6		

Credit: Valerie Taylor at Argonne National Lab

8

ILLINOIS INSTITUTE OF TECHNOLOGY

Scale Up: Problem Size & Processors

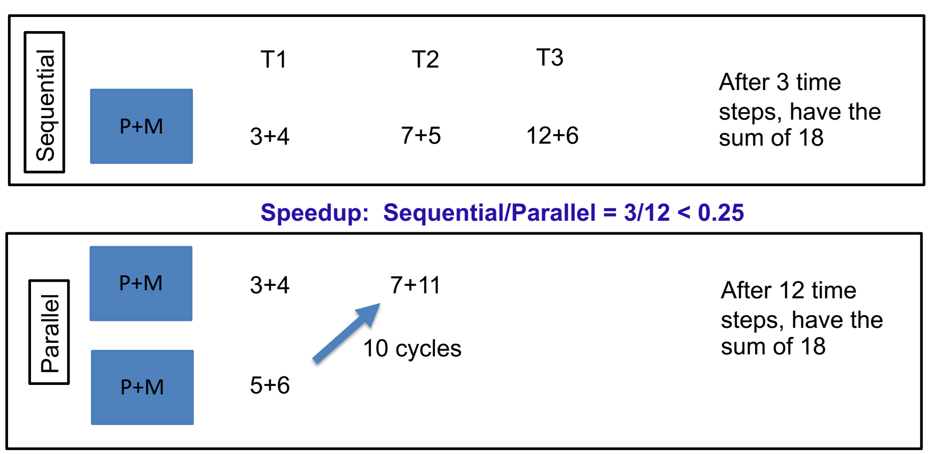
Numbers to Add	Speedup 2 Processors
4	1.5
20	1.9
40	1.95
200	1.99
400	1.995

Number of Processors (400 Numbers)	Speedup
2	1.995
4	3.95
16	14.25
32	23.47
64	33.25

9

ILLINOIS INSTITUTE OF TECHNOLOGY

Parallel Processing: Add Communication



10

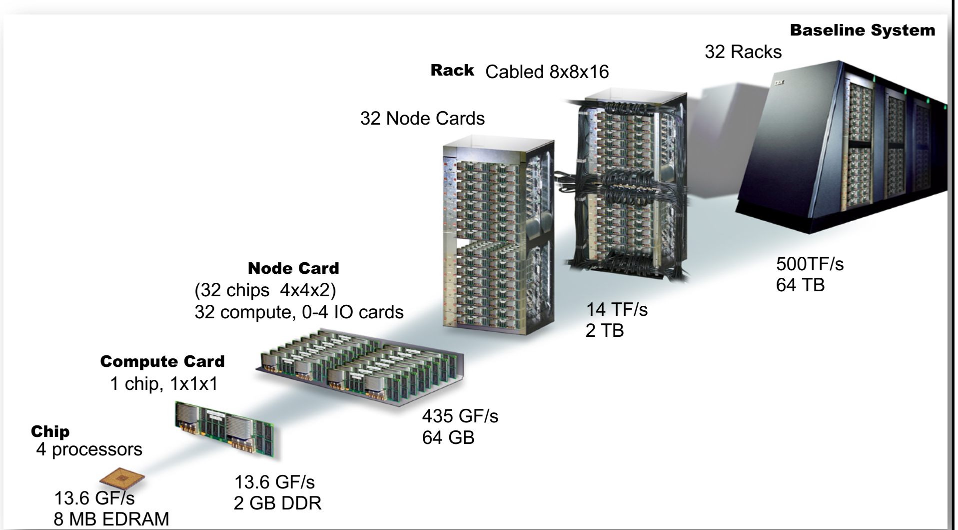
Scale Up: With Communication

Numbers to Add	Speedup 2 Proc	With Comm
4	1.5	0.25
20	1.9	0.95
40	1.95	1.30
200	1.99	1.81
400	1.995	1.90

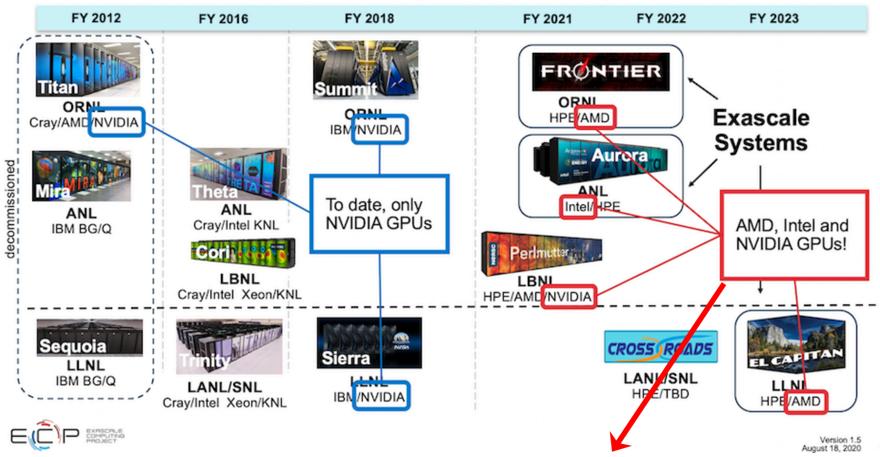
Number of Processors (400 Numbers)	Speedup	With Comm
2	1.995	1.90
4	3.95	3.30
16	14.25	8.31
32	23.47	5.96
64	33.25	5.41

☞ Parallel programming is challenging!

What is Supercomputer?

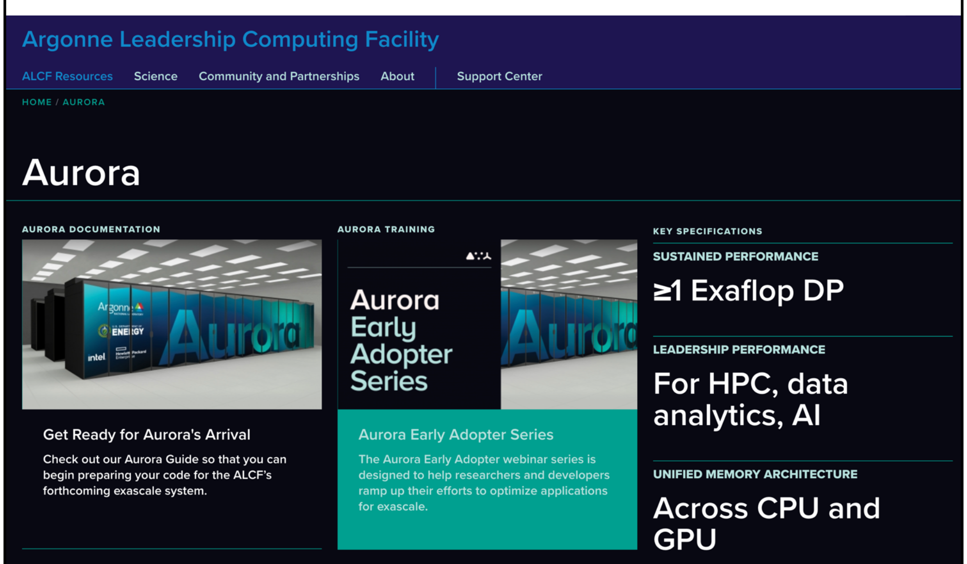


DOE HPC Roadmap to ECP Target Exascale Systems



heterogeneous multi-core system

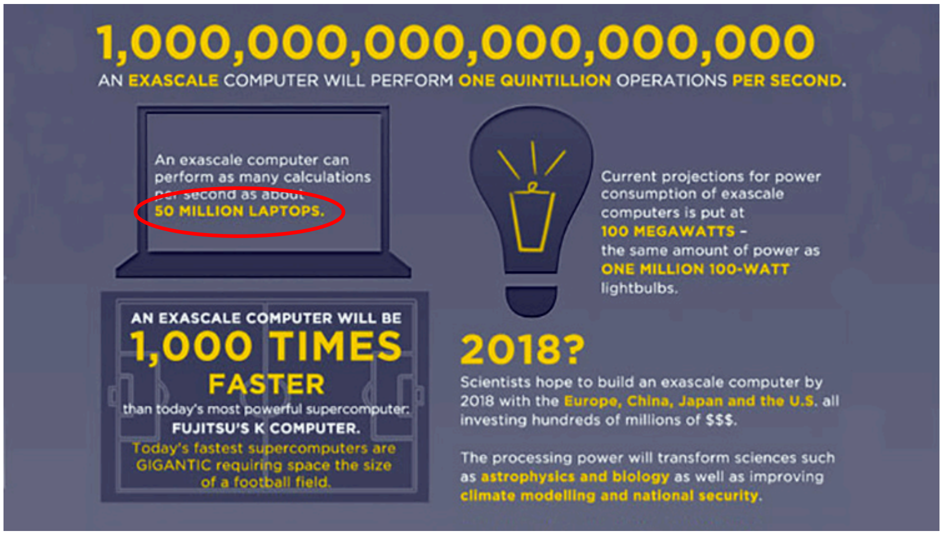
ILLINOIS INSTITUTE OF TECHNOLOGY



https://alcf.anl.gov/aurora

ILLINOIS INSTITUTE OF TECHNOLOGY

What is Exascale?

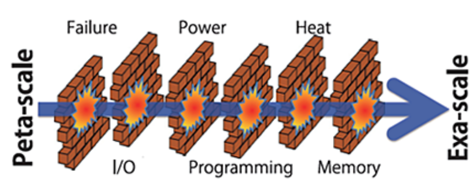


15

ILLINOIS INSTITUTE OF TECHNOLOGY

Research Challenges

- Energy and power challenge
- Memory and storage challenge
- Concurrency and locality challenge
- Resilience challenge



Various walls to go through

Peter Kogger, et al., "ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems", DARPA, 2008.

16

My Research

- Develop **system software** for
  - Resource management and job scheduling
  - Power and energy efficiency
  - Fault tolerance (resilience)
  - Networking and communication modeling
  - Performance analysis and modeling

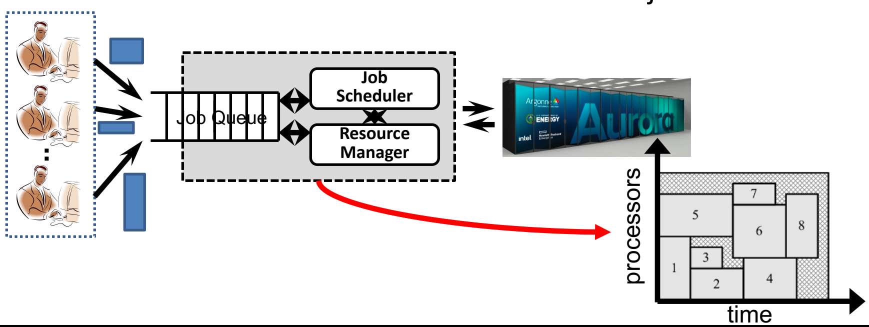
Current Research Projects

- **MINT**: **M**ulti-resource **IN**telligen**T** management of hybrid workloads *(NSF funded)*
- **SEEr**: **S**calabe, **E**nergy-**E**fficient HPC environment for AI-enabled science *(NSF funded)*
- **IRON**: **I**nterference **R**educti**ON** on massively parallel platforms *(NSF funded)*
- **MUMMI-R**: Experimental-based Research on Effective Models of Parallel Application Execution Time, Power, and Resilience *(NSF funded)*
- **Cobalt**: A High Performance, Multi-Dimensional Batch Scheduler for Pre-exascale and Beyond Systems *(DOE funded)*
- More details on my webpage:
  - http://www.cs.iit.edu/~lan/

ILLINOIS INSTITUTE OF TECHNOLOGY

Workload Management

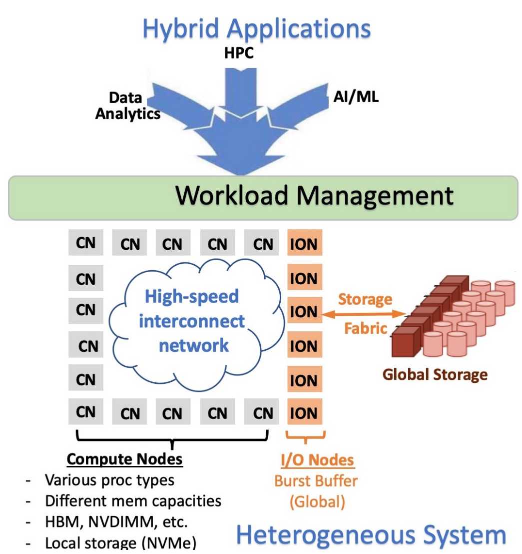
- Resource management and job scheduling
  - A system software sitting between users and hardware platform
  - Users specify job requirement, e.g., computation and time
  - User jobs are placed into queue to wait for execution
  - Decide when and where to execute user jobs



19

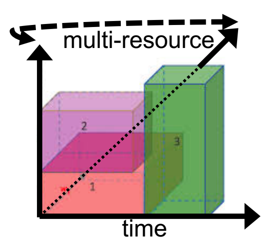
ILLINOIS INSTITUTE OF TECHNOLOGY

Application & System Evolutions



Open questions:

- **Hybrid workloads** with various resource requirements
  - Compute-, memory-, communication-, IO-intensive
- **Heterogeneous system** with different processors & deep memory/storage hierarchies
- Various **user demands**
  - Response time, on-demand, wait time, …

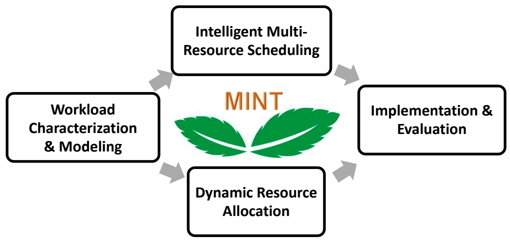


20

ILLINOIS INSTITUTE OF TECHNOLOGY

MINT (Multi-resource INtelligent managemenT of hybrid workloads)

To catalyze a harmonic HPC, AI, and data analytics environment



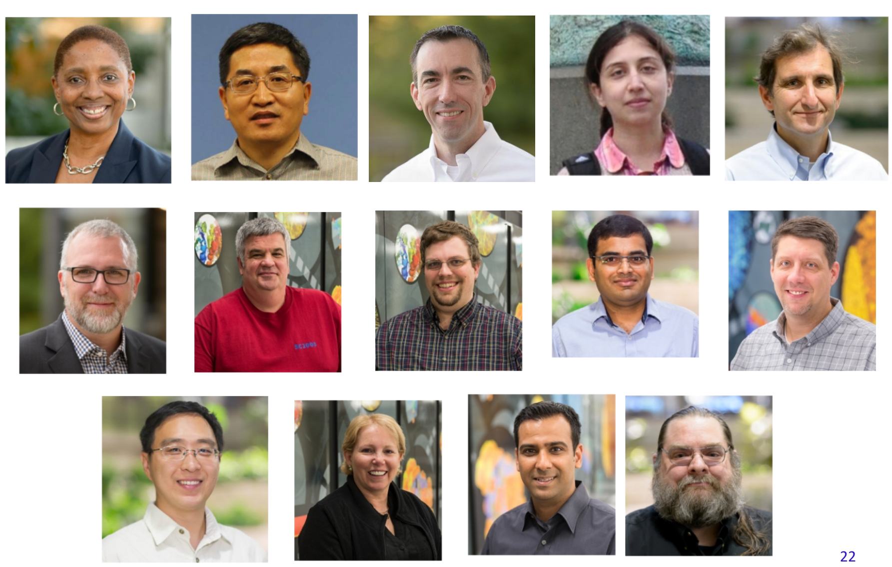
An **intelligent** workload management framework

- To promote the use of large-scale supercomputers for emerging data-centric applications (**HPC4AI**)
- To exploit advanced AI technologies, especially multi-objective reinforcement learning, to empower job scheduling and resource allocation in HPC (**AI4HPC**)

21

ILLINOIS INSTITUTE OF TECHNOLOGY

Collaborators



22

ILLINOIS INSTITUTE OF TECHNOLOGY

Student Opportunities

- Sahil Sharma (BS/MS, 9/2019- 5/2021) (Project: DNPC)
- Avery Peck (BS/MS, 1/2020- 12/2020)
- Sergio Servantez (BS, 1/2019-5/2019) (Project Link)
- Zhen Huang and Black Ehrenbeck (BS, 1/2019-5/2019) (Project Link)
- Sergio Servantez (BS, 8/2018-12/2018) (Project Link)
- Zhen Huang, Blake Ehrenbeck, Brianna Bransfield (BS, Student Cluster Competition at SC2018, 8/2018-11/2018)
- Zhen Huang (BS, CS Scholarship, 1/2018-5/2018)
- Blake Ehrenbeck (BS, NSF REU, 7/2017-9/2017)
- Arushi Rai (BS, NSF REU, 7/2017-9/2017)
- Shreyas Moudgalya (BS, 10/2016-11/2016)
- Aleksandra Kukielko (BS, NSF REU, 5/2016-8/2016)
- Jia Hao He (BS, NSF REU, 2/2016-3/2016)
- Tarun Gidwani (BS, NSF REU, 8/2014-10/2014)
- Asad Patel (BS, NSF REU, 8/2014-12/2014)
- Runran Wang (BS, REU summer, 2010)
- Kunlun Guo (BS, REU summer, 2010)
- Janusz Nosek (BS, REU summer, 2009)
- Soo Min Park (BS, 2007)

Requirements:

- CS351 & CS451
  - Five former CS451 students had summer internships at Argonne
- C/C++, Python
- Various research projects
- If interested, send me an email with your CV and transcripts!

23

ILLINOIS INSTITUTE OF TECHNOLOGY

Contact: Zhiling Lan

Email: lan@iit.edu

http://www.cs.iit.edu/~lan