TEXAS TECH UNIVERSITY™

CS4379: Parallel and Concurrent Programming
CS5379: Parallel Processing

Lecture 1

Dr. Yong Chen

Associate Professor

Computer Science Department

Texas Tech University

TEXAS TECH UNIVERSITY

Course Info

- **Lecture Time**: TR, 12:30-1:50
- **Lecture Location**: ECE 217
- **Sessions**: CS4379-001, CS4379-002, CS5379-001, CS5379-D01
- **Instructor:** Yong Chen, Ph.D., Associate Professor
- **Email:** yong.chen@ttu.edu
- **Phone:** 806-834-0284
- **Office**: Engineering Center 315
- **Office Hours**: 2-4 p.m. on Wed., or by appointment
- **TA: TBA**
- **More info:**
  - http://www.myweb.ttu.edu/yonchen
  - http://discl.cs.ttu.edu; http://cac.ttu.edu/; http://nsfcac.org

About Me

- Associate Professor in Computer Science, Director of the Data-Intensive Scalable Computing Laboratory (DISCL), Site Director of the Cloud and Autonomic Computing (CAC) Center
  - http://discl.cs.ttu.edu/
  - http://cac.ttu.edu/
  - http://nsfcac.org

- My research lab focuses on data-intensive computing, parallel and distributed computing, high-performance computing, cloud computing, computer architectures and system software

Wordle of My Publication Titles

Acknowledgement: http://www.wordle.net/

About My Research Group: DISCL@TTU

- Mission statement
  - Broad research interests in parallel and distributed computing, high-performance computing, cloud computing with a focus on building scalable computing systems for data-intensive applications
- Four associated faculty members
- Seven Ph.D. student members
- Eight Masters/UG student members
- Website: http://discl.cs.ttu.edu/

- Projects
  - *xBGAS (Extended Base Global Address Space)*
  - *Tuning Extreme-scale Storage Stack through Deep Learning (NSF)*
  - *Empowering Data-driven Discovery with a Lightweight Provenance Service for HPC*
  - *Uncovering Vulnerabilities in Parallel File Systems for Reliable High Performance Computing*
  - *OpenHPC: Open Source, Extensible High Performance Computing Platform*
  - *Compute on Data Path: Combating Data Movement in High Performance Computing (NSF)*
  - *Unistore: A Unified Storage Architecture (NSF/Nimboxx)*
  - *Development of a Data-Intensive Scalable Computing Instrument for HPC (NSF)*

DISCL@TTU Resources

- A state-of-the-art 40-node 800-core NSF-DISCI (Data-Intensive Scalable Computing Instrument) system with Xeon Phi co-processors, flash memory storage, and Qlogic QDR Infiniband
  - Funded by NSF (National Science Foundation) MRI (Major Research Instrumentation) program

TEXAS TECH UNIVERSITY

DISCL@TTU Resources (cont.)

- DISCFarm cluster
  - discfarm.hpcc.ttu.edu
  - 16-node 128-core
  - Dual quad-core processors each node
  - RAID-5 storage drives, 10 SSDs
- GPU boards: GTX480, Quadro FX 5800, more donations
- iMac workstations, MBA laptops
- Dell workstations, printers
- Standalone servers/testbed
- More details
  - http://discl.cs.ttu.edu/doku.php?id=resources

NSF Industry/University Cooperative Research Center at Texas Tech University (I/URCRC)

TEXAS TECH UNIVERSITY

THE UNIVERSITY OF ARIZONA

MISSISSIPPI STATE UNIVERSITY

VCU VIRGINIA COMMONWEALTH UNIVERSITY

- Current & past member

DMTF

HAPPY STATE BANK

SOLIEL

Raytheon

CovenantHealth

DELL EMC

DISA DEFENSE INFORMATION SYSTEMS AGENC DEPARTMENT OF DEFENSE

intel

IBM

NORTHROP GRUMMAN

STACK VELOCITY

Outline

- Questions?

- About me and my research group
- Syllabus highlights
- Introduction to parallel computing

Syllabus Highlights – Course Description

- The advent of multicore processors has completely changed the landscape of computing
- The serial computing era when programmers could transparently and automatically take advantage of the microprocessors' performance improvement that follows the Moore's law has gone
- Parallel processing has now become universal
  - Multicore computing and GPGPU computing on chip
  - Large-scale cluster computing
  - Grid computing
  - Cloud computing

Syllabus Highlights – Course Description

- This course introduces the principles of parallel computing and covers parallel and concurrent programming
- Covers general issues in parallel computing and programming
  - Parallel computer architectures
  - Parallel algorithms
  - Parallel programming models

Syllabus Highlights –Textbooks

- Reference textbook:
  - *Introduction to Parallel Computing (2nd Edition)*, by Ananth Grama, George Karypis, Vipin Kumar, Anshul Gupta
  - Publisher: Addison Wesley
  - ISBN-10: 0201648652
  - ISBN-13: 978-0201648652

Syllabus Highlights –Textbooks

- Reference textbooks (additional):
  - *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering*, by Ian Foster
    - Publisher: Addison Wesley
    - ISBN-10: 0201575949
    - ISBN-13: 978-0201575941
    - Online version: http://www.mcs.anl.gov/~itf/dbpp/text/book.html
  - *MPI: The Complete Reference (2-volume set)*, by Marc Snir, et al.
    - Publisher: The MIT Press
    - ISBN-10: 0262692163
    - ISBN-13: 978-0262692168
  - *The Sourcebook of Parallel Computing,* by Jack Dongarra et al.
    - Publisher: Morgan Kaufmann
    - ISBN-10: 1558608710
    - ISBN-13: 978-1558608719

Syllabus Highlights – Prerequisites

- Pre-requisites
  - CS3375 Computer Architecture
  - CS3364 Design and Analysis of Algorithms
  - Operating systems – CS4352 Operating Systems or equivalent
  - Unix/Linux experience
  - Programming (C/C++)
- Students should have done programming in C on a UNIX/Linux platform

Syllabus Highlights – Key Topics

- Overview of parallel computing
- Performance metrics and evaluation for parallel systems
- Parallel algorithms design and applications
- Shared memory parallel architectures and multithreaded programming
- Distributed memory parallel architectures and communications, message passing programming
- For CS5379: Exploration of latest research and development in parallel processing

Course Schedule

- Week Topic
- W1 Syllabus highlights and introduction of parallel computing
- W2 Parallel computer architectures: classification based on control mechanism/address space, interconnection networks
- W3 Evaluating network topologies, parallelism, data dependence, and control dependence
- W4 Parallelization, parallel performance and evaluation, scalable computing, scaled speedup, scalability of parallel systems
- W5 Using parallel computers at High Performance Computing Center on campus
- W6 Principles of parallel algorithm design, parallel algorithm models, matrix-matrix multiplication, Gaussian elimination algorithms
- W7 Sorting, quicksort, parallelizing quicksort; search, depth-first search, parallelizing DFS
- W8 Programming shared address space platforms; thread basics

Course Schedule (cont.)

- W9 PThreads programming
- W10 OpenMP programming model
- W11 Distributed address space and message passing; basic communication operations, One-to-all broadcast and all-to-one reduction, all-to-all broadcast and reduction
- W11 All-reduce, scatter and gather, all-to-all personalized communication, principles of message-passing programming and MPI basics
- W12 Blocking v.s. non-blocking, collective communications, communication modes, deadlocks, datatypes
- W13 Topology, communicators, understanding the performance and behavior
- W14 Review and assignment discussions, tour to High Performance Computing Center on campus
- W15 Graduate students project presentations

Topics and/or dates may be changed during the semester at the instructor’s discretion because of scheduling issues, developments in the discipline, or other contingencies.

Syllabus Highlights – Computer Usage

- A compute cluster will be used for the development of projects and assignments
- Detailed information regarding the access will be announced later

Syllabus Highlights – Grading Policy

- All submissions are graded according to the assignment guidelines, course policies, verbal instructions/explanations
- Quality, creativity, novelty, readability and level of effort

- For CS4379-001, 002
- Written assignments: 30% (four assignments, weighing 7.5% each)
- Quizzes: 35% (five quizzes, weighing 7% each)
- Programming projects: 35% (four programming projects, weighing 5% for the first project and 10% for each of the later three projects)

- For CS5379-001, CS5379-D01
- Written assignments: 25% (four assignments, weighing 6.25% each)
- Quizzes: 30% (five quizzes, weighing 6% each)
- Programming projects: 30% (four programming projects, weighing 4%, 8%, 8%, and 10%, respectively)
- Course research project: 15% (including written report and oral presentation)

TEXAS TECH UNIVERSITY

Syllabus Highlights - Grades

- Course Letter Grade Assignments

- For all CS4379-001/002, CS5379-001, CS5379-D01
- A=[86, 100]
- B=[76, 86)
- C=[66, 76)
- D=[50, 66)
- F=[0, 50)

Syllabus Highlights – Course Policies

- Code of Student Conduct
- Students are expected to comply with the Texas Tech Code of Student Conduct in all aspects of this class.
- In order to assure that all students have the opportunity to gain from time spent in class, unless otherwise approved by the instructor, students are prohibited from engaging in any other form of distraction, such as reading newspapers, working on other classes, taking cell phone calls, and working on laptop computers.
- Violations of conduct including academic dishonesty, foul language, and classroom citizenship are eligible to be reported to Student Judicial Services.

Syllabus Highlights – Course Policies

- **Attendance:** Attendance is mandatory for all lectures
- **Absence due to religious observance/officially approved trips:** need to notify the instructor
- **Late Work:** assignments are due when specified, but will be accepted late (with a 10-20% penalty) till a specified hard deadline

Syllabus Highlights – Course Policies

- **Accommodations:**
- The university is committed to the principle that in no aspect of its programs, shall there be differences in the treatment of persons because of race, creed, national origin, age, sex, or disability and that equal opportunity and access to facilities shall be available to all. If you require special accommodations in order to participate, please contact the instructor during office hours or by e-mail yong.chen@ttu.edu. Students should present appropriate verification from Student Disability Services. No requirement exists that accommodations be made prior to completion of this approved university process.

Caveats

- Parallel computing is a hard topic
- This course can be a challenging course if you don't have enough required background and don't work hard
- If you want to get an easy grade, please consider staying away from this course
- On the other hand, if you complete this course with outstanding performance, you might be offered assistantship opportunities and your hard work can be paid off in your future career endeavor

Additional Information

- Will use PPT slides to present lectures, instead of blackboard/whiteboard

- Will focus on major and critical points
  - Parallel computing domain contains a lot of materials
  - A symmetric but more challenging world than sequential processing
  - Lectures are primarily aligned with three directions
    - Architectures
    - Programming
    - Algorithms

- Will use Blackboard system to post lectures, assignments, and collect submissions

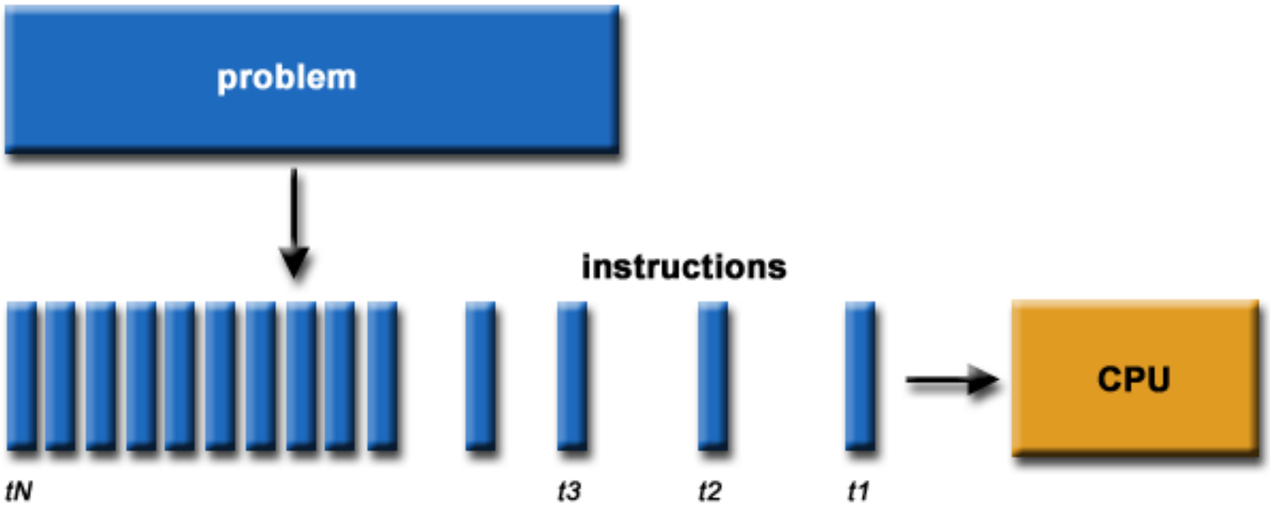
- If you write me an email for this class, please start the email subject with [CS4379] or [CS5379]

Outline

- Questions?
- About me and my research group
- Syllabus highlights
- Introduction to parallel computing

What is Parallel Computing

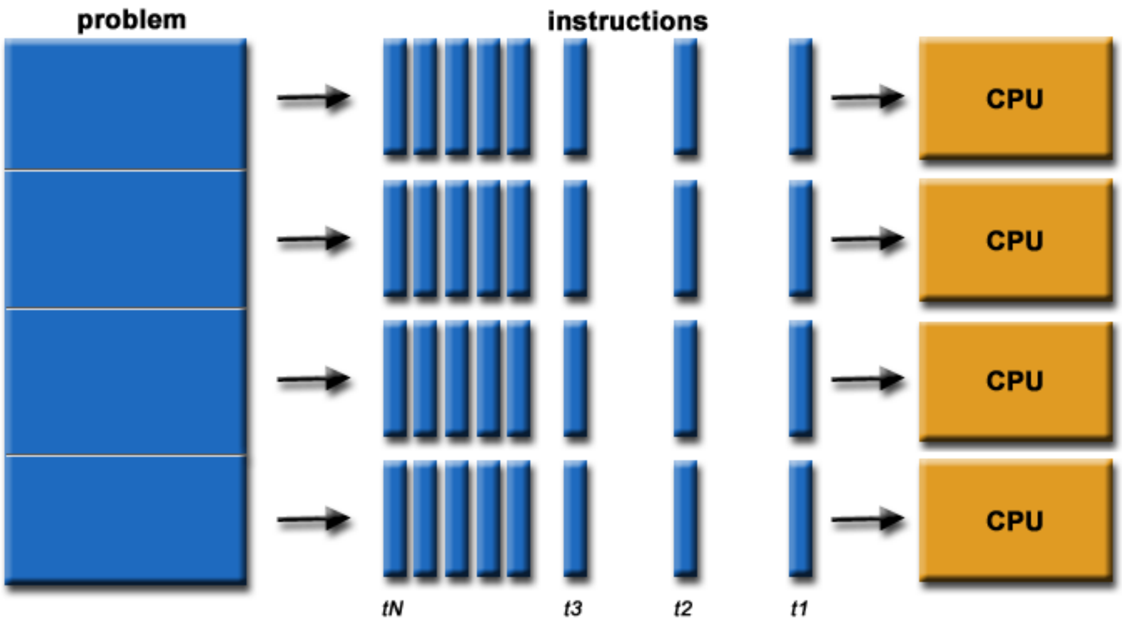
- Traditionally, software has been written for serial computation
  - To be run on a single computer having a single processor
  - A problem is broken into a discrete series of instructions
  - Instructions are executed one after another
  - Only one instruction may execute at any moment in time.



Courtesy: Blaise Barney, LLNL

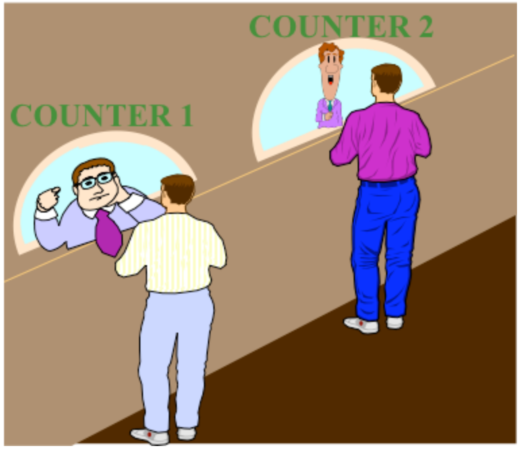
What is Parallel Computing

- In the simplest sense, parallel computing is the simultaneous use of multiple compute resources to solve a problem
  - To be run using multiple CPUs
  - A problem is broken into discrete parts that can be solved concurrently
  - Each part is further broken down to a series of instructions
  - Instructions from each part execute simultaneously on different CPUs



Daily Life Examples

- Wal-Mart checkout lines
- House construction
- Car manufacturing
- Call center
- …



TEXAS TECH UNIVERSITY

Why Parallel Computing

- Desire to solve bigger and more realistic applications
  - "Grand Challenge" problems
  - Web search engines/databases processing millions of trans per second

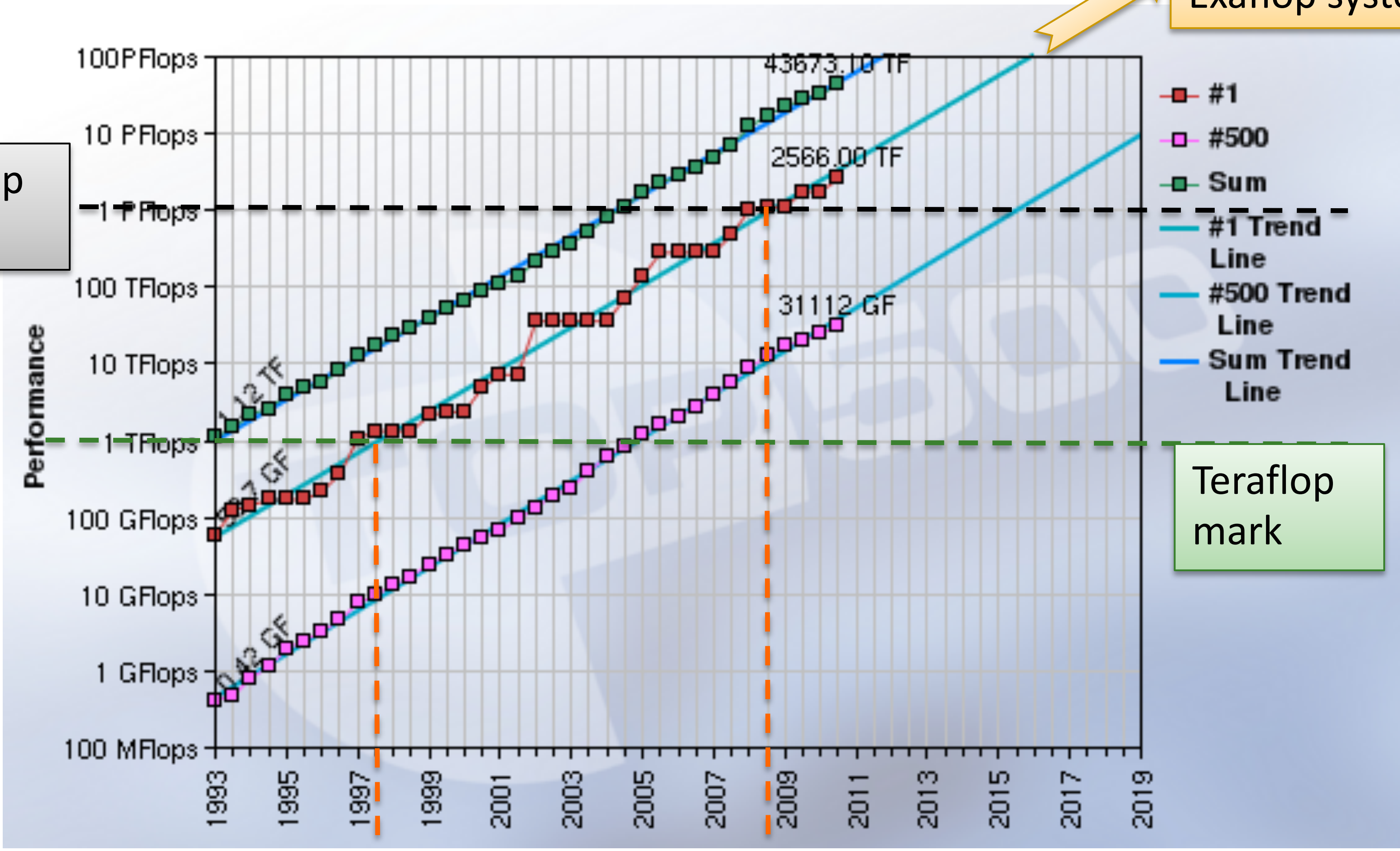
- Fundamental limits are being approached
  - E.g. limits to miniaturization - a limit will be reached on transistors integrated (post-Moore's Law era)
  - E.g. power consumption limit (frequency scaling stopped ~2005 -> multicore processors are the norm nowadays, "parallel computing" everywhere)
  - If we want faster computers, we have to build "parallel computers"

- More cost effective solution
  - It is increasingly expensive to make a single processor faster
  - Pollack's rule – the return of investment (i.e. performance increase) is roughly proportional to the square root of the investment (i.e. the number of transistors increased on chip)

TOP500 Supercomputers

Exaflop system: ~2022

Petaflop mark



Teraflop mark

1998 2008

Source: TOP500 site

1000x performance improvement over 10 years. ~10x faster than Moore's law!

TEXAS TECH UNIVERSITY™

Top 10 Machines as of 11/2010 (SC'10)

Rank	Site	Computer	Cores	Rmax	Rpeak
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C NUDT	186,368	2,566,000	4,701,000
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.	224,162	1,759,000	2,331,000
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU Dawning	120,640	1,271,000	2,984,300
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows NEC/HP	73,278	1,192,000	2,287,630
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz Cray Inc.	153,408	1,054,000	1,288,627
6	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 Bull SA	138,368	1,050,000	1,254,550
7	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband IBM	122,400	1,042,000	1,375,776
8	National Institute for Computational Sciences/University of Tennessee United States	Kraken XT5 - Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.	98,928	831,700	1,028,851
9	Forschungszentrum Juelich (FZJ) Germany	JUGENE - Blue Gene/P Solution IBM	294,912	825,500	1,002,701
10	DOE/NNSA/LANL/SNL United States	Cielo - Cray XE6 8-core 2.4 GHz Cray Inc.	107,152	816,600	1,028,659
110	Texas Tech University United States	Hrothgar - Poweredge Cluster, Xeon X5660 2.8Ghz, Infiniband DDR / 2010 Dell	7524	68.27	84.54

TTU HPCC

http://www.top500.org

TEXAS TECH UNIVERSITY

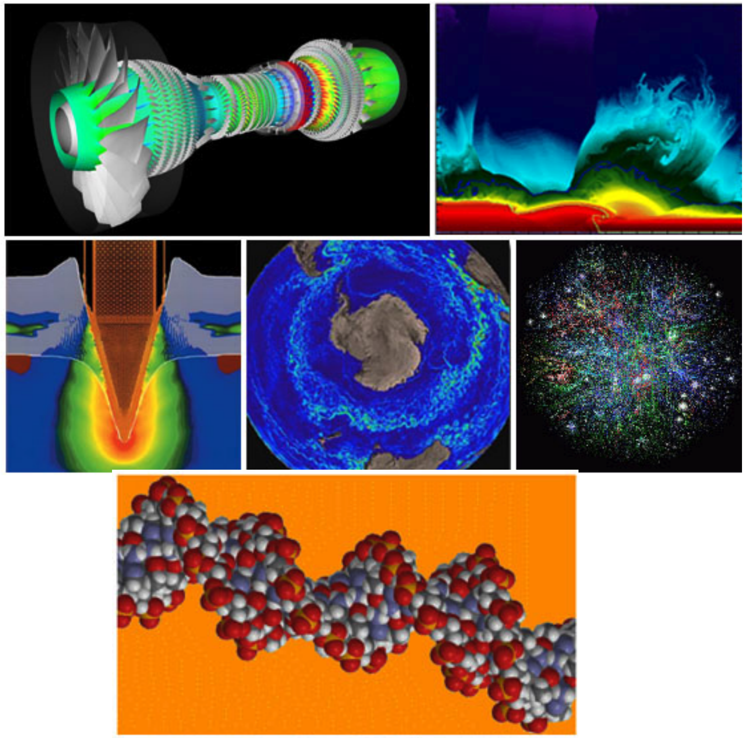
Top 10 Systems, as of 6/2018 (ISC'18)

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,282,544	122,300.0	187,659.3	8,806
2	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
3	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/NNSA/LLNL United States	1,572,480	71,610.0	119,193.6	
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
5	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2550 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	32,576.6	1,649
6	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
7	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
8	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
9	**Trinity** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,968	14,137.3	43,902.6	3,844
10	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939

Summit@ORNL ~200-petaflop (peak), ISC'18 June 2018 list

Parallel Computing Applications, Historically

- Atmosphere, Earth, Environment
- Physics - applied, nuclear, particle, condensed matter, high pressure, fusion, photonics
- Bioscience, Biotechnology, Genetics
- Chemistry, Molecular Sciences
- Geology, Seismology
- Mechanical Engineering
- Electrical Engineering
- Circuit Design, Microelectronics
- Computer Science, Mathematics



Parallel Computing Applications, Today

- Today, commercial applications provide an equal or greater driving force in the development of faster computers
- These applications require the processing of large amounts of data in sophisticated ways
- Databases, data mining
- Web search engines
- Web based business services
- Medical imaging and diagnosis
- Management of national and multi-national corporations
- Financial and economic modeling
- Advanced graphics and virtual reality
- Entertainment industry
- Networked video and multi-media technologies
- Collaborative work environments
- Parallel computing is universal nowadays with multicore/manycore

Readings

- Reference book ITPC, Chapter 1

Questions?

Questions/Suggestions/Comments are always welcome!

Write me: yong.chen@ttu.edu
Call me: 806-834-0284
See me: ENGCTR 315

If you write me an email for this class, please start the email subject with [CS4379] or [CS5379].