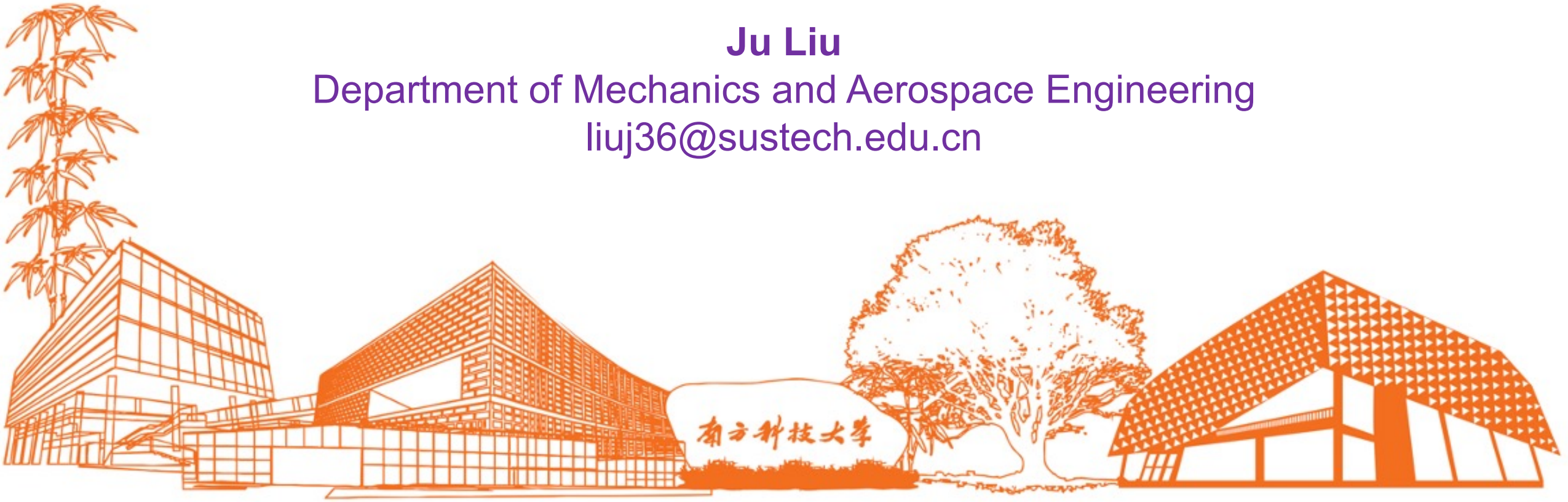
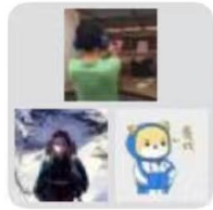
MAE 5032 High Performance Computing: Methods and Applications

Lecture 1: Introduction

Ju Liu
Department of Mechanics and Aerospace Engineering
liuj36@sustech.edu.cn





MAE5032 HPC

What is it?

Only that which is well done is quickly done.

~ Augustus

AZ QUOTES

HPC

- High Performance Computing (HPC) refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, medicine, or business.

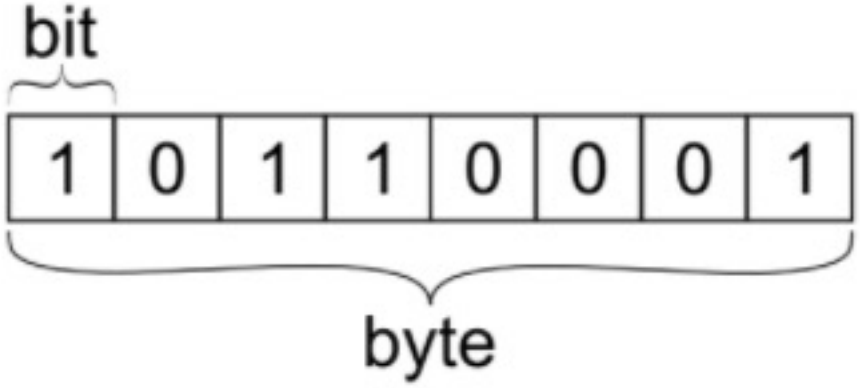
It involves knowledges on hardwars, softwares, applied math, and the application background.

An Overview of the current fastest computers

The Top500 List

Units of measure in HPC

- High Performance Computing (HPC) units are:
  - Flop : floating point operation, usually double precision unless noted
  - Flop/s : flowating point operations per second
  - Bytes : size of data (a double precision floating point number is 8 bytes;



Units of measure in HPC

- High Performance Computing (HPC) units are:
  - Flop : floating point operation, usually double precision unless noted
  - Flop/s : flowating point operations per second
  - Bytes : size of data (a double precision floating point number is 8 bytes; one byte contains 8 bits);
- Typical sizes are millions, billions, trillions…

Kilo	Kflop/s = 10^3 flop/sec	Kbyte = $10^3 \sim 2^{10}$ = 1024 bytes (KiB)
Mega	Mflop/s = 10^6 flop/sec	Mbyte = $10^6 \sim 2^{20}$ bytes (MiB)
Giga	Gflop/s = 10^9 flop/sec	Gbyte = $10^9 \sim 2^{30}$ bytes (GiB)
Tera	Tflop/s = 10^{12} flop/sec	Tbyte = $10^{12} \sim 2^{40}$ bytes (TiB)
Peta	Pflop/s = 10^{15} flop/sec	Pbyte = $10^{15} \sim 2^{50}$ bytes (PiB)
Exa	Eflop/s = 10^{18} flop/sec	Ebyte = $10^{18} \sim 2^{60}$ bytes (EiB)
Zetta	Zflop/s = 10^{21} flop/sec	Zbyte = $10^{21} \sim 2^{70}$ bytes (ZiB)
Yotta	Yflop/s = 10^{24} flop/sec	Ybyte = $10^{24} \sim 2^{80}$ bytes (YiB)

Current fastest machines are etaflop systems.

The TOP500 Project

- Listing the 500 most powerful (public) computers in the world

- Yardstick: Floating Point Operations per Second (FLOP/s) Rmax of Linpack
  - Solve Ax = b, with the matrix A being a dense matrix with random entries
  - Dominated by dense matrix-matrix multiply

- Updated twice a year:
  - ISC'xy in June in Germany
  - SCxy in November in US

- All information available from the TOP500 website at www.top500.org

Top 10 of the Top500, Nov. 2021

Clock rate: the frequency at which the transistor switches on and off. Typically between 1 an 3 GHz.

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**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	2,414,592	148,600.0	200,794.9	10,096
3	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	**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
5	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70,870.0	93,750.0	2,589
6	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63,460.0	79,215.0	2,646
7	**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
8	**JUWELS Booster Module** - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite, Atos Forschungszentrum Juelich (FZJ) Germany	449,280	44,120.0	70,980.0	1,764
9	**HPC5** - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, DELL EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252
10	**Voyager-EUS2** - ND96amsr_A100_v4, AMD EPYC 7V12 48C 2.45GHz, NVIDIA A100 80GB, Mellanox HDR Infiniband, Microsoft Azure Azure East US 2 United States	253,440	30,050.0	39,531.2	

More of the Top500, Nov. 2021

433	**AFI-NITY** - PRIMERGY CX2550 M4, Xeon Gold 6148 20C 2.4GHz, Infiniband EDR, Fujitsu Tohoku University, Institute of Fluid Science Japan	35,200	1,691.0	2,703.4	
434	**TaiYi** - ThinkSystem SD530, Xeon Gold 6148 20C 2.4GHz, Intel Omni-Path, Lenovo Southern University of Science and Technology China	32,400	1,686.5	2,488.3	
435	**NA-IT1** - ZettaScaler3.0, AMD EPYC 7702P 64C 1.5GHz, PEZY-SC3, Infiniband EDR, PEZY Computing / Exascaler Inc. NA Simulation Japan	822,400	1,684.8	2,353.8	22

Summit (#2 machine) system overview

- Peak performance of 200 petaflops for modeling & simulation
- Peak of 3.3 ExaOps for data analytics and artificial intelligence

Each node has
- 2 IBM POWER9 processors
- 6 NVIDIA Tesla V100 GPUs
- 608 GB of fast memory
- 1.6 TB of NVMe memory

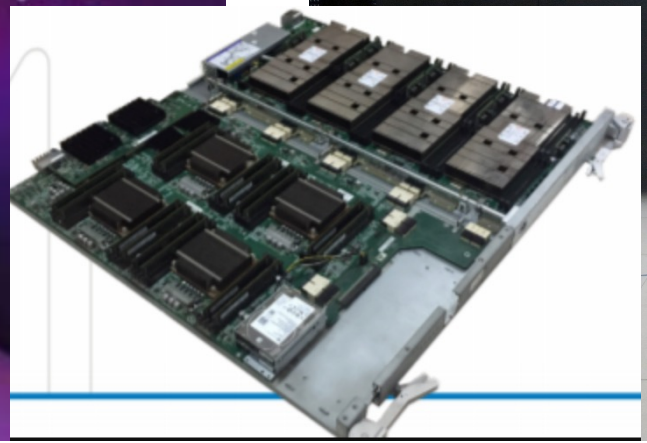
The system has
•4608 nodes
•Dual-rail Mellanox EDR InfiniBand network
•250 PB IBM Spectrum Scale file system transferring data at 2.5 TB/s

TianHe-2A (#7 machine) system overview

- Peak performance of 100 petaflops for modeling & simulation
- Power 18482 kW
- OS: Kylin Linux

Each node has
- 24 Xeon E5 CPU 2.2GHz
- 3 x 57 coprocessors (Xeon phi / Matrix-2000)
- 64 + 3 x 8 = 88 GB Memory

The system has
- 16000 compute nodes
- Self-developed TH express-2
- 12 PB file system



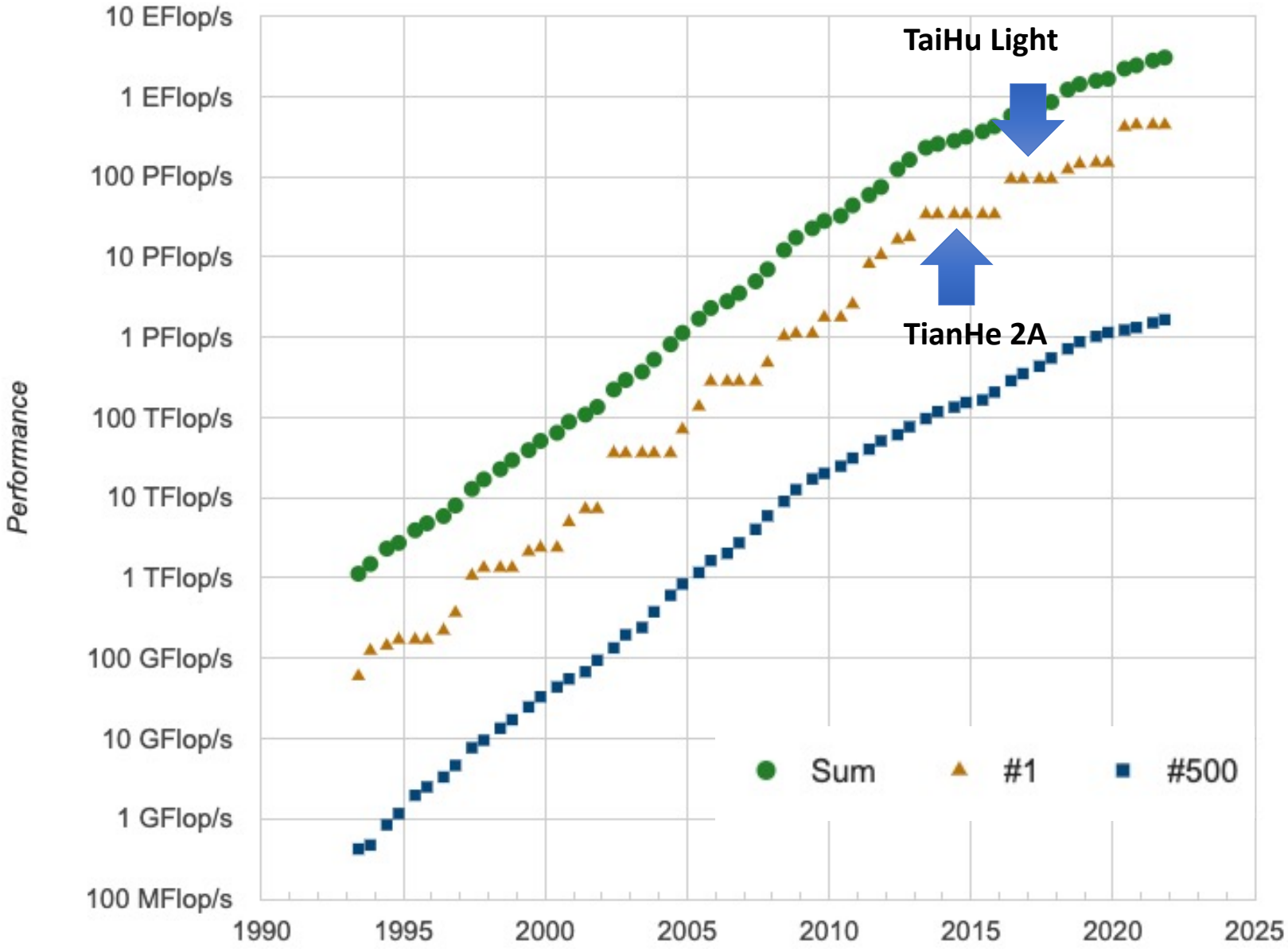
TaiYi (#434 machine) system overview

- Peak performance of 2.5 petaflops for modeling & simulation

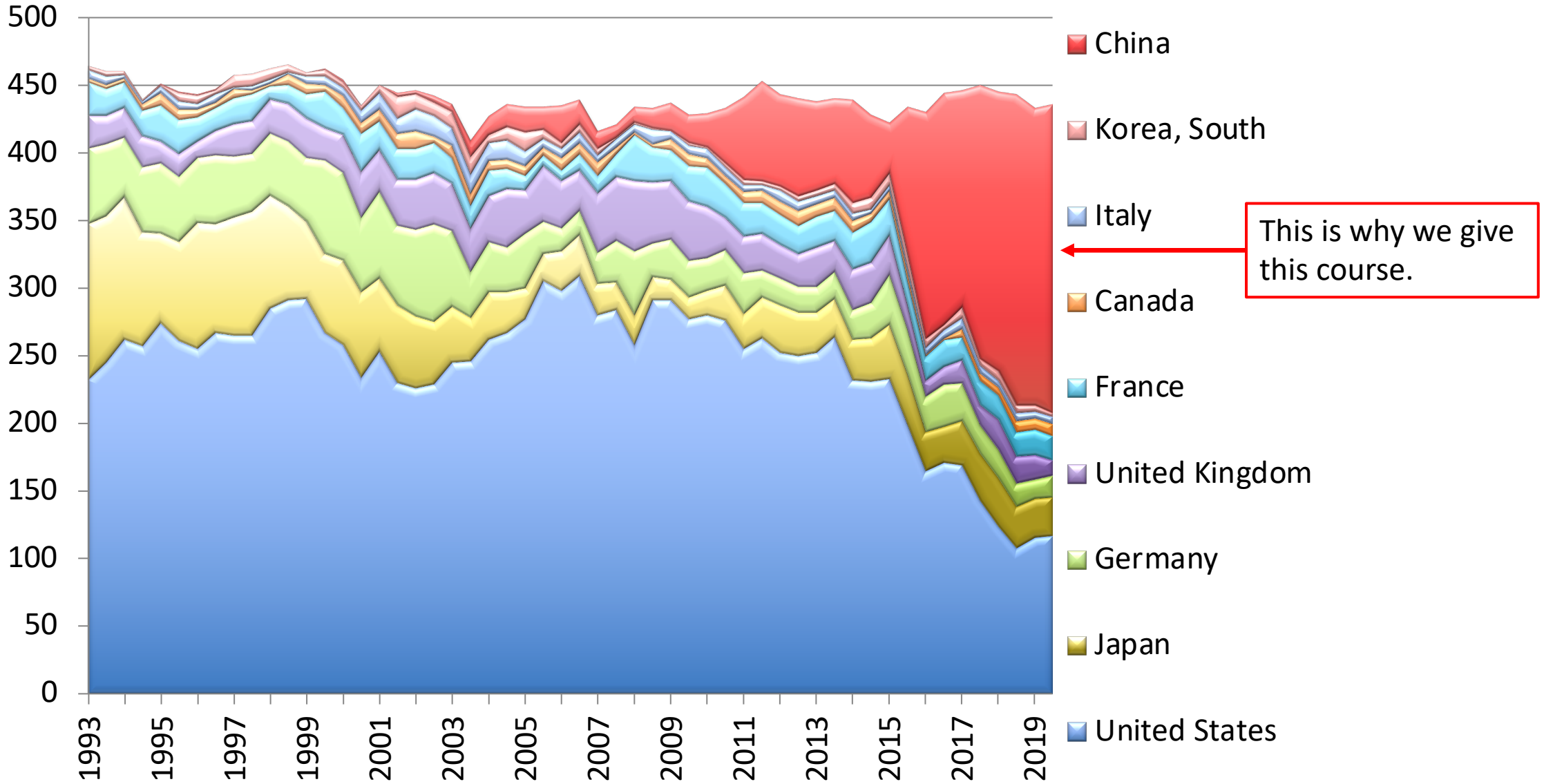
Each node has
- 2 Xeon Gold 6140 CPU
- 384 GB Memory

The system has
- 815 compute nodes
- Intel Omni-path network with 100 Gbps

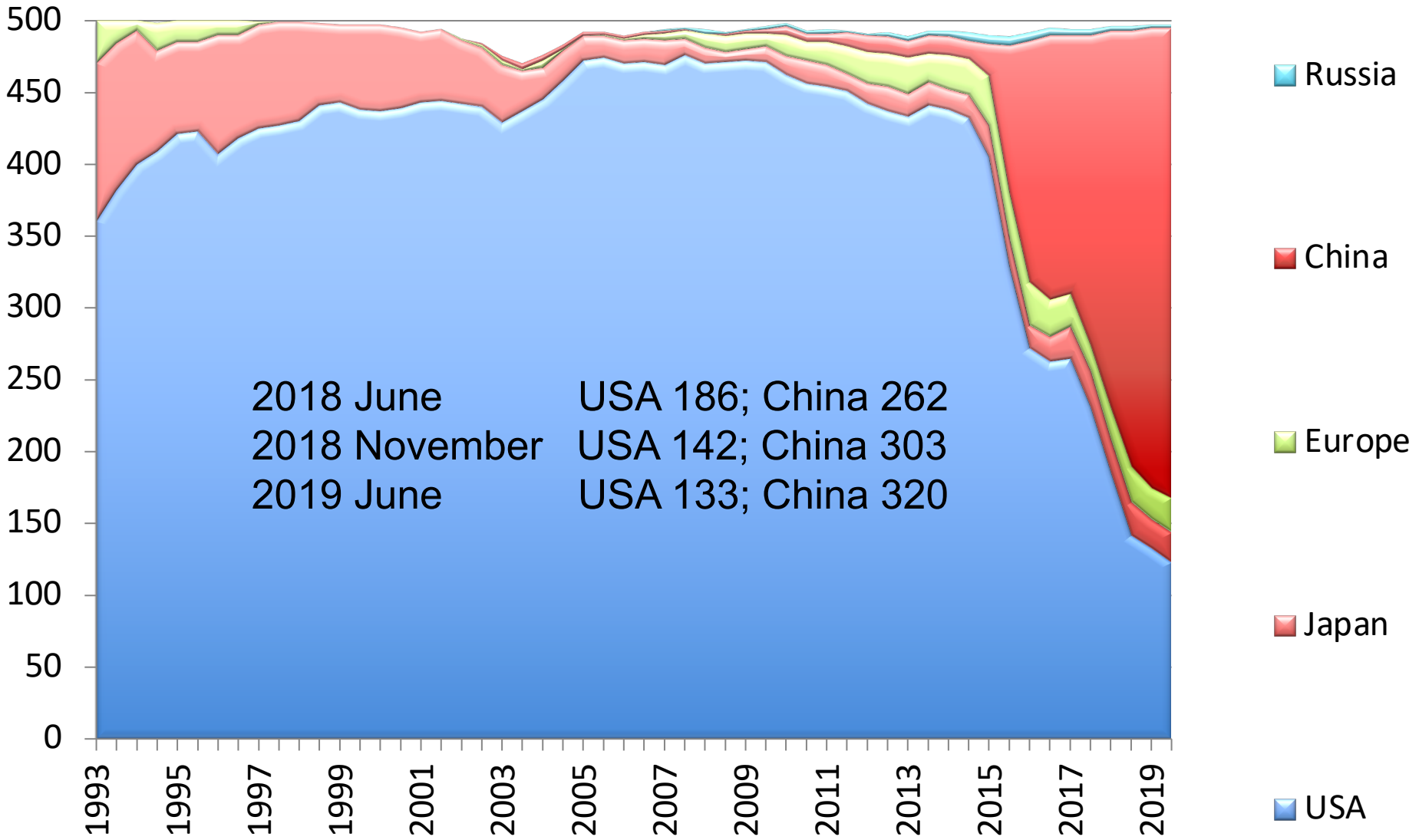
Performance over time



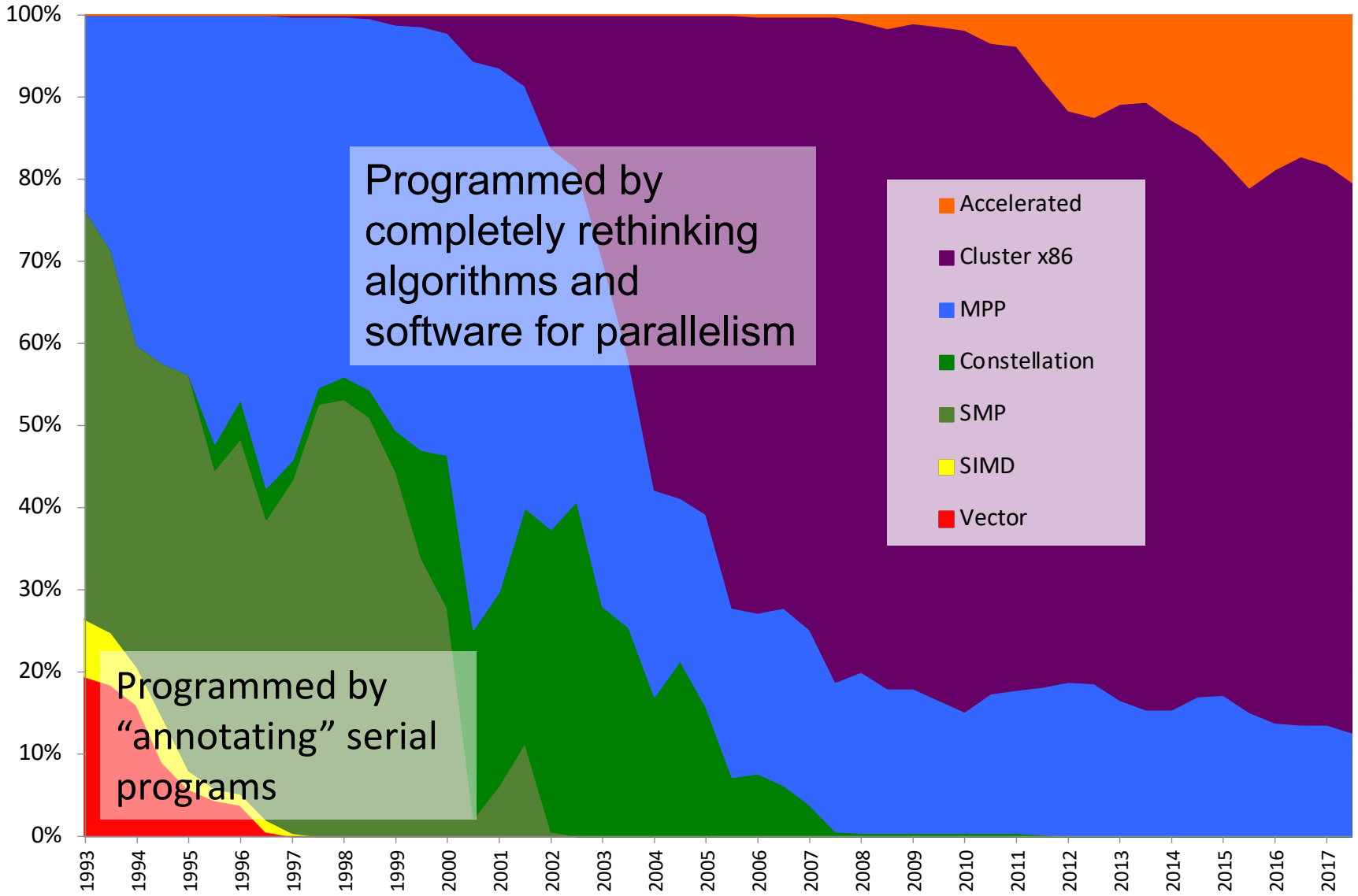
Countries


500
450
400
350
300
250
200
150
100
50
0
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
China
Korea, South
Italy
Canada
France
United Kingdom
Germany
Japan
United States
This is why we give this course.

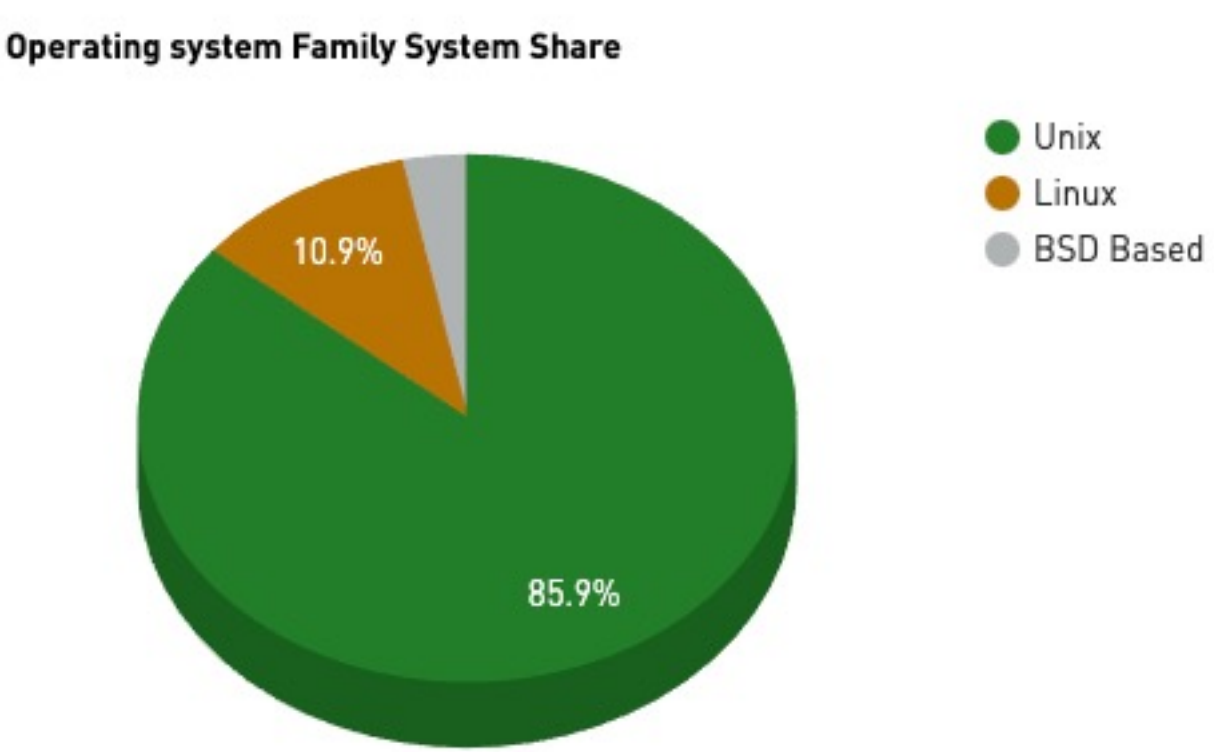
Producers


500
450
400
350
300
250
200
150
100
50
0
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
2018 June USA 186; China 262
2018 November USA 142; China 303
2019 June USA 133; China 320
Russia
China
Europe
Japan
USA

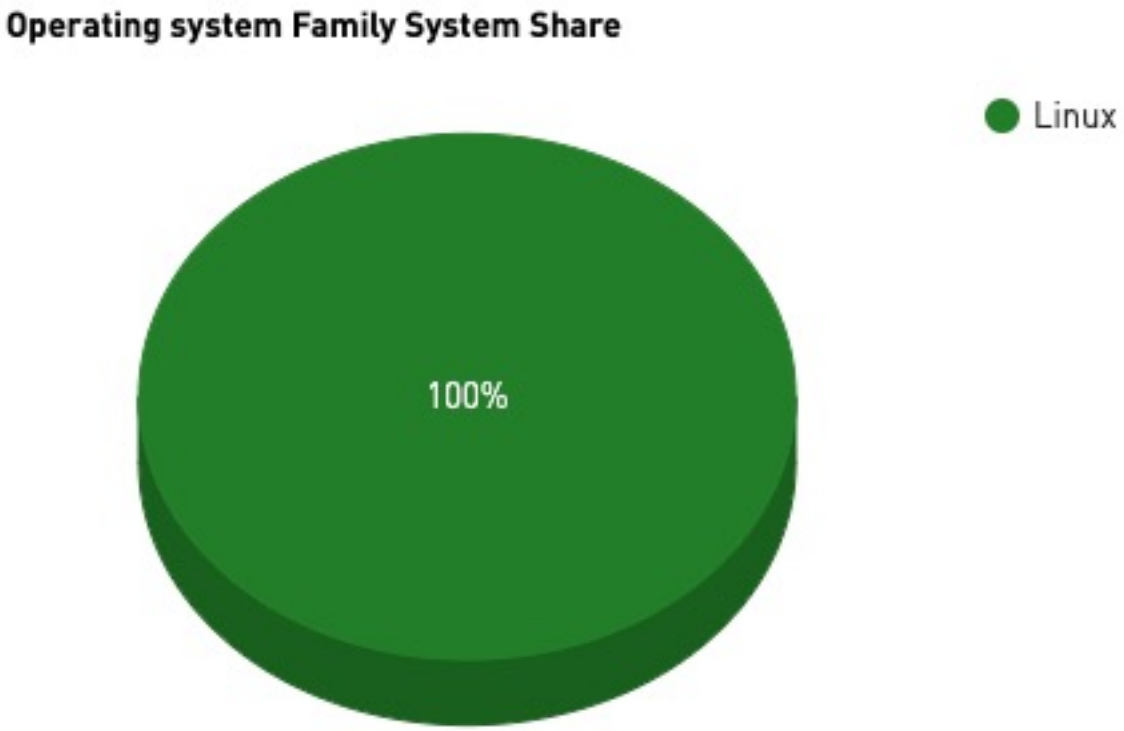
From vector supercomputers to massively parallel accelerator systems



Operating system



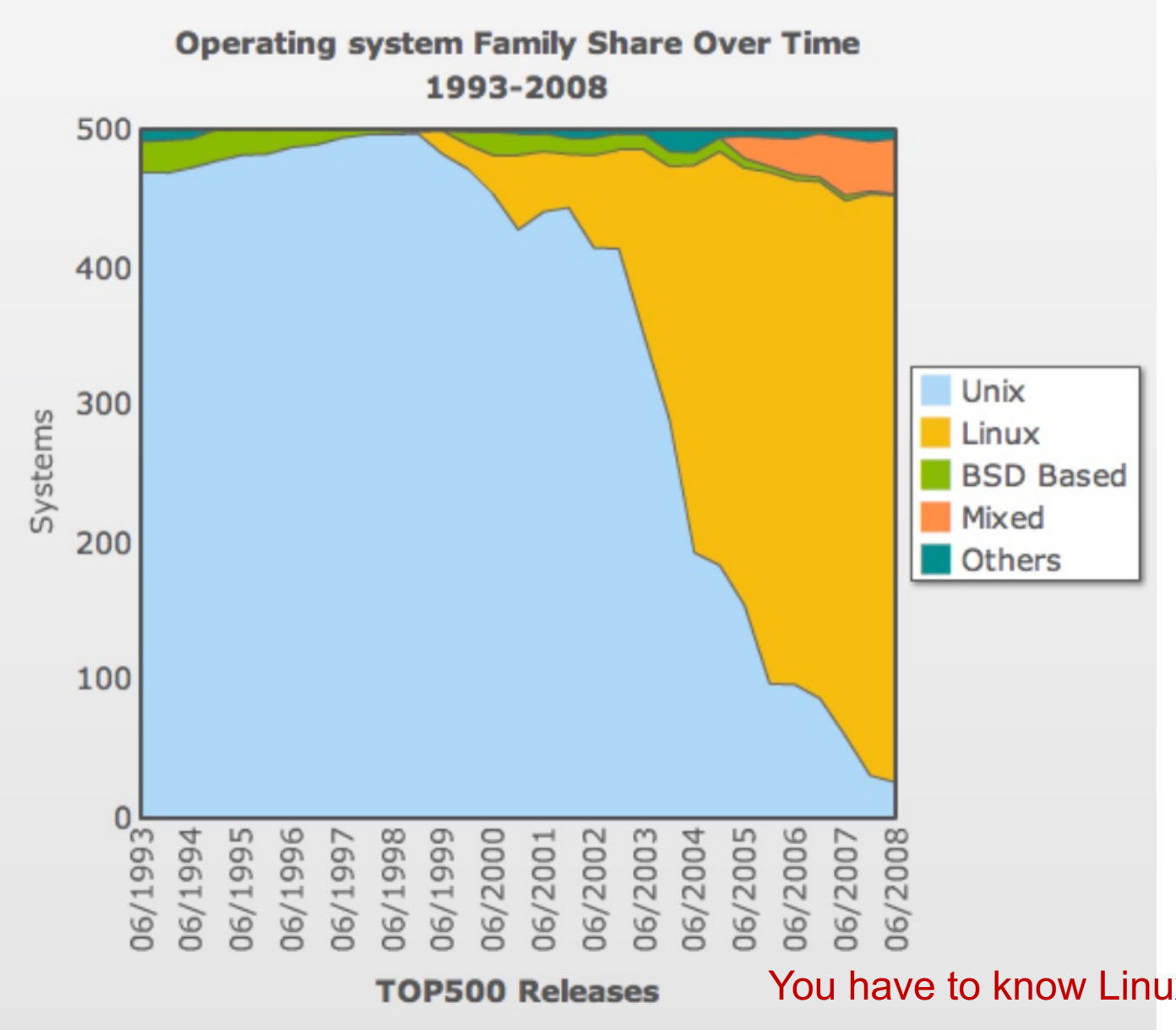
Nov. 2000



Nov. 2021

You have to know Linux for technical computing!

Operating system



You have to know Linux for technical computing!

Computer accounts

- Tai-Yi
  - We have very powerful academic system on campus
  - Linux system
  - Get an account on it from your advisor.

- You own machine
  - More convenient for development
  - You need a Linux installed on it
  - Mac system is fine.

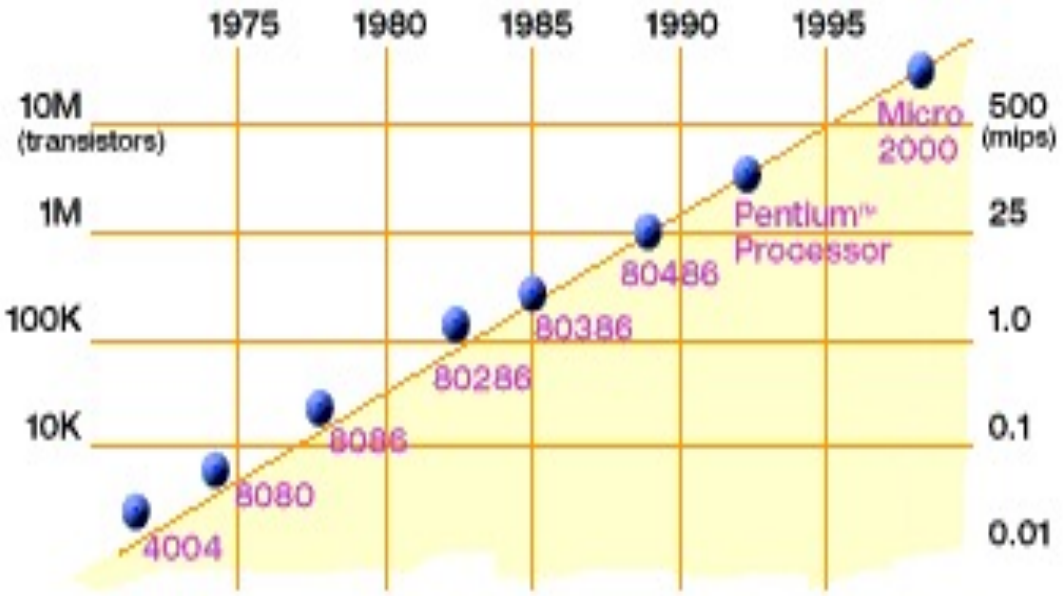
Homework 0:
Install Linux on your computer, and get access to TaiYi.

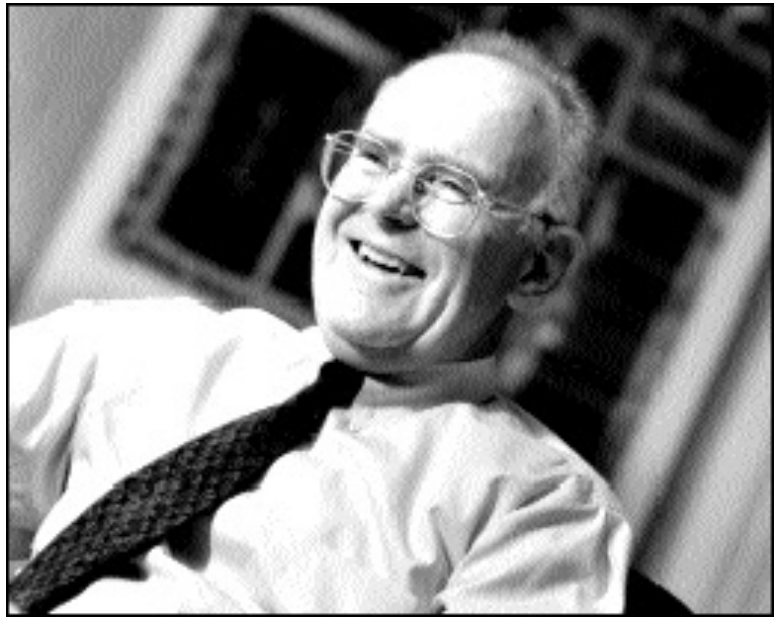
Why all computers are parallel computers (since 2005)

Tunnel vision by experts

- “I think there is a world market for maybe five computers.”
  - Thomas Watson, chairman of IBM, 1943.
- “There is no reason for any individual to have a computer in their home”
  - Ken Olson, president and founder of Digital Equipment Corporation, 1977.
- “640K [of memory] ought to be enough for anybody.”
  - Bill Gates, chairman of Microsoft,1981.
- “On several recent occasions, I have been asked whether parallel computing will soon be relegated to the trash heap reserved for promising technologies that never quite make it.”
  - Ken Kennedy, CRPC Directory, 1994

Microprocessor capacity





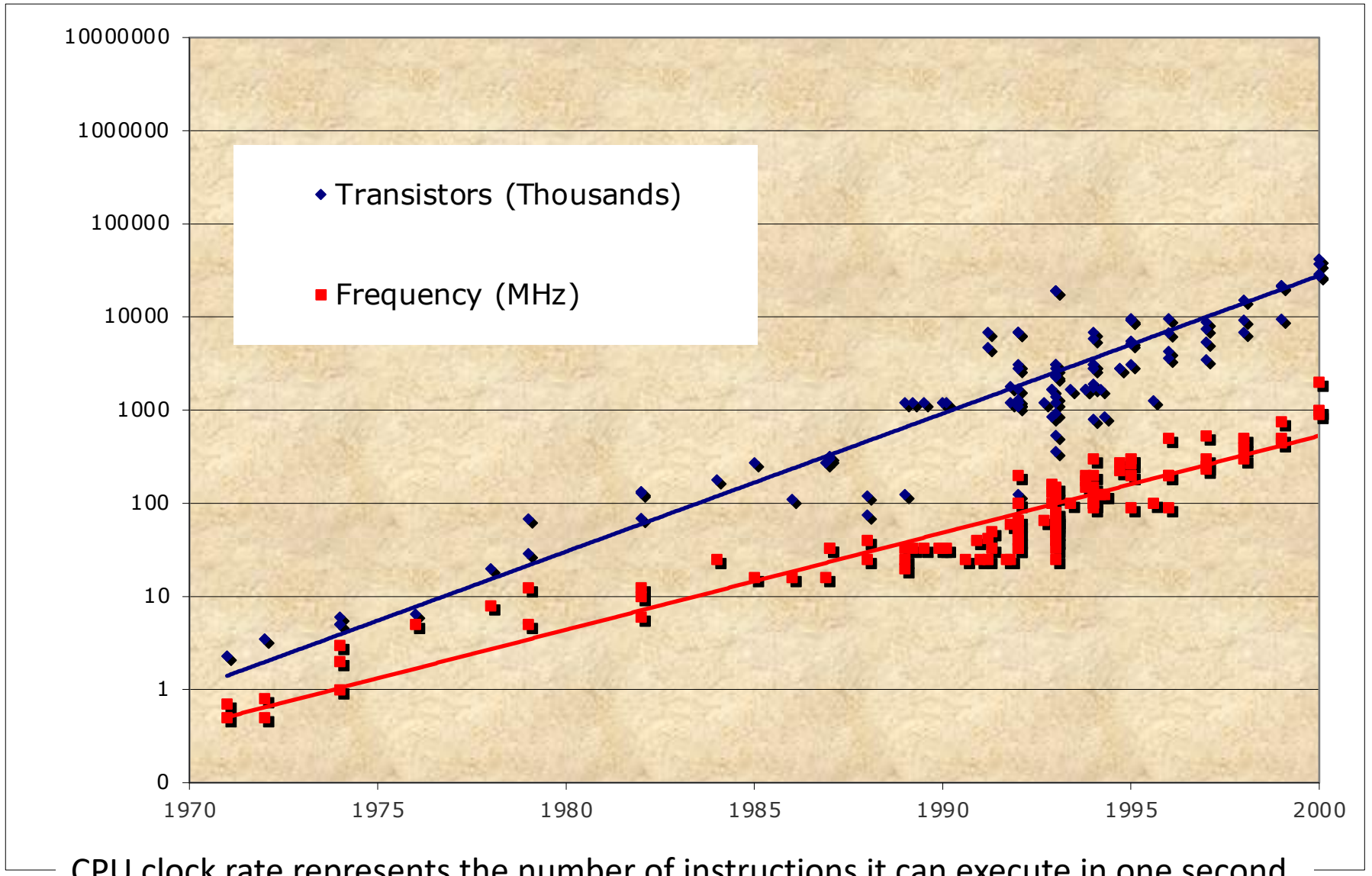
2X transistors/Chip Every 1.5 years

Called "Moore's Law"

Microprocessors have become smaller, denser, and more powerful.

Gordon Moore (co-founder of Intel) predicted in 1965 that the transistor density of semiconductor chips would double roughly every 18 months.

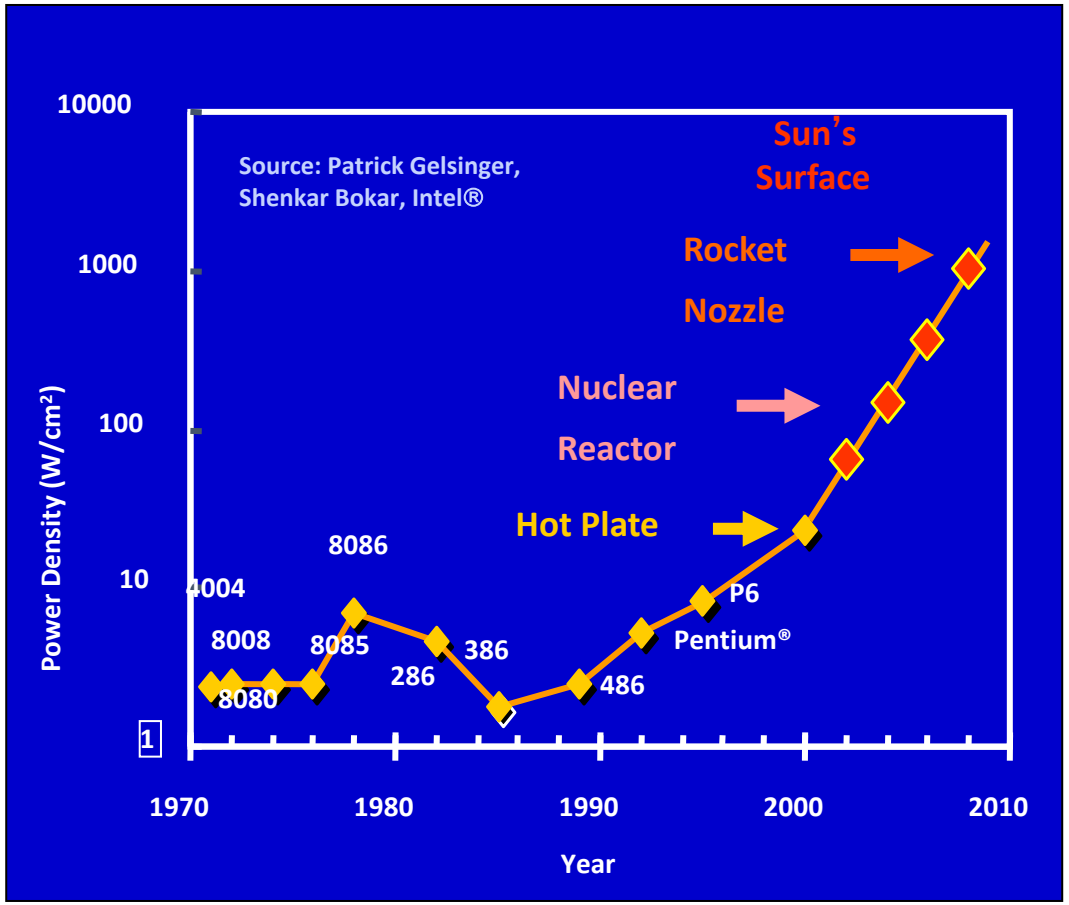
Microprocessor 1970 - 2000



CPU clock rate represents the number of instructions it can execute in one second.

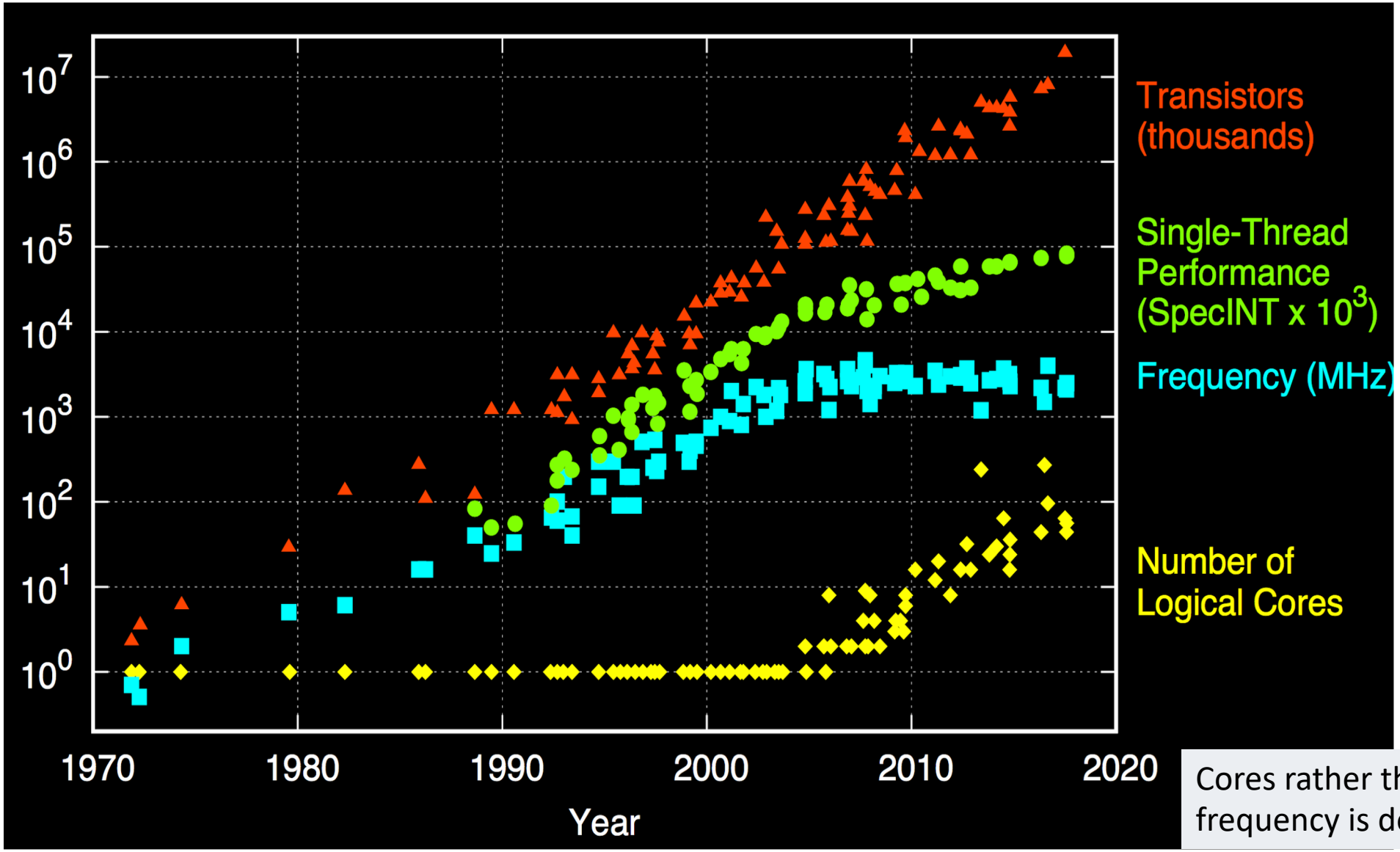
Microprocessor 1970 - 2000

- Concurrent systems are more power efficient
  - Dynamic power is proportional to V^2fC
  - Increasing frequency (f) also increases supply voltage (V) → cubic effect
  - Increasing cores increases capacitance (C) but only linearly
  - Save power by lowering clock speed



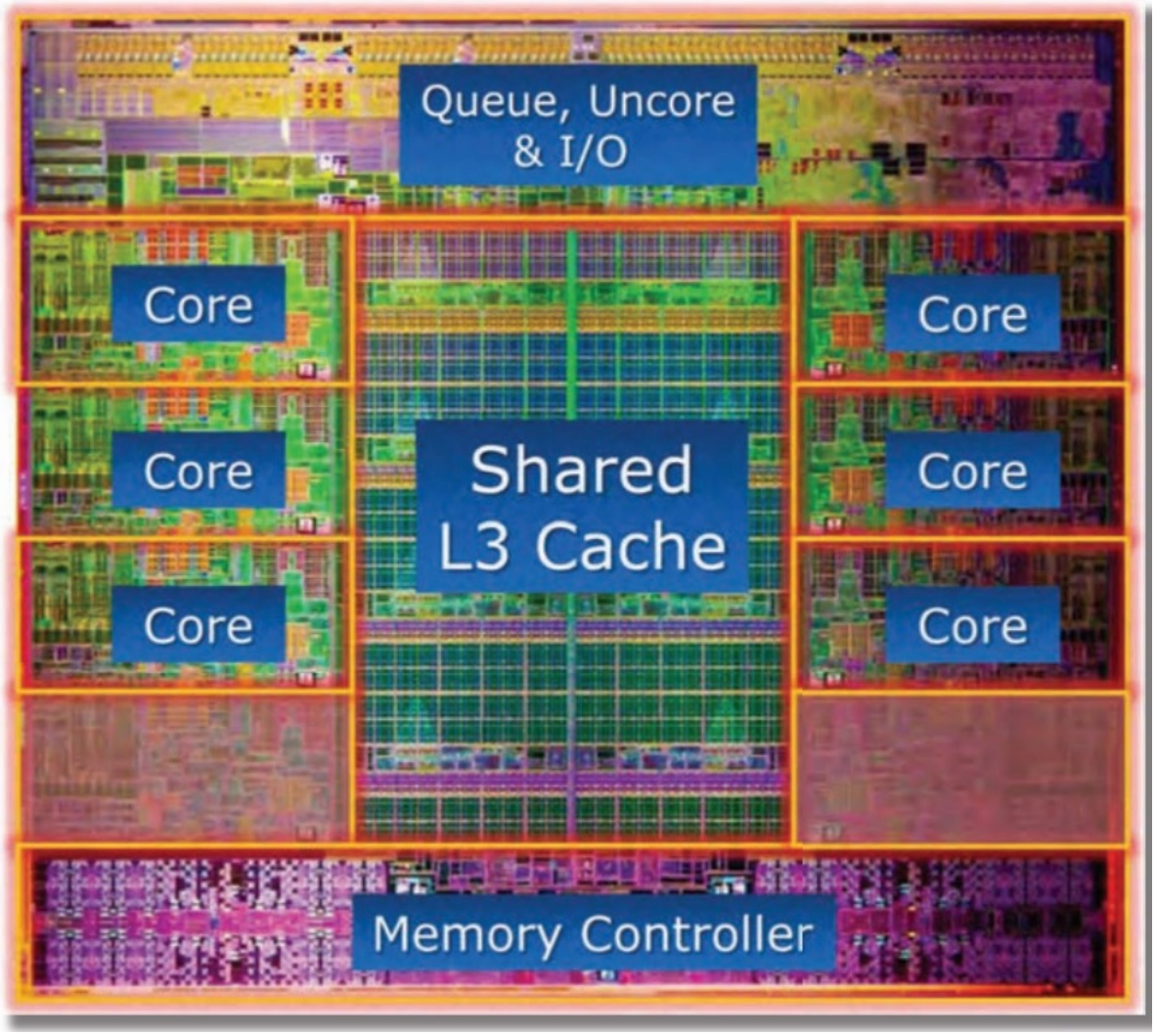
- High performance serial processors waste power
- More transistors, but not faster serial processors

Parallelelism



Cores rather than clock frequency is doubling

Multicore architecture



- Intel Core i7 Sandy Bridge E 2011
  - 6 cores
  - 3.3 GHz
  - 15 MB L3 Cache

To scale performance, manufacturer put many processing cores on the microprocessor chip.

Each generation of Moore's law potentially doubles the number of cores.

Software must be adapted to utilize this hardware efficiently.

Moore's Law reinterpreted

- Number of cores per chip can double every two years
- Clock speed will not increase (possibly decrease)
- Need to deal with systems with millions of concurrent threads
- Need to deal with inter-chip parallelism as well as intra-chip parallelism
- But Moore's Law is not forever… industry consortium predicts end in 2021

You need to know how to write software

Gordon Bell Prizes

Established in 1987 with a cash award of $10,000 (since 2011), funded by Gordon Bell, a pioneer in HPC.
For innovation in applying *HPC to applications in science, engineering, and data analytics*.

In 2016, 2020, and 2021, Gordon Bell prize were awarded to research teams from China.

Gordon Bell Prizes

The 2016 award goes to Chao Yang, et al. They simulated weather at 0.5 kilometer resolution with 770 billion unknowns.

Calculations were done on TaiHu Light with 10.4 million processors at a rate of 7.95 Pflops/s, versus 93 Pflops/s for Linpack.

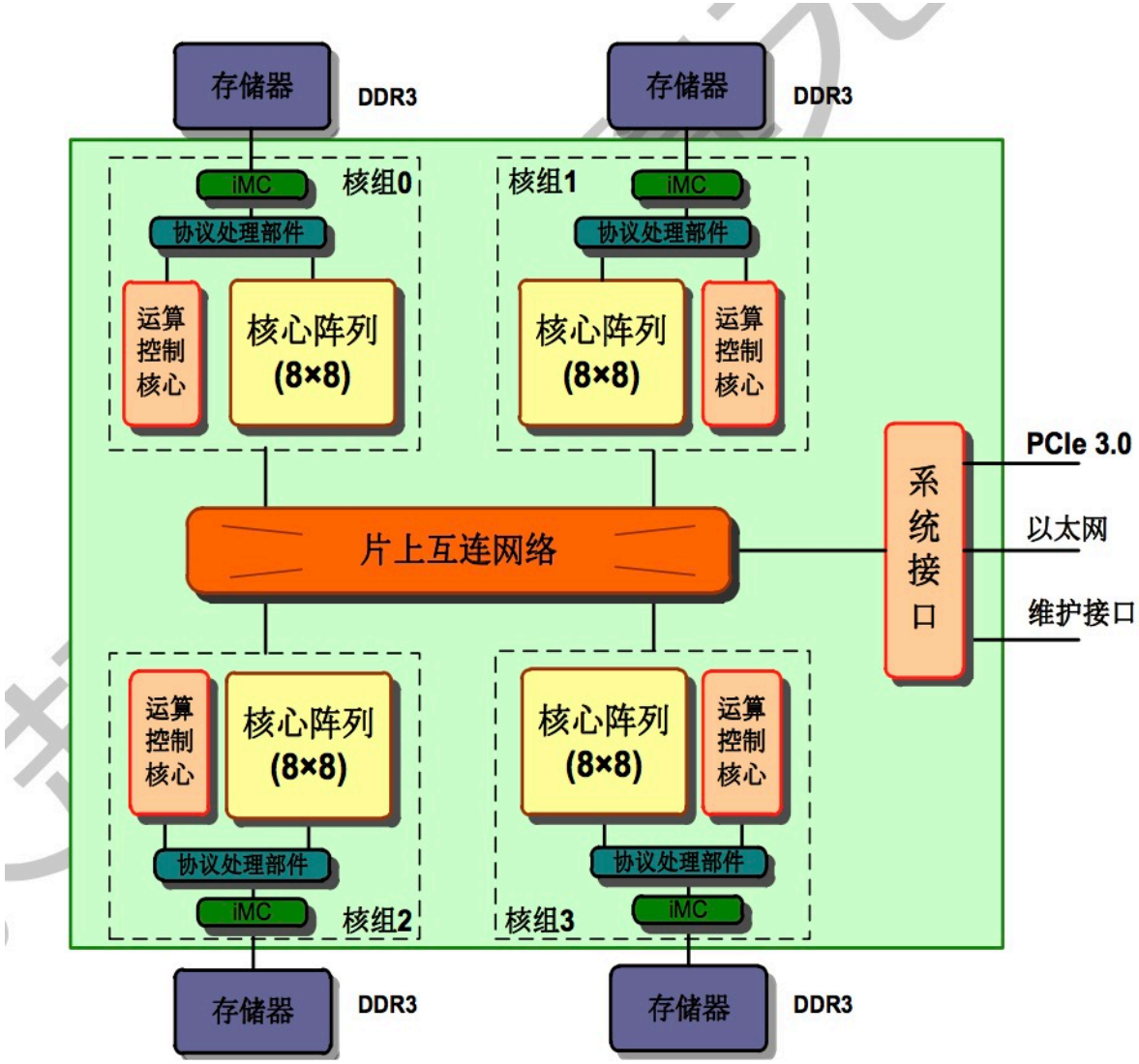
Gordon Bell Prizes

申威26010众核处理器：

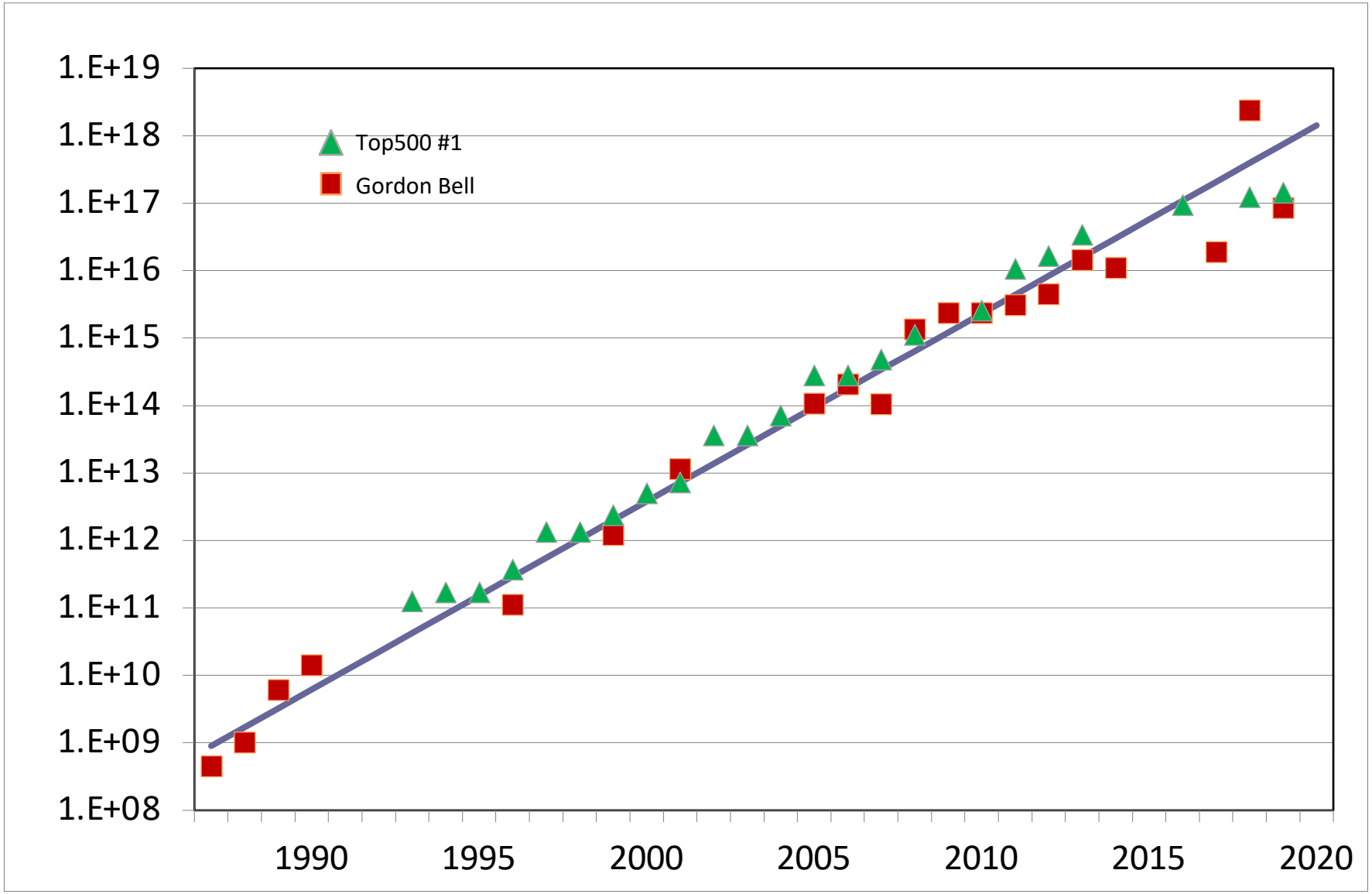
包括四个核组；
每个核组包括1个主核和64个从核；
每个核组8GB内存。

40960个众核处理器。

You need to know (1) how to program on a processor board; (2) how to coordinate 40960 processor boards.

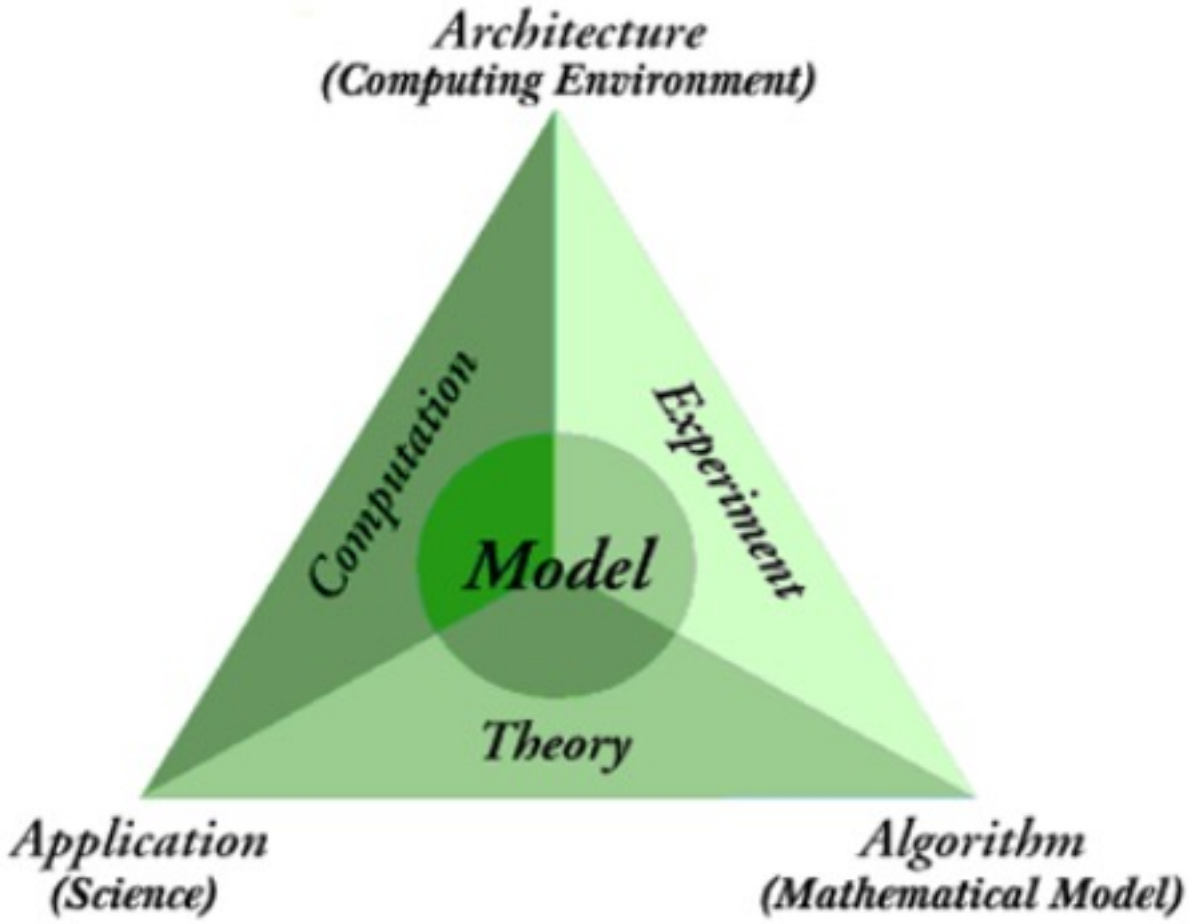


Gordon Bell Prizes



Science and Engineering Problems using HPC

The third pillar



The third pillar

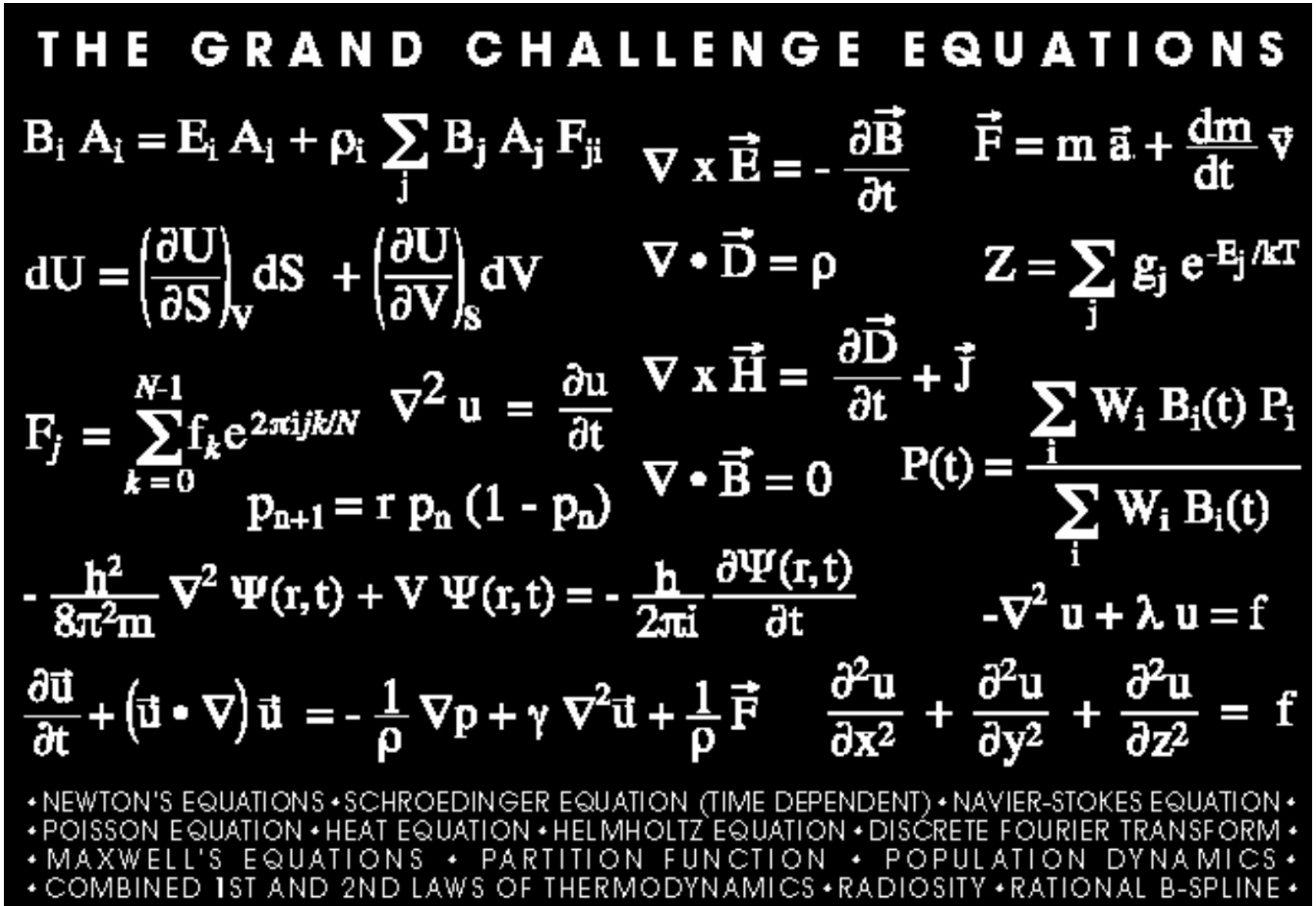
- In science, we use mathematics to understand physical systems
- Different fields of science explore different domains of the universe, and have their own sets of equations, encapsulated in theories
- Determining the theories an governing equation requires observation or experimentation and testing hypotheses

The third pillar

- Why should we care about scientific computing?
  - Computational research has emerged to complement experimental methods in basic research, design, optimization, and discovery in all facets of engineering and science.
  - In certain areas, computational simulations are the only possible approach to analyze a problem:
    experiments may be cost prohibitive;
    experiments may be impossible.

- Simulation capabilities rely heavily on the underlying computer power (e.g. the amount of memory, total compute processors, and processor performance).

The third pillar



The third pillar

- A definition: the efficient computation of constructive methods in applied mathematics
  - Applied math: getting results out of application areas;
  - Numerical analysis: results need to be correctly and efficiently computed;
  - Computing: the algorithms need to be implemeneted on modern hardware.

Simulation in Science and Engineering

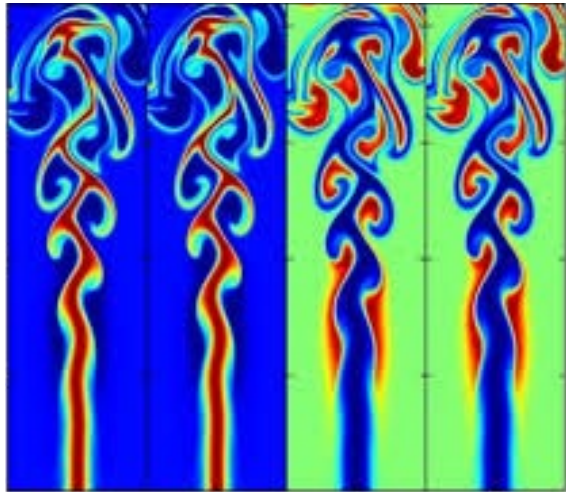
High performance computing (HPC) simulation to understand things that are:

- too big
- too small
- too fast
- too slow
- too expensive or
- too dangerous

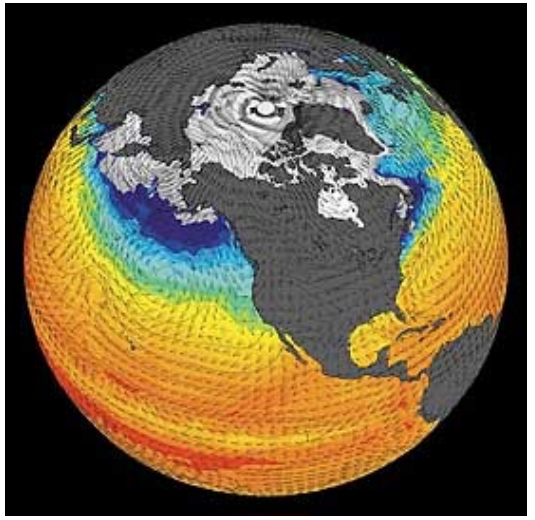
for experiments

Understanding the universe

Proteins and diseases



Energy-efficient jet engines

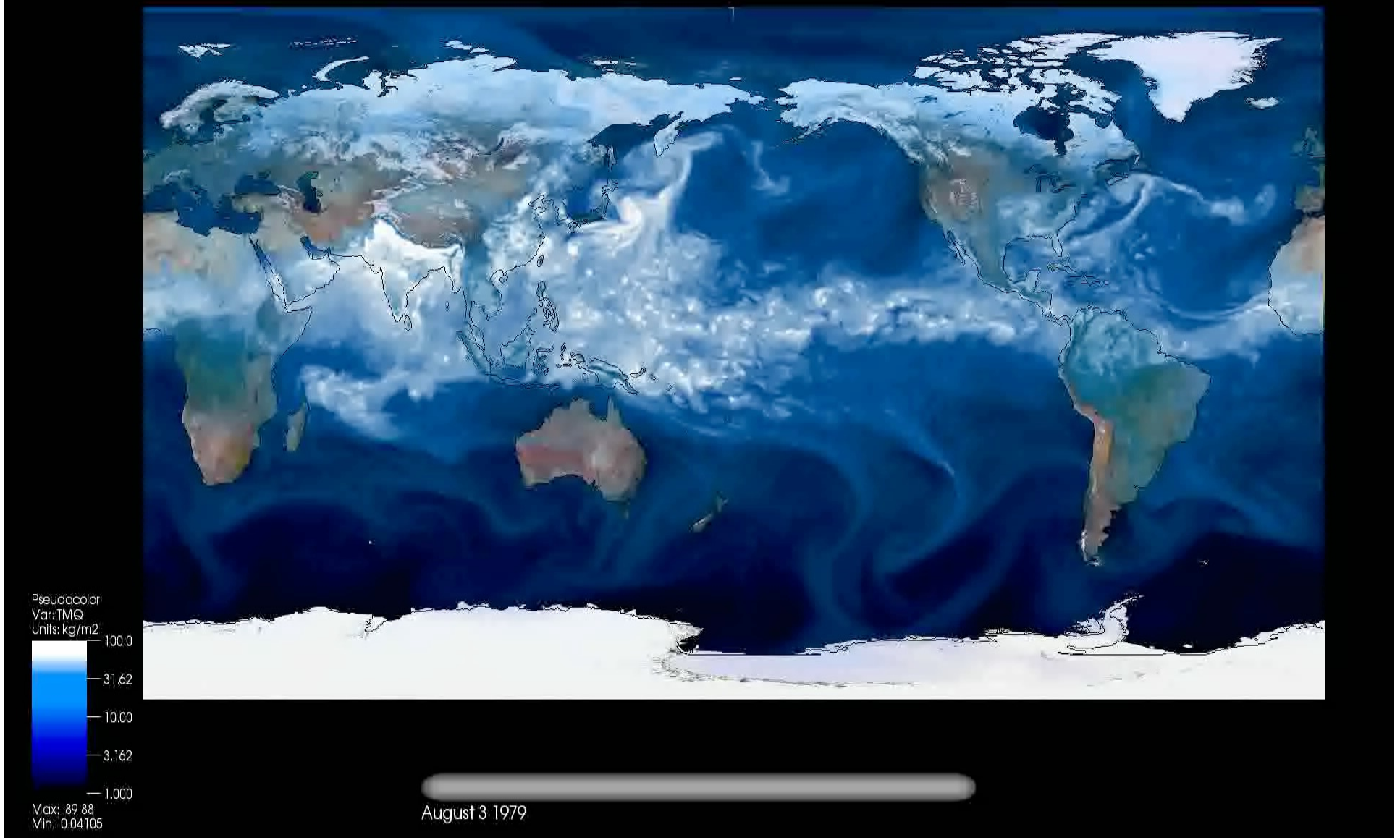


Climate change

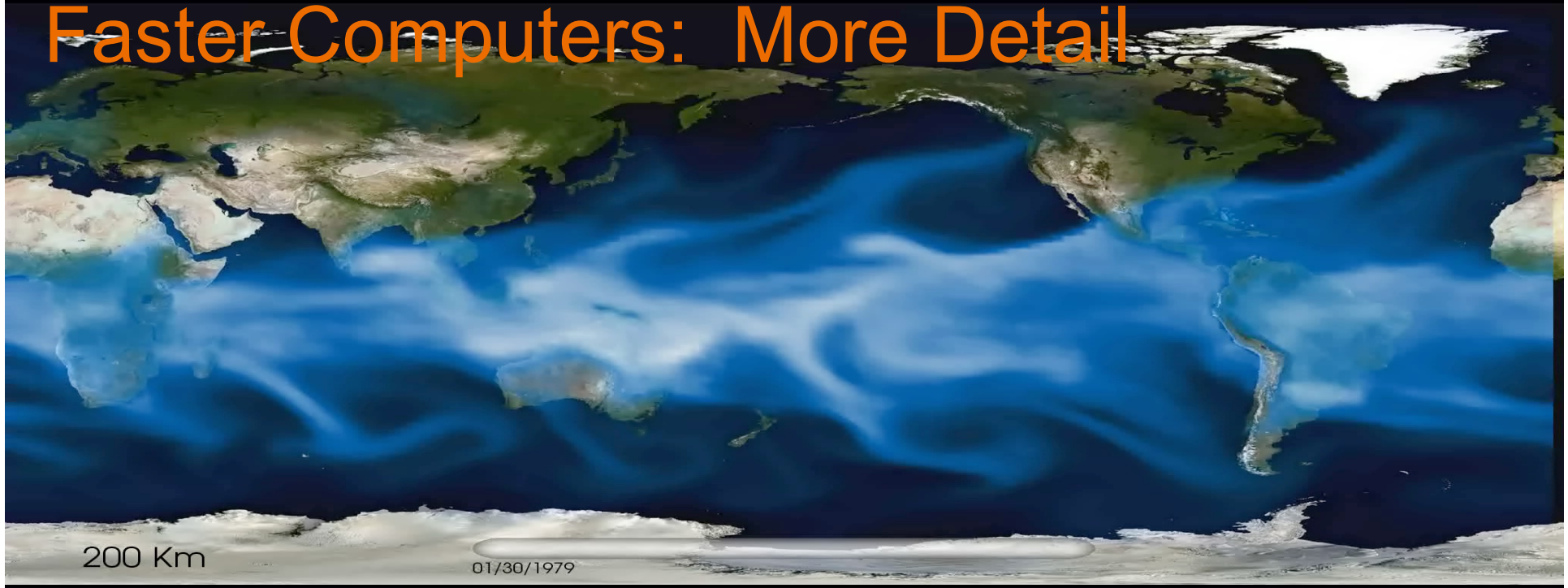
Need: Weather Modeling and Forecasting

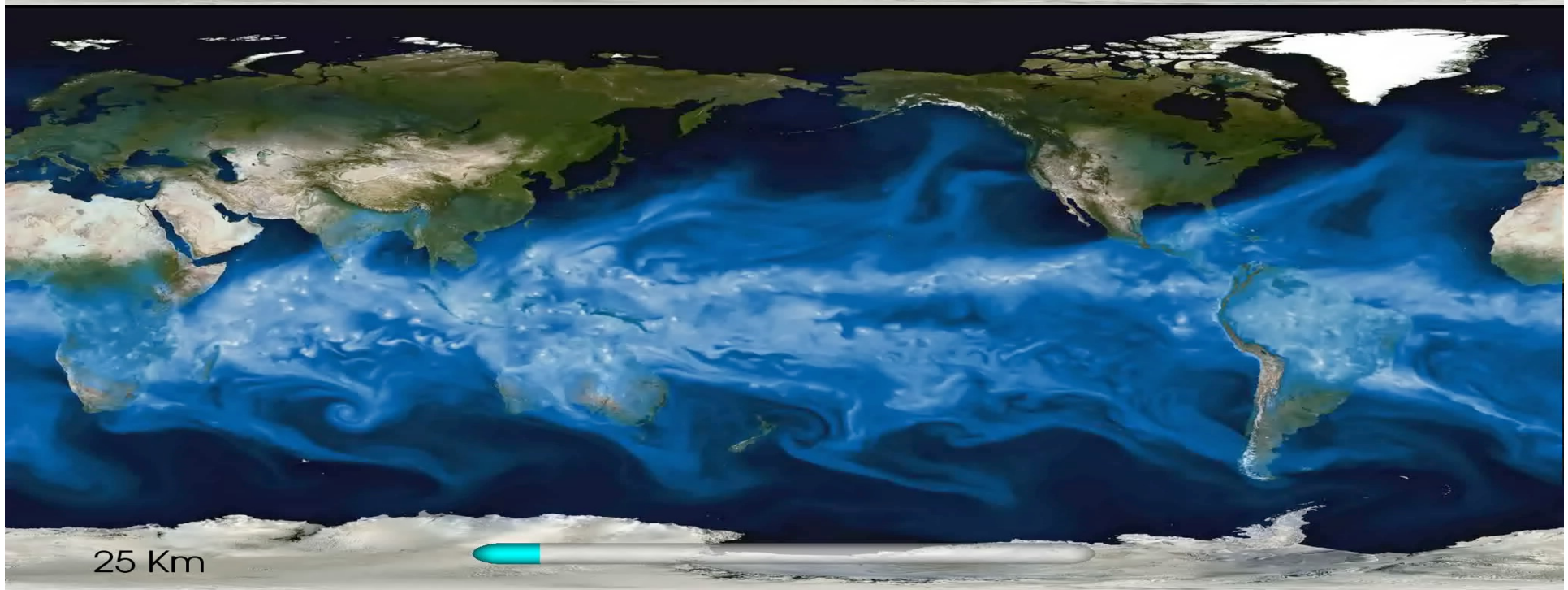
- For modeling a hurricane region:
  - Assume region of interest is 1000 X 1000 miles, with height of 10 miles.
  - Partition into segments of 0.1 x 0.1 x 0.1 miles: 10^10 grid points
  - Simulate 2 days, with 30-minute timesteps: 100 total timesteps
- Assume the computations at each grid point require 100 instructions.
  - A single timestep then requires 10^12 instructions.
  - For two days we need 10^14 instructions
  - For serial computer with 10^8 instructions/sec, this takes 10^6 seconds (10 days!) to predict next 2 days!!
- THIS REQUIRES PARALLELISM FOR PERFORMANCE TO PREDICT
  - Also requires lots of memory which implies parallelism
- Il major weather forecast centers (US, Europe, Asia) have supercomputers with 1000s of processors.

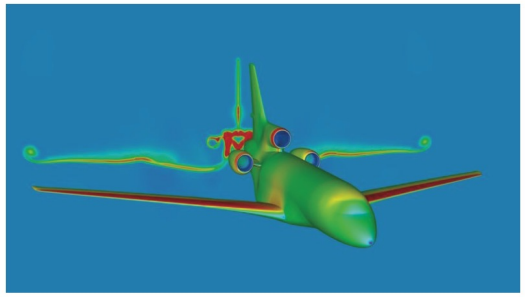
Simulations Show the Effects of Climate Change in Hurricanes



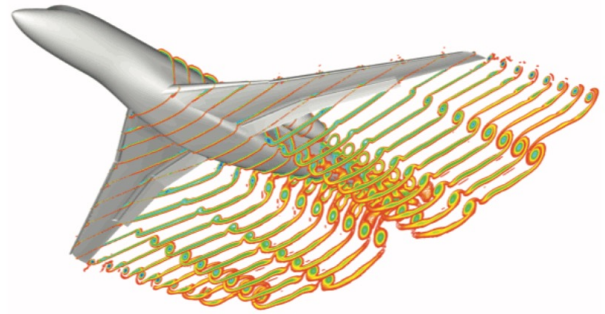
Faster Computers: More Detail

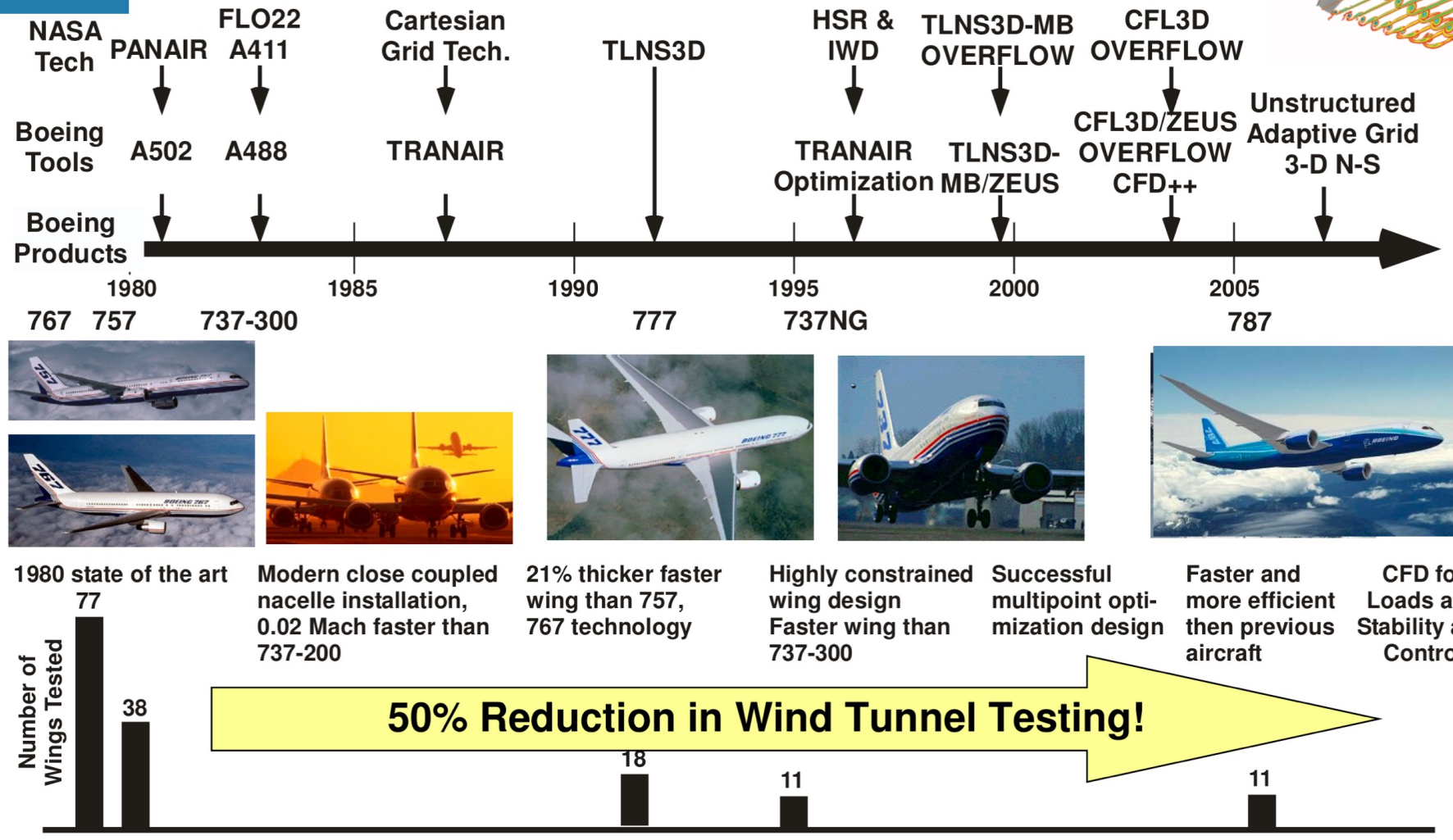




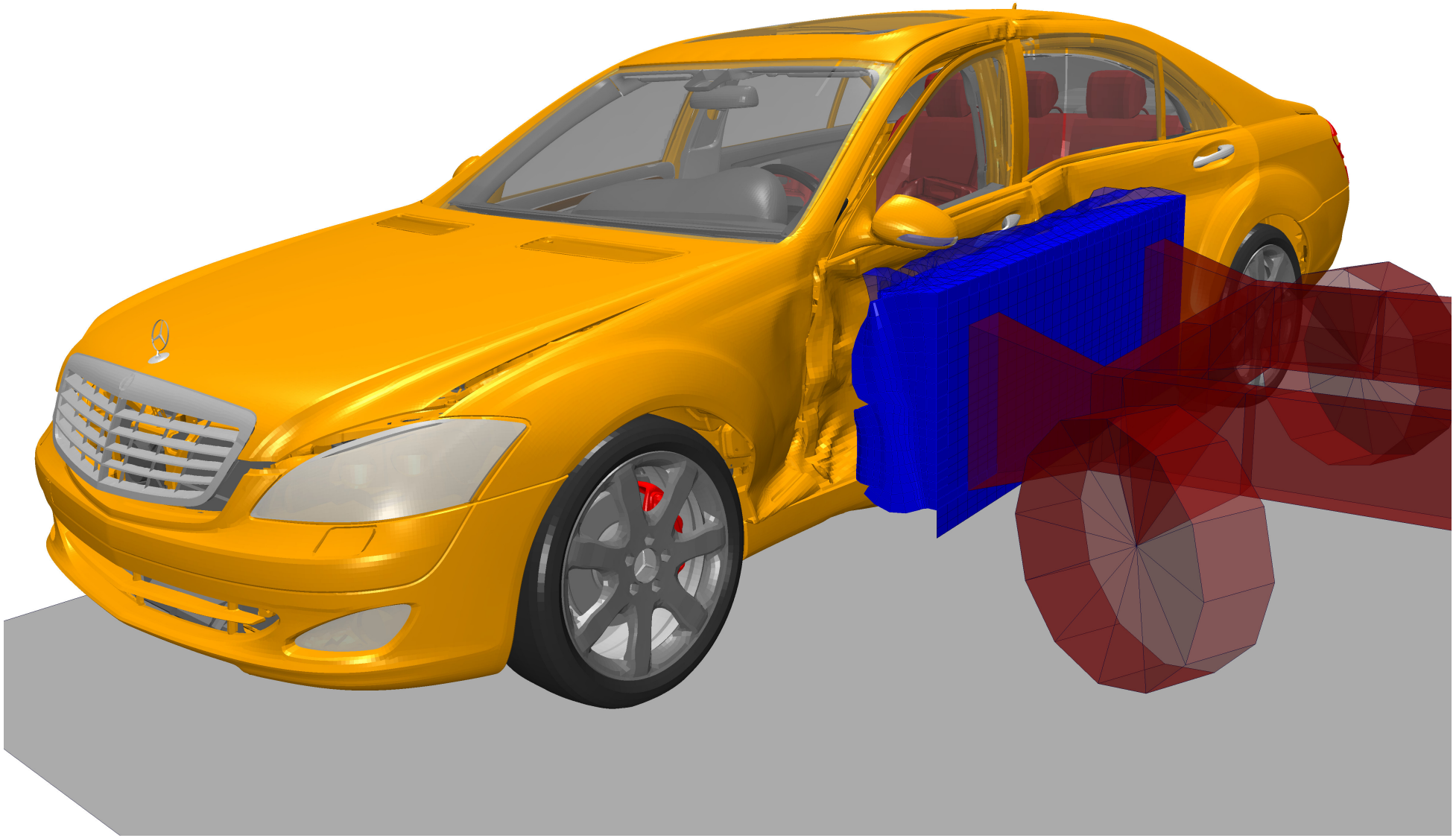


Some applications



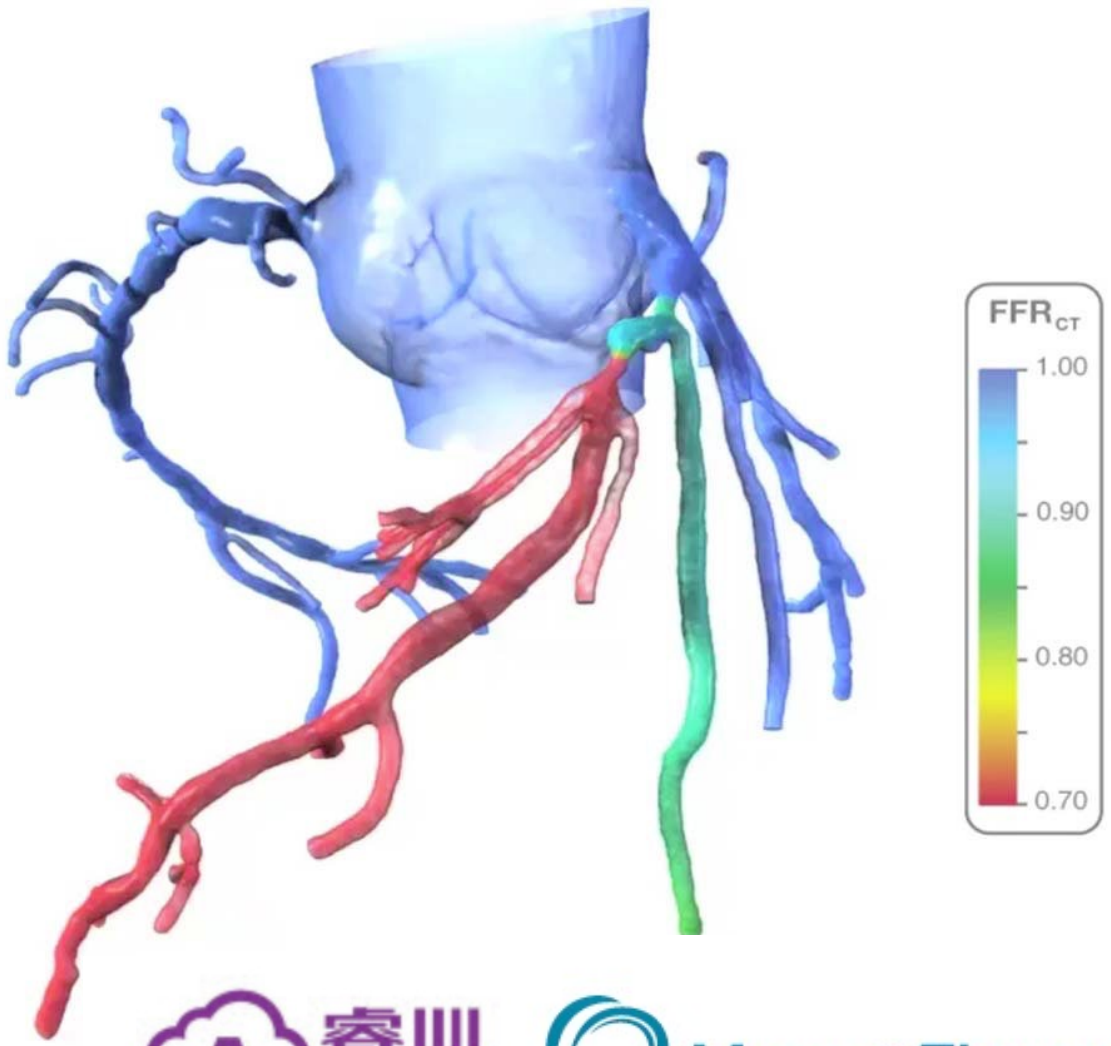


Some applications



Some applications

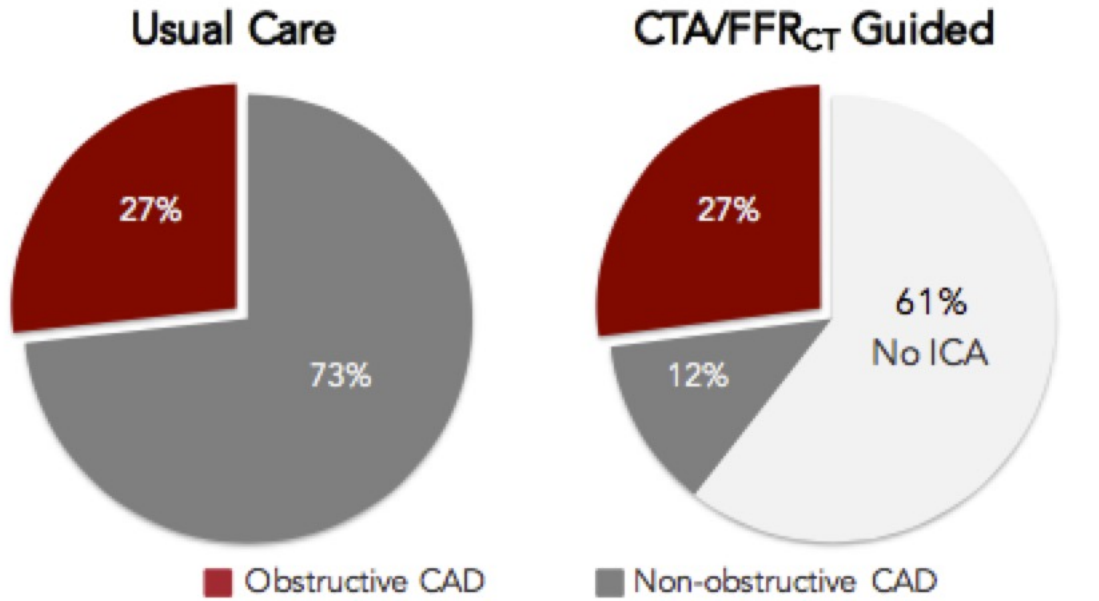
Some applications



睿心 RAYSIGHT
Healthcare • Intelligent • Accurate

HeartFlow®

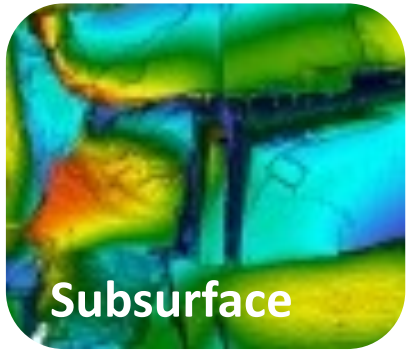
Invasive Catheterization (ICA) with No Obstructive Disease

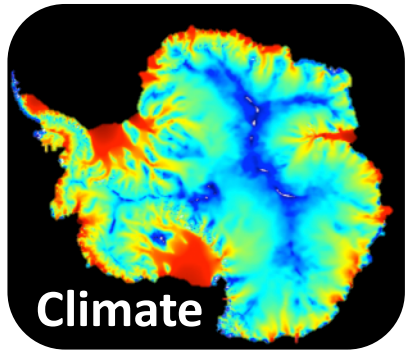


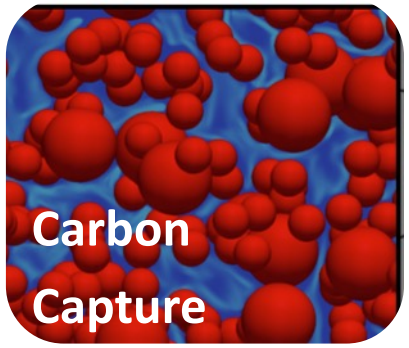
83% reduction of ICAs that found no obstructive CAD
No adverse clinical events in patients in whom ICA was cancelled.

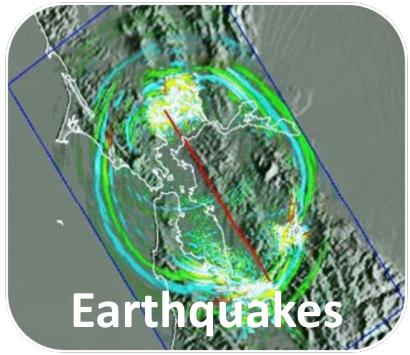
P.S. Douglas, et al. European Heart Journal 2015.

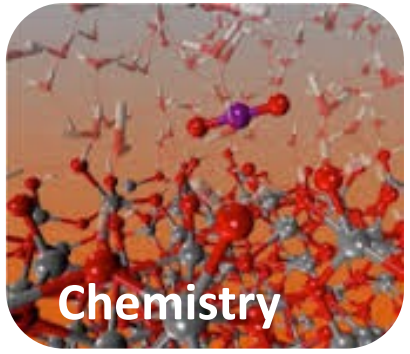
Some applications

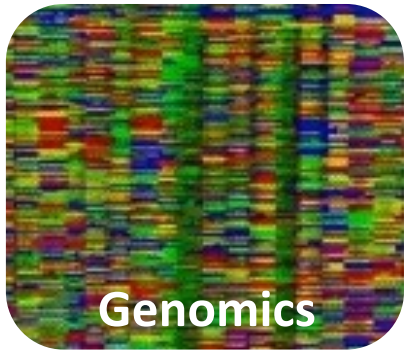


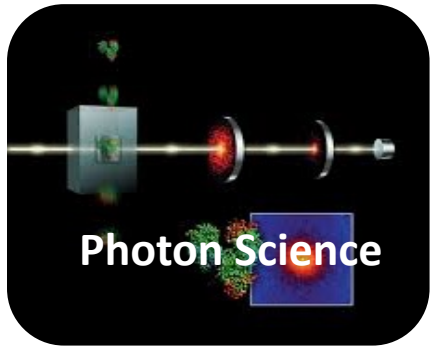












What is in this course?

Course goals

- UNIX exposure
  - Command line
  - Compilers
  - Libraries
- Some numerical background
  - Linear algebra
  - Numerical PDE
- Good practices of software development
  - Version control
  - Build systems
  - Debugging skills
- Parallel computing
  - MPI
  - PETSc

What you should get out of the course

Understanding of:

Part – 1

- Serial computer architecture
- Linux basics
- Version control of software (git)
- Compilers and cross-platform compiling tools (CMake)
- Debugging and code profiling

Part – 2

- Numerical analysis basics (linear algebra, PDEs, etc.)
- Machine learning basics

What you should get out of the course

Part – 3

- Engineering computing libraries
- Visualization of data
- I/O of data

Part – 4

- Parallel computer architecture
- Parallel computing basics
- MPI programming

Part – 5

- GPU programming (if time permits)

Methods and Practice

Practice is important.

You are (strongly) encouraged to bring your laptop to class.

If you do not have access to computers, please come talk to me immediately after this lecture.

There will be a little theory, which is not too hard.

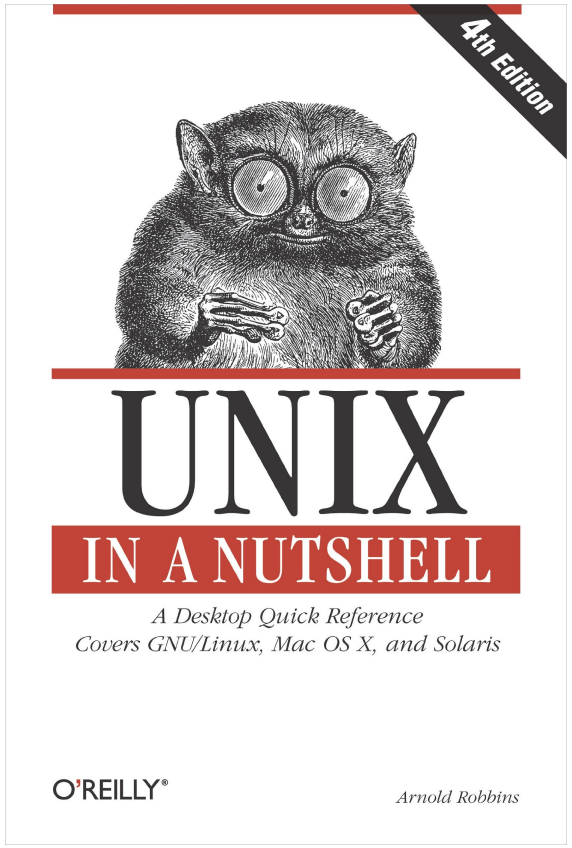
Slides will be shared with you on blackboard.

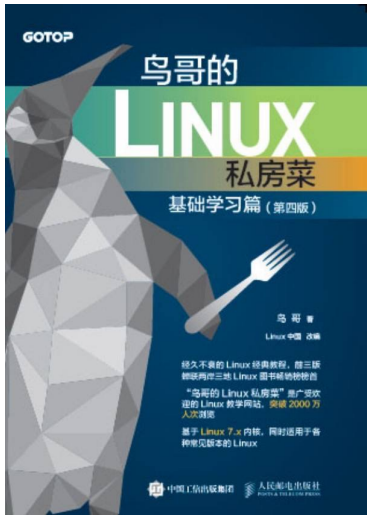
References

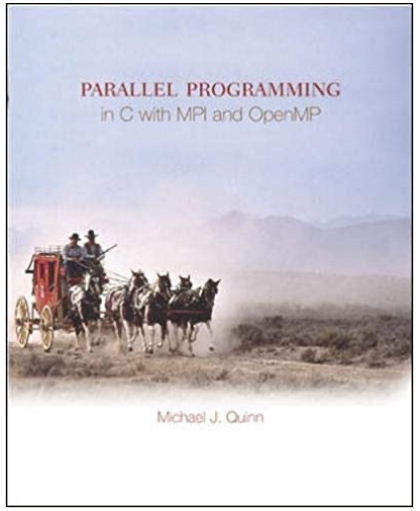
We recommend a few books for general reading.

Tutorials may be distributed in class on specific topics (Makefile, VTK, PETSc, etc.)

Slides will become available.

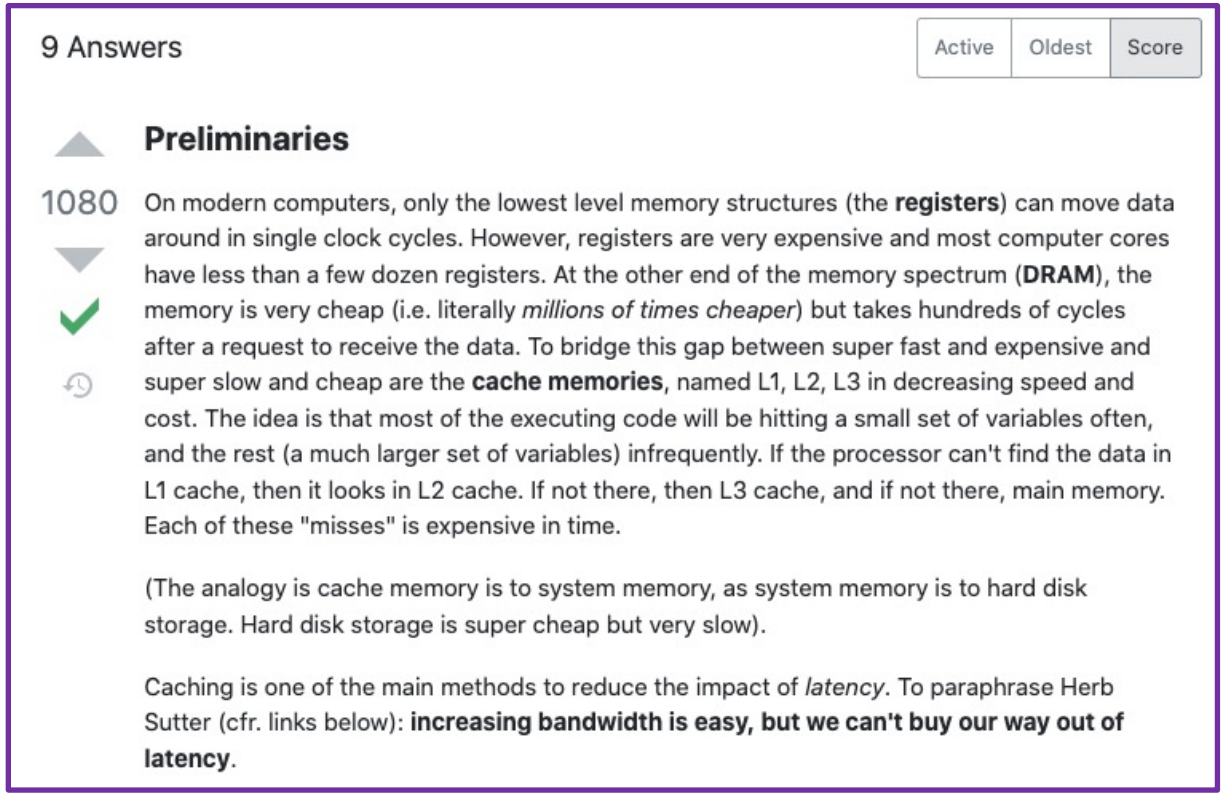






Some recommendations

- Use a good search engine and Stack Overflow to learn things



Grades

- The final grade will be based on

➢Homework (40%) (Around 8 assignments)
➢Mid-term exam (20%)
➢Projects (40%) Proposal due on April 27.

We have `practice' in the course name. There will be frequent assignments on practical computing and programming.

You need to know C/C++/Fortran.

If you use Fortran, come and discuss with me (I am not a Fortran programmer ☺).

Matlab/Python is not OK.

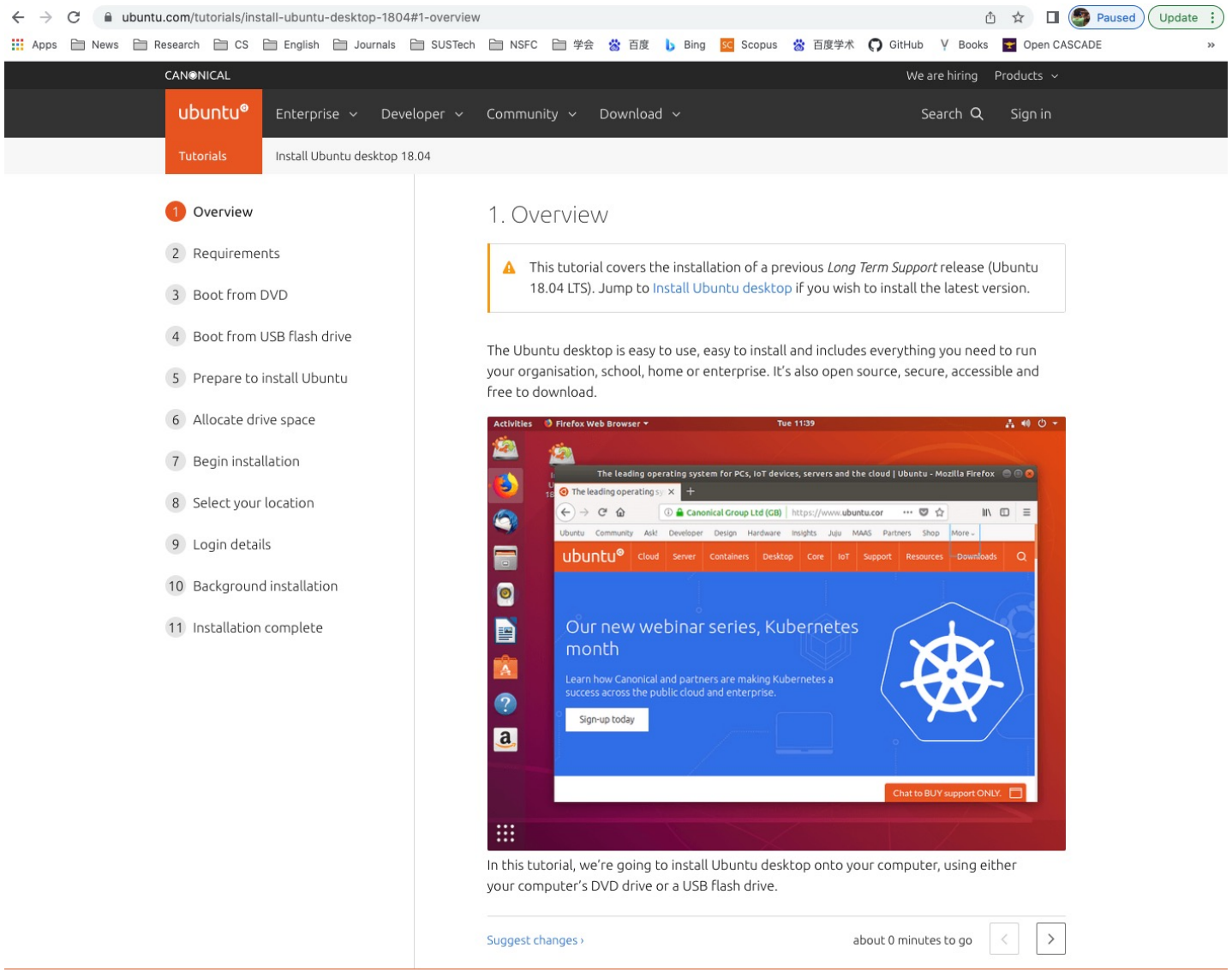
Some policies

- All codes and written work should be your own
  - Homework will be assigned on a regular basis (approximately once every two weeks).
  - Make your own code repositories private and do not share after the semester ends (will talk about git repository later).
  - We do NOT take late submission of HW.
  - It is allowed to revise your HW and get points back. (Do NOT abuse this.)

- The final project will be done in teams
  - We encourage teams of 2 to 3 students for final projects – exceptions by approval
  - Each one's contribution will be reflected on the Git commit history

Homework 0

- Install Linux Ubuntu 18 on your machine.
- Go to ubuntu.com/tutorials/install-ubuntu-desktop-1804
- Teaching assistants can help you get a USB install drive



Homework 0

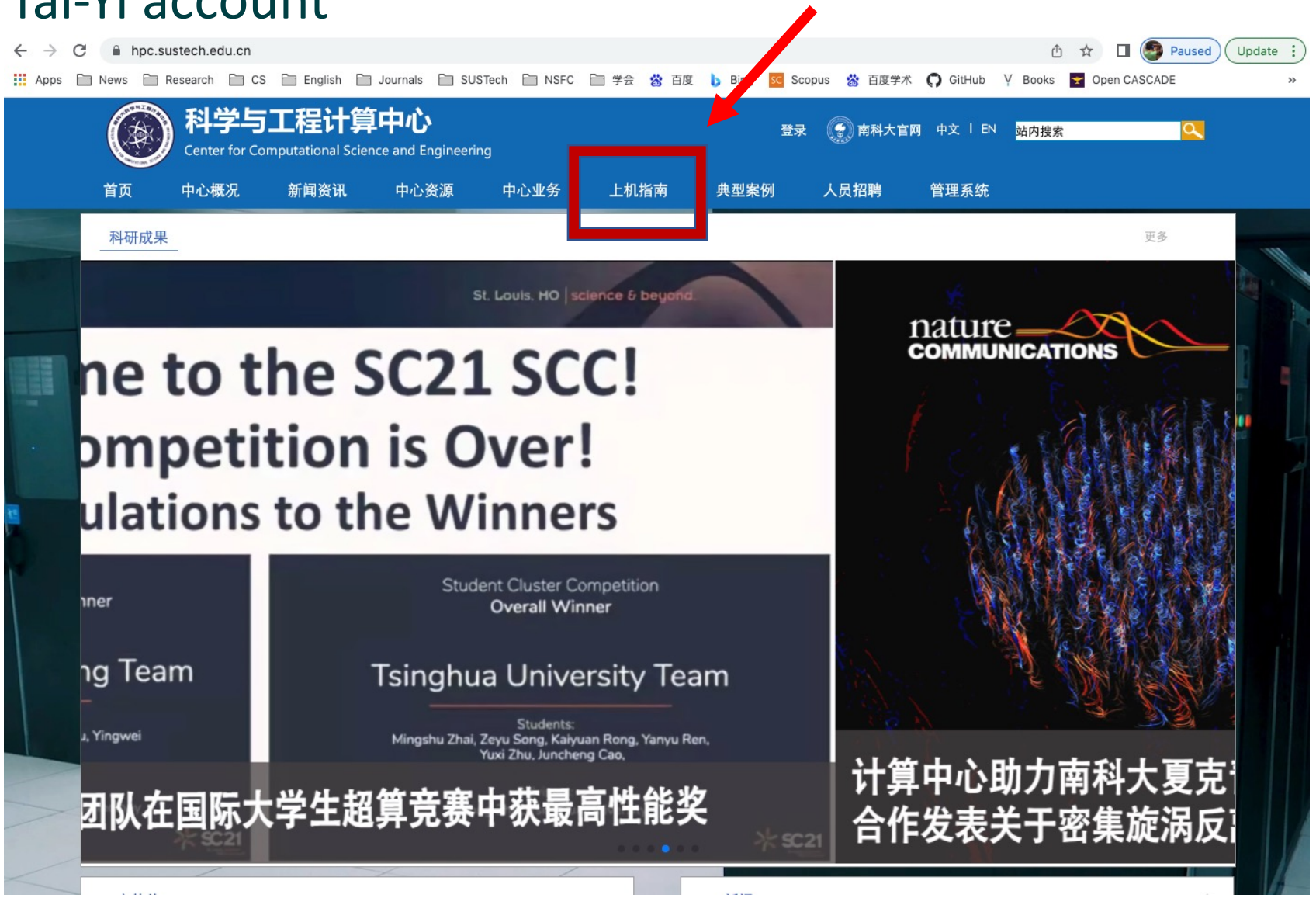
- You do not have to delete your Windows system

  https://www.itzgeek.com/how-tos/linux/ubuntu-how-tos/how-to-install-ubuntu-18-04-alongside-with-windows-10-or-8-in-dual-boot.html

- You need to prepare Windows for dual-boot
  - Extra partitions to allocate Linux ubuntu

Homework 0

- Get your Tai-Yi account



Homework 0

- Get your Tai-Yi account

- You need to have your advisor get an account first, and he/she will have to add you to his/her allocation

- Remote login by SSH

  ssh username@172.18.6.175

at your command line. username is your Tai-Yi account.

```
juliu::Kolmogorov { ~ }
-> ssh mae-liuj@172.18.6.175
mae-liuj@172.18.6.175's password:
```

- Due in two weeks.