UN5390: Scientific Computing I

Dr. Gowtham

Director of Research Computing, IT
Adj. Asst. Professor, Physics and ECE

EERC B39 · (906) 487-4096 · g@mtu.edu · @sgowtham

Week #01: 2016/08/30 and 2016/09/01

Cross-listed as BE5390, EE5390 and MA5390

Tips to succeed

A quick guide to making the most of our time and efforts

http://dilbert.com/strip/2009-09-07/

Tips to succeed

* Treat the course as a job or an internship opportunity
  * Brush your teeth, take a shower, and dress appropriately
  * Show up and do things on time, and come prepared
  * Communicate your potential absence ahead of time
  * Work hard and smart, and keep your advisor happy

Getting and citing help

Actively inquire if someone needs help. Ask for help when necessary but be sure to cite it. Reciprocate the favor and/or pay it forward.

Tips to succeed Continued

- Treat every assignment as a proposal seeking external funding
  - Follow the guidelines and submission procedure
  - Get started as soon as possible, and leave enough time for revisions
  - Scripting, programming, and writing tasks can be very time consuming

Check the grammar

Its, It's, Quiet, Quite, Their, There, They're, Your, You're ... are all different. Attempt to use short, simple, and effective sentences – with appropriate punctuation – to convey your ideas.

Acceptable programming methodologies

You have the freedom to choose a programming language

You have the responsibility to learn what it can or cannot do.

* Compiled/Interpreted languages from scratch

  e.g., C/C++, FORTRAN, Julia, Mathematica, MATLAB, Python, R, etc.

  Required for all graded assignments

* Interpreted languages with built-in routines/modules

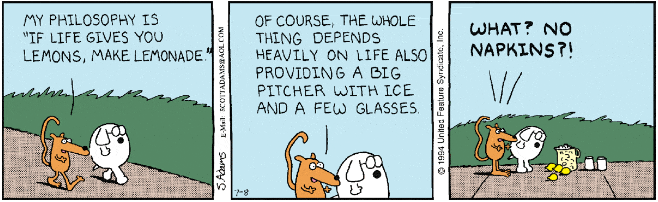
  e.g., Mathematica (`DSolve`, `Integrate`), MATLAB (`ode45`, `trapz`), etc.

  Compiled languages with third-party routines/modules e.g., C/C++, FORTRAN with BLAS, LINPACK, NAG, SCALAPACK, etc.

  May be used for project work and *free time* exercises

My teaching philosophy

A brief explanation of methods underlying my madness



http://dilbert.com/strip/1994-07-08/

My teaching philosophy

#1: Freedom within discipline

The more disciplined (i.e., doing things correctly on time every time) you are, more freedom (i.e., opportunities) you will earn to do the things you want – within the framework of this course.

#2: You'll be treated like what you could be and should be, and ...

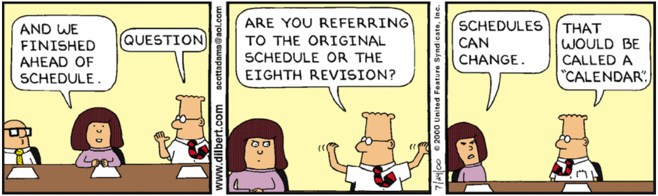
not as you are. You will often be put outside of your comfort zone to expand your capabilities. You will be expected to adapt quickly.

#3: You won't learn on an empty stomach

Get enough rest and eat something before lectures/meetings.

Revision Control System

Travel back and forth between revisions



http://dilbert.com/strip/2000-07-24/

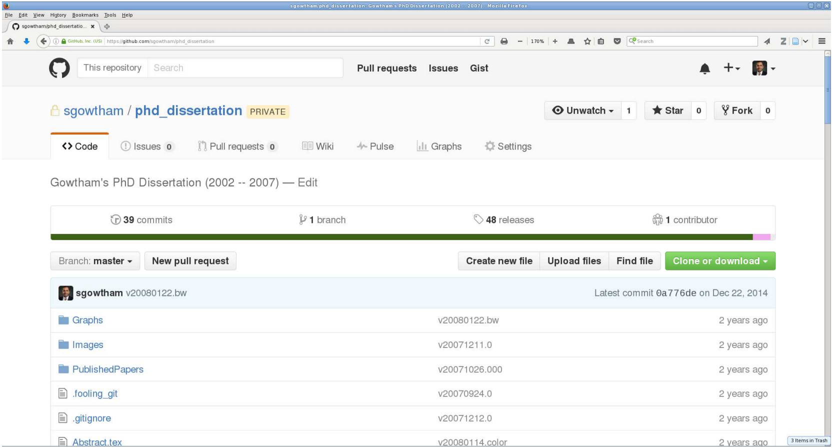
My dissertation How it looked without a formal revision control system

```
sgowtham@feynman:~/research/Dissertation/20080122.color
File  Edit  View  Search  Terminal  Help
[sgowtham@feynman Dissertation]$ ls
20070924.0  20071025.0    20071123.0  20071203.0    20071216.0  20080114.0
20070924.1  20071026.000  20071124.0  20071204.0    20071217.0  20080114.bw
20070925.0  20071030.0    20071125.0  20071205.000  20071218.0  20080114.color
20070927.0  20071030.1    20071126.0  20071211.0    20071219.0  20080121.bw
20070928.0  20071119.0    20071127.0  20071211.1    20071220.0  20080121.color
20071002.0  20071120.0    20071128.0  20071212.0    20071220.1  20080122.bw
20071022.0  20071121.0    20071129    20071213.0    20080107.0  20080122.color
20071023.0  20071122.0    20071129.0  20071214.0    20080109.0
[sgowtham@feynman Dissertation]$ cd 20080122.color
[sgowtham@feynman 20080122.color]$ ls
Abstract.tex          Chapter6.tex             MTUPhDThesis.sty
Abstract.txt          Chapter7.bib             MTUPhDThesis.sty.0
Acknowledgements.tex  Chapter7.tex             MyThesis.bib
Appendix.tex          Dedication.tex           MyThesis.dvi
Beowulf_Cluster.bib   Future_Work.bib          MyThesis.pdf
Beowulf_Cluster.tex   Future_Work.tex          MyThesis.tex
Bibliography.tex      Graphs                   Nano_Bio_Physics.bib
Chapter1.bib          Images                   Nano_Bio_Physics.tex
Chapter1.tex          Index.tex                nextpage.sty
Chapter2.bib          Introduction.bib         PublishedPapers
Chapter2.tex          Introduction.tex         README.PLEASE
Chapter3.bib          ListOfFigures.tex        TableOfContents.tex
Chapter3.tex          ListOfPublications.bib   Theoretical_Details.bib
Chapter4.bib          ListOfPublications.tex   Theoretical_Details.tex
Chapter4.tex          ListOfTables.tex         TOC.pdf
Chapter5.bib          Makefile                 TOC.tex
Chapter5.tex          Metal_Oxide_Clusters.bib
Chapter6.bib          Metal_Oxide_Clusters.tex
[sgowtham@feynman 20080122.color]$
```

My dissertation Impact of not using a formal revision control system

* Did not have to spend time learning something new near graduation
* Spent a lot of time incorporating edits from advisor and advisory committee members, and between versions
* An incomplete sentence, and missed out on thanking six good friends (and their parents) in the final printed copy as a result of picking an incorrect version to continue editing
* Lifelong shame of being inept and ungrateful

My dissertation How it would have looked with a formal revision control system



Git

Git

A distributed RCS with an emphasis on speed, data integrity, and support for distributed, non-linear workflows, and single/multiple users working on single/multiple projects.

Every working copy is a full-fledged repository with complete history and full version-tracking capabilities, independent of network access or a central server.

Potential applications

Systems administration, software development, manuscript preparation, event planning, etc.

http://git-scm.com
Linus Benedict Torvalds (1965 – present): Finnish American software engineer

Git and GitHub

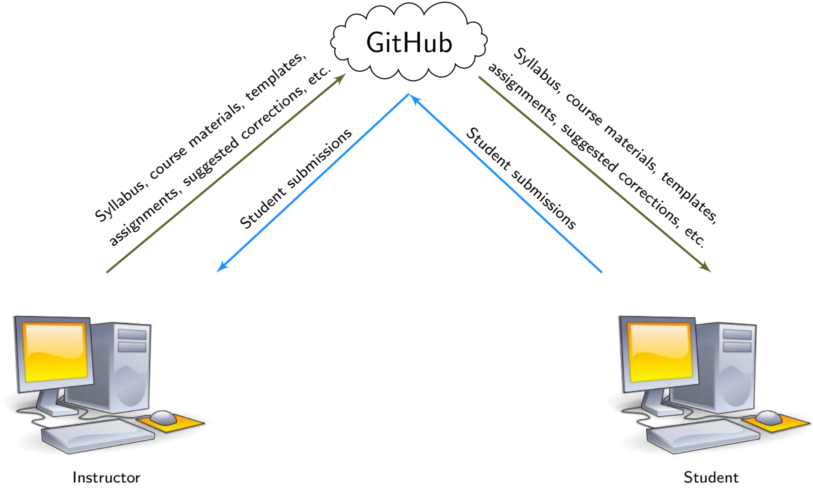
GitHub, world's largest code host

A safe, secure and social web-based hosting service for software development projects that use Git revision control system. GitHub's copy will be treated as the most trustworthy repository for UN5390.

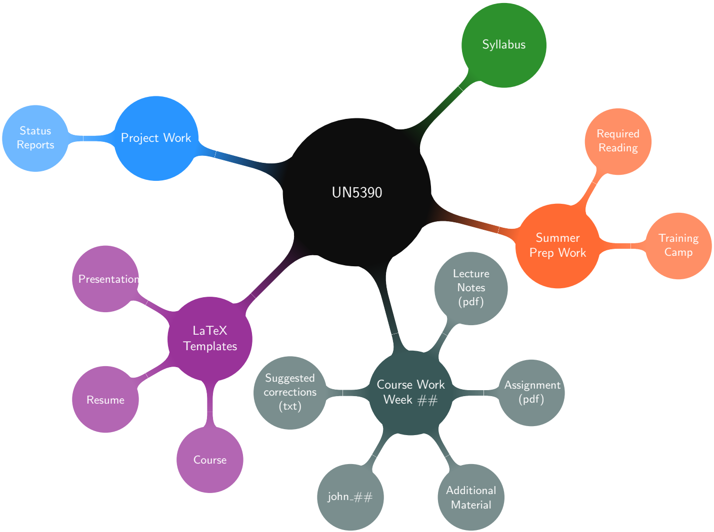
- The learning curve can be steep
- A form of data backup that keeps track of the workflow
- Easily move back and forth between revisions
- A readily available portfolio for potential employers
- Saves space, time, $, and creates opportunities

http://github.com

Git, GitHub, and UN5390



Structure of the course repository



Replace `john` with your Michigan ISO username. Replace `##` with the appropriate week number.

The honor system of *who writes where?*

- For the instructor: everything under ${UN5390} is *read-write* except

  CourseWork/Week_##/john_##/
  LaTeXTemplates/Course/
  LaTeXTemplates/Resume/

- For students: everything under ${UN5390} is *read-only* except

  CourseWork/Week_##/john_##/
  LaTeXTemplates/Course/
  LaTeXTemplates/Resume/

Replace john with your Michigan ISO username. Replace ## with the appropriate week number.
${HOME}/git_work/un5390_f2016_john will henceforth be abbreviated as ${UN5390}.

${UN5390}/.gitignore

- Every Git repository should have one at its very top level
- A *read-only* file for students per the honor system
- List of files, folders and file types that should not be in the repository
  - OS- and language-specific temporary files
  - System files and symbolic links
  - Program executables and other binary files
  - Files with large data sets and/or sensitive information
  - A class of entities can be specified with wild card characters

Git workflow

Clone the repository (in an IT-managed Linux workstation)

```
git config --global user.name "John Sanderson"
git config --global user.email "john@mtu.edu"
git config --global core.editor vim
git config --list

mkdir -p ${HOME}/git_work
cd ${HOME}/git_work
git clone \
  git@github.com:MichiganTech/un5390_f2016_john.git
un5390
ls -latrh
tree
```

Replace `john` with your Michigan ISO username, and John Sanderson with your real/preferred name.
\ is the continuation character in BASH and indicates that the command continues into the next line.
Cloning the repository needs to be done only once per machine.

Git workflow

Test write permissions (in an IT-managed Linux workstation)

```
cd ${UN5390}/CourseWork/Week_01/john_01
git pull
touch test_file.txt
git add test_file.txt
git commit -m "Adding test_file.txt for testing GitHub"
git push origin master
## Visit GitHub.com and check if test_file.txt exists
git pull
git rm test_file.txt
git commit -m "Removing test_file.txt from GitHub"
git push origin master
## Visit GitHub.com and check if test_file.txt is gone
```

Replace `john` with your Michigan ISO username.

Git workflow

Receiving [updated] course material, corrections, etc.

```
cd ${UN5390}
git pull
```

Discarding edits

Suppose that the *read only* files which contain inadvertent edits are `${UN5390}/.gitignore` and `${UN5390}/README.md`.

```
cd ${UN5390}
git checkout -- .gitignore
git checkout -- README.md
```

The cause of this problem is usually not paying enough attention. An easy cure is to cultivate the habit of typing `pwd` every time you get into or change directory, and observe where you are before doing anything.

Git workflow

Submitting a partially completed assignment

```
cd ${UN5390}/CourseWork/Week_01
git pull
git add john_01
git commit -m "Submitting problem #1 in assignment #01"
git push origin master
```

Submitting a completed assignment

```
cd ${UN5390}/CourseWork/Week_01
git pull
git add john_01
git tag -a a01 -m "Submitting assignment #01"
git push origin a01
```

Replace `john` with your Michigan ISO username. Replace `01` appropriately for subsequent assignments.
Review and follow the assignment submission workflow in `Assignment_01.pdf`.

Viewing commit history

Commands (text; one per line)

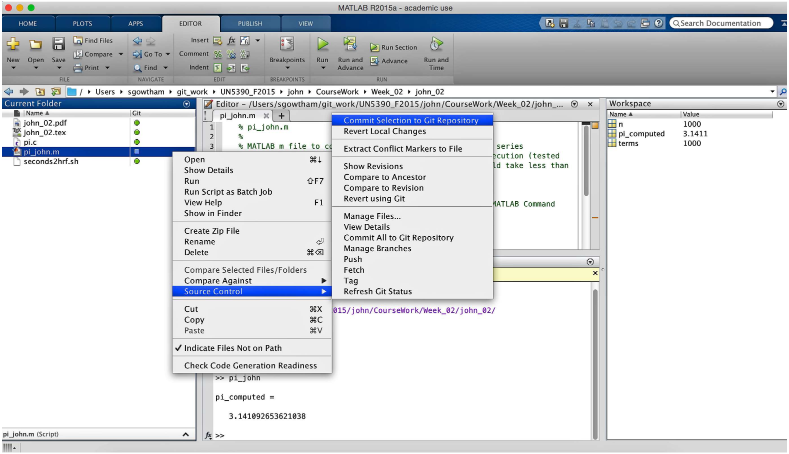
```
cd ${UN5390}
git log --pretty=format:"%h - %an, %ad :  %s"
```

Commands (graphical; one per line)

```
cd ${UN5390}
gource --hide dirnames,filenames --seconds-per-day 0.1 \
   --auto-skip-seconds 1 -1280x720 -o - | \
   ffmpeg -y -r 60 -f image2pipe -vcodec ppm -i - \
   -vcodec libx264 -preset ultrafast -pix_fmt \
   yuv420p -crf 1 -threads 0 -bf 0 gource.mp4
```

http://git-scm.com/book/en/Git-Basics-Viewing-the-Commit-History
Gource: Google project page | Linux kernel development 1991-2012 ; not installed in `colossus.it` or `guardian.it`.

Git and MATLAB R2014b and beyond



http://www.mathworks.com/products/matlab/whatsnew.html

Additional references

* Git

  Reference | Book | Videos | External links
  Tagging | Forking | Branching and merging

* Git – structuring commit messages

  On commit messages | Writing good commit messages
  How to write a git commit message

* GitHub

  Interactive tutorial | Cheat sheet | Online training

* Twitter

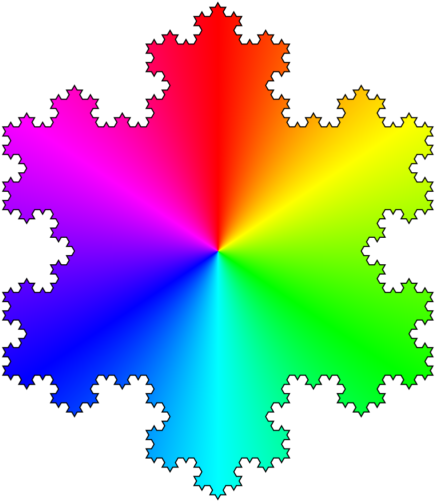
  @GitHub | @GitHubEducation | @GitHubStatus

Before we meet again

- Review the syllabus, course material through week #01, notations, active participation, free time exercises, tips, opportunities, and mathematical results
- Review/Complete the training camps (#01 – #10)
- Get started on assignment #01
- Get to know one of your classmates, someone you didn't know before

Continuous integration/improvement

You will be expected to incorporate material from a following week into an ongoing assignment - e.g., week #02 material into assignment #01. Getting started on assignments and practicing the course material on time will facilitate turning in your submissions on or ahead of time.



End of Tuesday lecture.

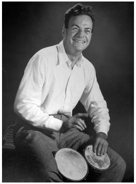
A Brief History of Computing

Curiosity (about the universe) in thoughtful action using computers

http://dilbert.com/strip/2012-07-11/

Freedom to doubt is born of a struggle against authority

We have found it of paramount importance that in order to progress, we must recognize our ignorance and leave room for doubt. Scientific knowledge is a body of statements of varying degrees of certainty – some most unsure, some nearly sure, but none absolutely certain.

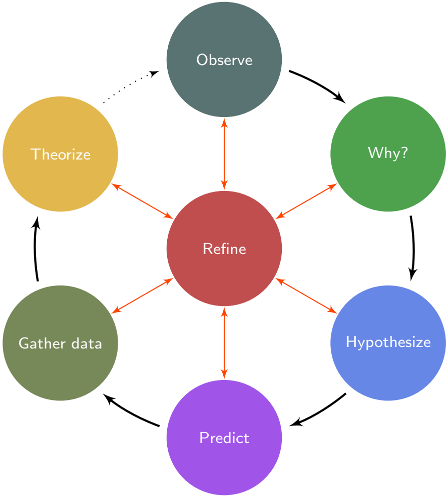


The guessing game

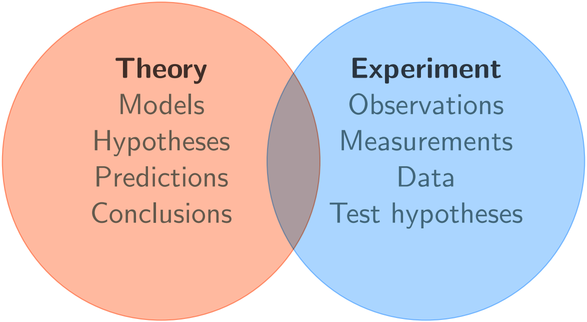
We look for a new law by the following process. First, we guess it. Then we compute the consequences of the guess, to see what, if this is right, if this law we guess is right, to see what it would imply and then we compare the computation results to nature, or we say compare to experiment or experience, compare it directly with observations to see if it works. If it disagrees with experiment, it's wrong.

Richard Phillips Feynman (1911 – 1988): American theoretical physicist, and 1965 Physics Nobel Laureate

Scientific method



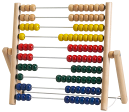
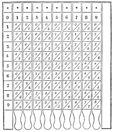
Scientific method



Computing

How can we know how much more (or less) we need
if we don't know how much we already have?

A goal-oriented discipline and responsibility requiring, benefiting from, or creating a mathematical sequence of steps known as an algorithm, with or without use of the formal tools to solve a problem.

Year	Discoverer/Discovery
1588	Joost Buerghi, natural logarithms
1614	John Napier, Napier's bones
1614	William Oughtred, slide rules
1642	Blaise Pascal, mechanical calculator
1671	Gottfried Leibniz, stepped reckoner
1774	Philipp Hahn, portable calculator
1775	Charles Stanhope, multiplying calculator
1786	Johann Mueller, idea of a difference engine
1801	Joseph-Marie Jacquard, automatic loom controlled by punch cards
1820	Charles Xavier Thomas de Colmar, arithmometer
1822	Charles Babbage, a prototype of difference engine
1832	Semen Koraskov, punch cards for information storage/search
1834	Charles Babbage, design of analytical engine with punch cards

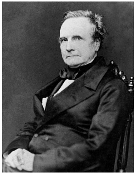
Computing Impact of everyday life

Year	Discoverer/Discovery
1835	Joseph Henry, electromechanical relay
1843	Per Georg Scheutz, third-order difference engine with printer
1847	Charles Babbage, difference engine #2
1848	George Boole, binary/Boolean algebra
1853	Per Georg Scheutz, tabulating machine – 4th-order difference engine
1869	William Jevons, practical logic machine
1884	Dorr Felt, comptometer
1886	Herman Hollerith, tabulating system
1889	Dorr Felt, printing desk calculator

Augusta Ada King, Countess of Lovelace

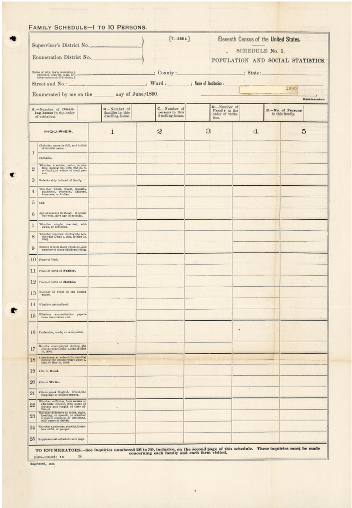
Ada's notes on Analytical Engine, Charles Babbage's early mechanical general-purpose computer, include sequential computation of the Bernoulli numbers – now recognized as the first algorithm intended to be carried out by a machine.

She also developed a vision of the capability of computers to go beyond mere number-crunching. Her poetically scientific mindset led to examining how individuals and society relate to technology as a collaborative tool.



Augusta Ada King (1815 – 1852): English mathematician and writer
Charles Babbage (1791 – 1871): English mathematician, engineer, philosopher and inventor
Jakob Bernoulli (1654 – 1705): Swiss mathematician

Computing The US Census of 1890, and the beginning of a new frontier



FAMILY SCHEDULE—1 TO 10 PERSONS.

Supervisor's District No. ____ [7—556 b.] Eleventh Census of the United States.

Enumeration District No. ____ SCHEDULE No. 1.

POPULATION AND SOCIAL STATISTICS.

Name of city, town, township, precinct, district, beat, or other minor civil division. ____; County: ____; State: ____

Street and No.: ____; Ward: ____; Name of Institution: ____ 1890

Enumerated by me on the ____ day of June, 1890. ____, Enumerator.

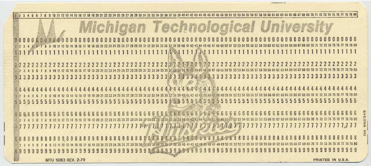
A.—Number of Dwelling-house in the order of visitation.		B.—Number of families in this dwelling-house.		C.—Number of persons in this dwelling-house.		D.—Number of Family in the order of visitation.		E.—No. of Persons in this family.	

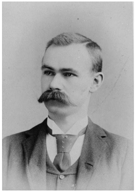
	INQUIRIES.	1	2	3	4	5
1	Christian name in full, and initial of middle name.					
	Surname.					
2	Whether a soldier, sailor, or marine during the civil war (U. S. or Conf.), or widow of such person.					
3	Relationship to head of family.					
4	Whether white, black, mulatto, quadroon, octoroon, Chinese, Japanese, or Indian.					
5	Sex.					
6	Age at nearest birthday. If under one year, give age in months.					
7	Whether single, married, widowed, or divorced.					
8	Whether married during the census year (June 1, 1889, to May 31, 1890).					
9	Mother of how many children, and number of these children living.					
10	Place of birth.					
11	Place of birth of **Father.**					
12	Place of birth of **Mother.**					
13	Number of years in the United States.					
14	Whether naturalized.					
15	Whether naturalization papers have been taken out.					
16	Profession, trade, or occupation.					
17	Months unemployed during the census year (June 1, 1889, to May 31, 1890).					
18	Attendance at school (in months) during the census year (June 1, 1889, to May 31, 1890).					
19	Able to **Read.**					
20	Able to **Write.**					
21	Able to speak English. If not, the language or dialect spoken.					
22	Whether suffering from **acute** or **chronic** disease, with name of disease and length of time afflicted.					
23	Whether defective in mind, sight, hearing, or speech, or whether crippled, maimed, or deformed, with name of defect.					
24	Whether a prisoner, convict, homeless child, or pauper.					
25	Supplemental schedule and page.					

TO ENUMERATORS.—See inquiries numbered 26 to 30, inclusive, on the second page of this schedule. These inquiries must be made concerning each family and each farm visited.

(19379—1,750,000.) 1 b 34

Reprinted, 1935





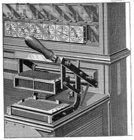
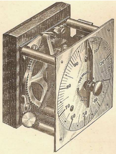


Fig. 8.—Circuit Closing Press.
Hollerith's Electric Sorting and Tabulating Machine.



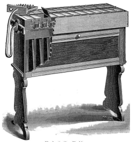
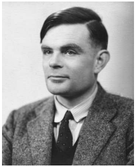


Fig. 3.—Sorting Machine.
Hollerith's Electric Sorting and Tabulating Machine.

Herman Hollerith (1860 – 1929): American statistician and inventor



Turing Machine (TM; 1936)

A mathematical model of computation that defines an abstract machine which can manipulate the symbols on a strip of tape according to a table of rules.

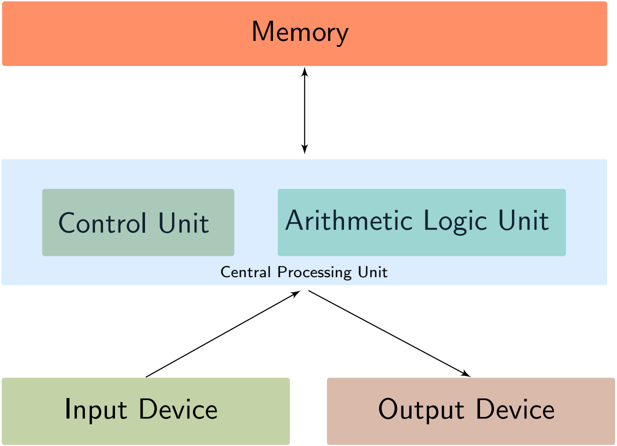
Given any computer algorithm, a TM can be constructed that is capable of simulating that algorithm's logic.

Universal Turing Machine (UTM; 1936-37)

A TM that can simulate an arbitrary TM on arbitrary input by reading both the description of the machine to be simulated as well as the input thereof from its own tape.

Alan Mathison Turing (1912 – 1954): British mathematician, computer scientist and philosopher

Computing John von Neumann architecture



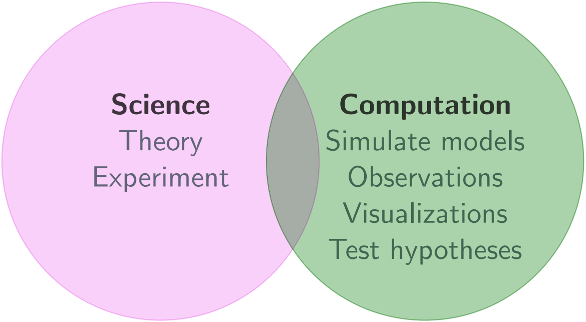
John von Neumann (1903 – 1957): Austrian-Hungarian/American mathematician
John von Neumann was one of only several contributors to the 1945 report, and its first (and incomplete) draft was inadvertently distributed by Herman Goldstine (security officer of the ENIAC project).

Computing Storage hierarchy/Data access time

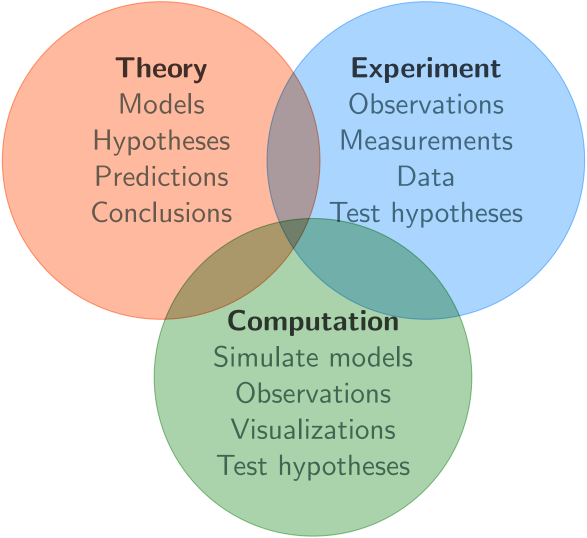
CPU registers	$\sim$1 machine cycle; very low capacity (B – kB); not easily expandable; volatile; very expensive
Cache	1-10 machine cycles; low capacity (kB – MB); not easily expandable; volatile; very expensive; 700 GB/sec (L1), 200 GB/sec (L2), 100 GB/sec (L3), 40 GB/sec (L4)
RAM	25-100 machine cycles; high capacity (GB); easily expandable; volatile; expensive; 10 GB/sec
Local storage	$\sim 10^6$ machine cycles; high capacity (GB – TB); easily expandable; non-volatile; inexpensive; 600 MB/sec (SSD)
Network storage	$10^6 - 10^8$ machine cycles; high capacity (TB – PB); easily expandable; non-volatile; inexpensive; 160 MB/sec
Offline storage	very slow access times; very high capacity (PB – EB); easily expandable; non-volatile; inexpensive

1 machine cycle for a 1.0 GHz processor is 1×10^{-9} second.

Scientific computing



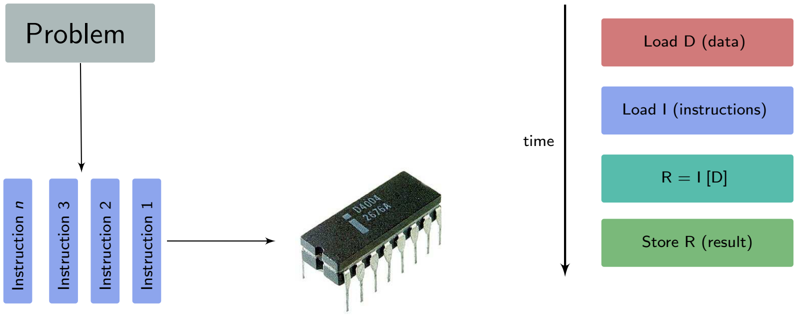
Scientific computing



Scientific computing

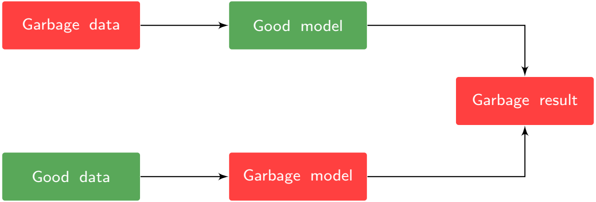
A multi-disciplinary field that uses advanced computing capabilities and involves the application of computer simulations to solve complex problems in various arts, science and engineering disciplines – to gain an understanding mainly through the analysis of mathematical models implemented on computers.

Scientific computing Divide and conquer

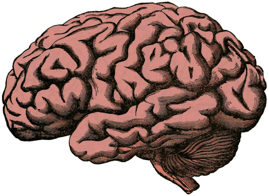


The world's first general purpose single chip micro-processor, invented by Intel in 1971, 4004 had a 12-bit instruction pointer, forty six 8-bit opcodes, and sixteen 4-bit registers. It was a 16-pin DIP part, containing about 2,000 transistors, and at 740 kHz, it was capable of performing 92,000 instructions per second.

Scientific computing Garbage In, Garbage Out (GIGO) paradigm



There is no substitute for thinking



Scientific computing Spectrum of applications

Store data

Interact with humans

Run programs

Interact with computers

A floating-point number has a decimal point in it

3.14159, 2.71828, 0.57721, 1.61803, 42×10^{-5}, etc.

A floating-point operation is any mathematical operation [addition (+), subtraction (-), multiplication (∗), division (/), exponentiation (^)] that involves floating-point numbers OR assignment of a floating-point number to a variable

$a = 3.14159$

$b = 2.71828$

$c = 1.61803$

$p = 8 * \left(a + c^{b}\right) + (b - 4 * c) - [c/(b + a)]^{-0.50}$

Scientific computing A measure of performance

The number of floating-point operations a computer can perform in one second, FLOPS, is a measure of its computing power

$$\boxed{\text{FLOPS} \;=\; \#\text{ of processors} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

Most scientific applications use floating-point operations

More FLOPS implies a more powerful computer

A double-preicison floating-point number requires 8 bytes of memory

Most scientific applications use double-precision numbers

More memory (RAM and HD) implies a more powerful computer

The number of floating-point operations per CPU cycle is a hardware characteristic.

Scientific computing A measure of performance

* Smart phones (iPhone 5s: 1 GB RAM, 64 GB storage, 75 GFLOPS)

FLOPS = # of processors $\times$ CPU speed $\times$ FLOPs/CPU cycle

Scientific computing A measure of performance

- Smart phones (iPhone 5s: 1 GB RAM, 64 GB storage, 75 GFLOPS)

- Laptops (MacBook Pro: 16 GB RAM, 500 GB SSD storage, 80 GFLOPS)

FLOPS = # of processors × CPU speed × FLOPs/CPU cycle

- **Smart phones** (iPhone 5s: 1 GB RAM, 64 GB storage, 75 GFLOPS)
- **Laptops** (MacBook Pro: 16 GB RAM, 500 GB SSD storage, 80 GFLOPS)
- **Workstations** (my office: 64 GB RAM, 2 TB SSD storage, 250 GFLOPS)

FLOPS $=$ # of processors $\times$ CPU speed $\times$ FLOPs/CPU cycle

Scientific computing A measure of performance

- ⁎ Smart phones (iPhone 5s: 1 GB RAM, 64 GB storage, 75 GFLOPS)
- ⁎ Laptops (MacBook Pro: 16 GB RAM, 500 GB SSD storage, 80 GFLOPS)
- ⁎ Workstations (my office: 64 GB RAM, 2 TB SSD storage, 250 GFLOPS)
- ⁎ University clusters (Superior: 6 TB RAM, 33 TB storage, 30 TFLOPS)

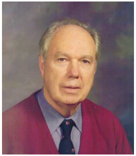
FLOPS = # of processors × CPU speed × FLOPs/CPU cycle

Scientific computing A measure of performance

- **Smart phones** (iPhone 5s: 1 GB RAM, 64 GB storage, 75 GFLOPS)
- **Laptops** (MacBook Pro: 16 GB RAM, 500 GB SSD storage, 80 GFLOPS)
- **Workstations** (my office: 64 GB RAM, 2 TB SSD storage, 250 GFLOPS)
- **University clusters** (Superior: 6 TB RAM, 33 TB storage, 30 TFLOPS)
- **Leadership supercomputing centers** (ORNL: 700 TB RAM, 40 PB storage, 17 PFLOPS)

FLOPS = # of processors × CPU speed × FLOPs/CPU cycle

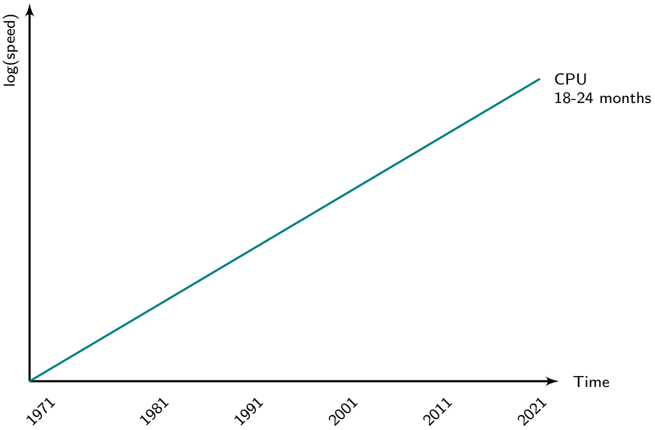
Scientific computing Flynn's taxonomy

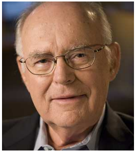


Data \ Instruction	Single	Multiple
Single	SISD (Serial)	MISD (Parallel)
Multiple	SIMD (Parallel)	MIMD (Parallel)

Michael Flynn (1934 – present): American computer scientist; Professor, Stanford University

Moore's Law





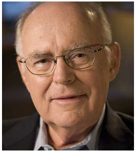
Gordon Earle Moore (1929 – present): American chemist and entrepreneur; co-founder, Intel

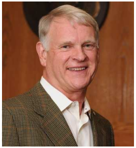
Moore's Law and Michigan Tech

A popular re-formulation

While component density gains were slowing by 1975, Intel's Dave House observed that individual components were themselves getting faster. He theorized that this meant computing power on a chip could double about every 18 months – slower than Moore's original 1965 prediction, but faster than the 1975 revision.

This is the form of the ~~Law~~ self-fulfilling prophecy that has become popular and has been carefully – almost slavishly – relied on by the semiconductor industry.





Gordon Earle Moore (1929 – present): American chemist and entrepreneur; co-founder, Intel
Dave House (1943 – present): American electrical engineer and entrepreneur (BS, Michigan Tech, 1965)
Excerpts From A Conversation With Gordon Moore (2005) | An Interview With Dave House (2005)

Brainstorm

The formula for FLOPS for a computer with one processor is given by

$$\boxed{\text{FLOPS} = \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

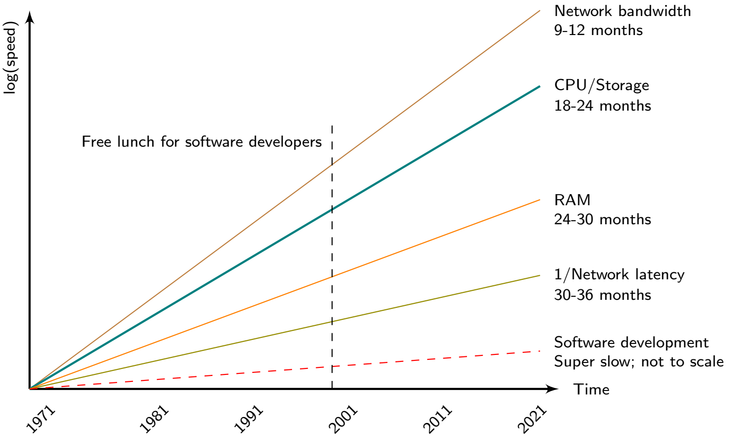
How can one go about increasing performance? What are the benefits and caveats associated with such an approach?

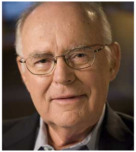
A feasible upper limit for CPU speed is ∼4 GHz

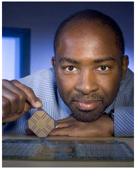
Increasing CPU's speed increases required electricity (∼60% increase for every 400 MHz increase in CPU speed) and generated heat. The cost of manufacturing a new processor with higher FLOPs/cycle and issues with how fast electrons can move across a chip are other concerns.

Updated with answers/feedback from Akhil Kurup, Jeff Brookins, and Ian Cummings.

Moore's Law in practice







Gordon Earle Moore (1929 – present): American chemist and entrepreneur; co-founder, Intel
Kunle Olukotun (1962 – present): American electrical engineer of Nigerian origin; Professor, Stanford University
IBM POWER4, in 2001, became the very first commercially released multi-core processor (2 processors, 1.1 GHz).

Brainstorm

If each double-precision element requires 8 bytes, what is the maximum order of a square matrix that can fit within 1 GB RAM?

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

My approach

A square matrix of order N has N^2 elements
A double-precision element requires 8 bytes
1 GB is 1024 x 1024 x 1024 bytes

8 is close enough to 10; a double-precision element requires 10 bytes
1024 is close enough to 1000 (i.e., 10^3); 1 GB is 10^9 bytes

$$\# \text{ of elements in 1 GB} = \frac{\# \text{ of bytes in 1 GB}}{\# \text{ of bytes per element}} = \frac{10^9}{10} = 10^8$$

$$N^2 = \# \text{ of elements in 1 GB} = 10^8$$

$$N = \sqrt{\# \text{ of elements in 1 GB}} = 10^4 = 10,000$$

If approximations weren't used, the value of N would be 11,585.
You may use this or a similar technique to solve *back of the envelope* exercises.

Brainstorm

Estimate the number of floating-point operations necessary to compute the product of two square and dense matrices of order N.

Assuming that N is very large (10^6), estimate the time required to solve this problem using one 2.0 GHz CPU with 4 FLOPs/cycle.

$$\boxed{\text{FLOPS} = \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

My approach

Start with a 2x2 symbolic matrix.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a*e+b*g & a*f+b*h \\ c*e+d*g & c*f+c*h \end{pmatrix}$$

Each element of the product matrix requires two multiplication (i.e., order of the matrix) and one addition (i.e., one less than the order of the matrix) operations – for a total of three floating-point operations.

There are four (i.e., 2^2, square of the order) elements, and a total of $2^2 \times 3 = 12$ operations.

You may use this or a similar technique to solve *back of the envelope* exercises.

My approach

Extend the idea for the 3x3 case.

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} j & k & l \\ m & n & o \\ p & q & r \end{pmatrix} = \begin{pmatrix} a*j+b*m+c*p & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix}$$

Each element of the product matrix requires three multiplication (i.e., order of the matrix) and two addition (i.e., one less than the order of the matrix) operations – for a total of five floating-point operations.

There are nine (i.e., 3^2, square of the order) elements, and a total of $3^2 \times 5 = 45$ operations.

You may use this or a similar technique to solve *back of the envelope* exercises.

My approach (continued)

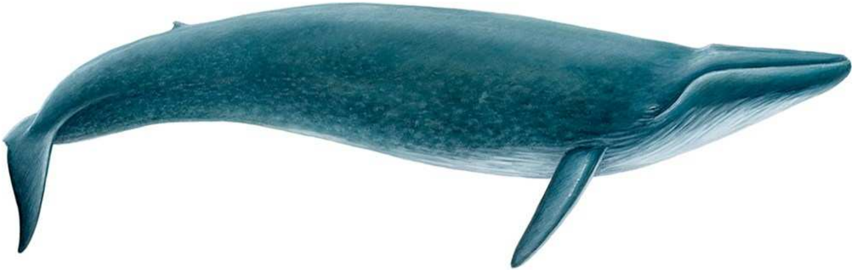
Extending to NxN case, each element of the product matrix requires N multiplication and $N-1$ addition operations – for a total of $(2N-1)$ floating-point operations. There are N^2 such elements, and a total of $N^2 \times (2N-1)$ operations.

One day has 86,400 seconds $\simeq$ 100,000 (i.e., 10^5)
One 2.0 GHz processor, 4 FLOPs/cycle $\Rightarrow$ 8 GFLOPS $\simeq 10^{10}$ FLOPS
When N is large, $N^2 \times (2N-1) = 2N^3 - N^2 \simeq N^3$ operations

For $N = 10^6$, $N^3 = 10^{18}$ floating-point operations
Computer can do 10^{10} floating-point operations per second
It'd take $10^{18}/10^{10} = 10^8$ seconds $= 10^3$ days $\simeq$ 3 years to complete

If approximations weren't used, the time required to complete matrix multiplication would be 7.93 years.

Supercomputing The need for size and speed



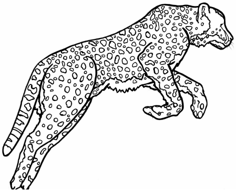


Image courtesy: Google images

Supercomputer

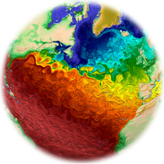
A very big and very fast computer with a lot of FLOPS

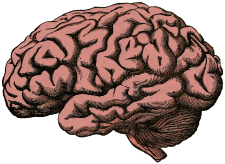
$$\boxed{\text{FLOPS} = \#\text{ of processors} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

A network of fairly powerful computers, with software to facilitate communication amongst them, working together to solve a complex problem acts as if it's a <u>super</u> computer

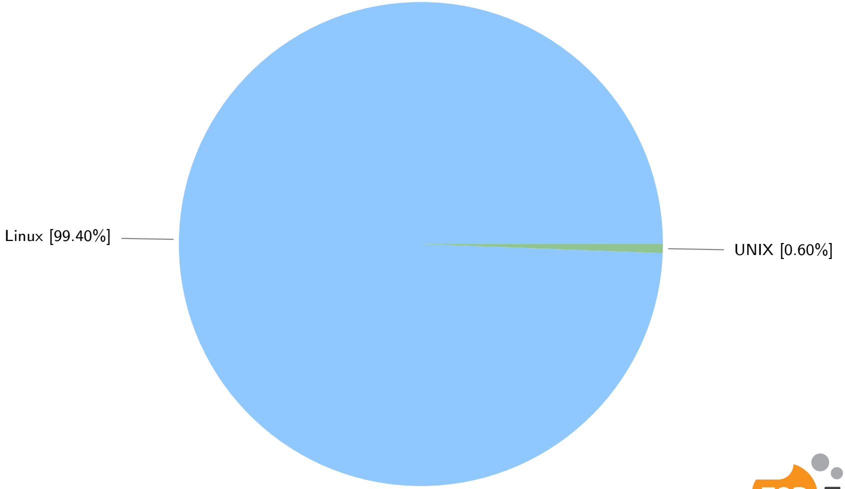
Supercomputing

Complex computational activities using a supercomputer



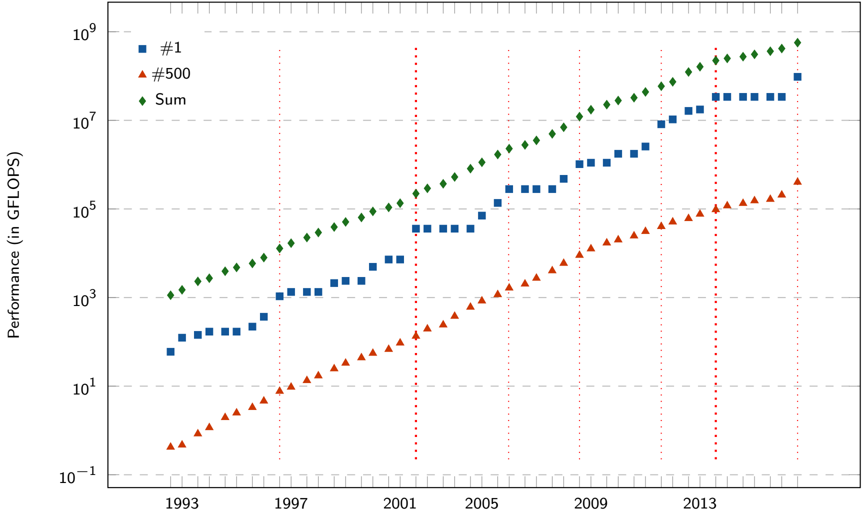



Supercomputing Top 500, the OS distribution



http://top500.org/lists/2016/06/

Supercomputing Top 500, and the effect of Moore's Law and multicore processors



Supercomputing Top 500, and the effect of Moore's Law and multicore processors

- All 500 in 2001 $\simeq$ #500 in 2016
- #350 current edition $\simeq$ #500 next edition
- #440 in 2001 $\simeq$ iPhone 5s
- Today's supercomputer $\simeq$ Laptop in 10-15 years
  - #250 in 2000, #375 in 2001 $\simeq$ MacBook Pro

http://top500.org

Supercomputing The top 10 in Top 500

#	Name	Location	Processors	TFLOPS Expt	TFLOPS Theory	Power (MW)
01	Sunway TaihuLight	2016, NSCW, China	10,649,600	93,014.6	125,435.9	15.371
02	Tianhe-2	2013, NSCG, China	3,120,000	33,862.7	54,902.4	17.808
03	Titan	2012, ORNL, USA	560,640	17,590.0	27,112.5	8.209
04	Sequoia	2011, LLNL, USA	1,572,864	17,173.2	20,132.7	7.890
05	K	2011, RIKEN, Japan	705,024	10,510.0	11,280.4	12.660
06	Mira	2012, ANL, USA	786,432	8,586.6	10,066.3	3.945
07	Trinity	2015, LANL, USA	301,056	8,100.9	11,078.9	–
08	Piz Daint	2013, CSCS, Switzerland	115,984	6,271.0	7,788.9	2.325
09	Hazel Hen	2015, HLRS, Germany	185,088	5,640.2	7,403.5	–
10	Shaheen II	2015, KAUST, Saudi Arabia	196,608	5,537.0	7235.2	2.834

http://top500.org/lists/2016/06/

Supercomputing A closer look at #1

- #1 in 2016/06
- National Supercomputing Center, China
- 10.6 million cores, 1.3 PB RAM, and 93 PFLOPS
- $270 million and 15 MW electricity

The power of human brain

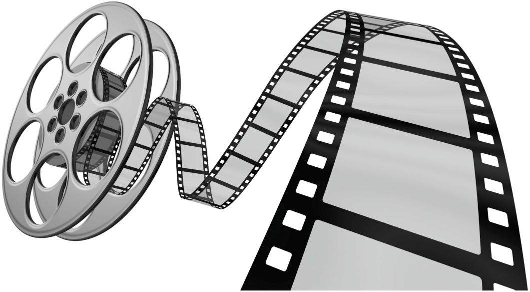
K, #5 in the list with 80,000+ processors and 1 PB RAM, required 40 minutes to simulate about one second worth activity of 1% of human brain ($\simeq$ 1.7 billion neurons and 10 trillion synapses; estimate is that an adult brain has 80-100 billion neurons).

http://top500.org/lists/2016/06/

Supercomputing

Impact on everyday life

An overview of the spectrum of supercomputing applications



This and subsequent videos will be available in Impact of Supercomputing.

- **National Strategic Computing Initiative**

  Executive order by the President of the US to maximize benefits of HPC research, development, and deployment

- **CORAL**

  Department of Energy's $425M investment in the joint Collaboration of Oak Ridge (Summit; 150+ PFLOPS), Argonne (Aurora, 180+ PFLOPS) and Livermore (Sierra, 150+ PFLOPS) national laboratories for next-generation supercomputing technologies

- **Brain Research through Advancing Innovative Neurotechnologies**

  NSF's BRAIN initiative to bridge the scales that span from atoms to thoughts and behavior, address neurobiological questions, and generate brain-inspired smart technologies to meet future societal needs

It's not just a keel and a hull and a deck and sails.
That's what a ship needs but not what a ship is.
But what a ship is ... what the Black Pearl really is ... is freedom.

– Captain Jack Sparrow, Pirates of the Caribbean: The Curse of the Black Pearl

Supercomputing Parting thoughts

Our universe contains answers to questions which we don't yet know how to ask. Science represents a temperment and a methodology that'll help us formulate some of these questions. And when we do, (super) computing might be a tool that'll help us get the answers with some degree of certainty.

Additional references

- ACM Education
- Intl Journal of Modeling, Simulation, and Scientific Computing
- Journal of Computational Science
- Scientific Computing World
- Scientific Programming
- SIAM Journal on Scientific Computing
- Springer Journal of Scientific Computing
- Twitter

  @ElsevierMath | @SCWMagazine | @SIAMConnect

Additional references

- Sketch Of The Analytical Engine Invented By Charles Babbage
  L. F. Menabrea
  Bibliothéque Universelle de Genéve, No. 82 (1842)
  Translator notes by A. Lovelace in Scientific Memoirs, vol. 3 (1843)

- On Computable Numbers, With An Application To The Entscheidungsproblem
  A. M. Turing
  Proceedings of the London Mathematical Society, vol. 42, p. 230 (1937)

- First Draft Of A Report On The EDVAC
  J. von Neumann (1945)

PDF in `AdditionalMaterial` folder.

Additional references

- As We May Think
  V. Bush
  Interactions, vol. 3, p 35 (1996); The Atlantic Monthly (1945)

- Computing Machinery And Intelligence
  A. M. Turing
  Mind, vol. LIX, p. 433 (1950)

- Cramming More Components Onto Integrated Circuits
  G. E. Moore
  Electronics, vol. 38, p.114 (1965)

- Some Computer Organizations And Their Effectiveness
  M. J. Flynn
  IEEE Transactions on Computers, vol. C-21, p. 948 (1972)

PDF in `AdditionalMaterial` folder.

Additional references

- 'Do Not Fold, Spindle Or Mutilate': A Cultural History Of The Punch Card
  S. Lubar
  Journal of American Culture, vol. 15, p. 43 (1992)

- Two Heads Are Better Than One
  W. Knight
  IEE Review, vol. 51, p. 32 (2005)

- MilkyWay-2 Supercomputer: System And Application
  X. Liao, L. Xiao, C. Yang, Y. Lu
  Frontiers of Computer Science, vol. 8, p. 345 (2013)

PDF in `AdditionalMaterial` folder.

Additional references

- Limits On Fundamental Limits To Computation
  I. L. Markov
  Nature, vol. 512, p. 147 (2014)

- 50 Years Of Moore's Law
  IEEE Spectrum Special Report (2015)

- First 2000 Years Of Computing | Timeline

- Intel Timeline: A History Of Innovation

- The Story Of Intel 4004

- Origins of the Computer Simulation Program
  Jeffrey Lewin, Biological Sciences, Michigan Tech (2012)

PDF in `AdditionalMaterial` folder.

Additional references

- Works Of Charles Babbage
- Hackers: Heroes Of The Computer Revolution
  S. Levy; Anchor Press/Doubleday (1984)
- Recording Gender: Women's Changing Participation In Computing
  J. Abbate; MIT Press (2012)
- Alan Turing: The Enigma
  A. Hodges; Princeton University Press (2002)
- The Innovators: How A Group Of Hackers, Geniuses, And Geeks Created The Digital Revolution
  W. Isaacson; Simon & Schuster (2014)

Additional references

- UNIVAC – Then And Now (CHM, video, 1960)
- Computer Pioneers – Pioneer Computers I (CHM, video, 1966)
- Computer Pioneers – Pioneer Computers II (CHM, video, 1966)
- The Imitation Game (iTunes, movie, 2014)
  Based on *Alan Turing: The Enigma* by A. Hodges
- Robert Noyce: The Man Behind The Microchip
- Moore's Law At 50 (Vimeo, video, 2015)

Additional references

* Computer History Museum
  1401 N. Shoreline Blvd., Mountain View, CA 94043
  (650) 810-1010
* Intel Museum
  2200 Mission College Blvd., Santa Clara, CA 95054
  (408) 765-5050
* The Tech Museum Of Innovation
  201 S. Market St., San Jose, CA 95113
  (408) 294-8324
* Science Museum
  Exhibition Rd., South Kensington, London, SW7 2DD

Additional references

* HP Garage, The Birthplace Of Silicon Valley
  367 Addison Ave., Palo Alto, CA 94309

* Birthplace Of Apple
  2066 Crist Dr., Los Altos, CA 94024

* Shockley Semiconductor Laboratory
  391 San Antonio Rd., Mountain View, CA 94040

* Fairchild Semiconductor
  844 E Charleston Road, Palo Alto, CA 94303

* Birthplace Of Intel
  365 E Middlefield Road, Mountain View, CA 94043

Additional references

* Twitter

  @ComputerHistory | @HPCCouncil | @HPCWire | @IBM | @IBMHPC
  @IBMResearch | @insideHPC | @Intel | @IntelHPC | @MITLL
  @NSF | @NSFGRFP | @SciNode | @ScienceMuseum
  @Supercomputing | @TACC | @TheTechMuseum
  @Top500Supercomp | @XSEDEScience

Compliance and Security

Institutional, federal and/or funding agency guidelines

The need

Who cares?

The University has an obligation to comply with laws, regulations, policies, and standards associated with information security to preserve the confidentiality, integrity, availability of, and satisfy legal and ethical responsibilities with regard to information assets owned or entrusted by the university.

Institutional policies and procedures

Information Security Compliance Policy

Information Security Compliance Policy is a university-wide approach to information security to help identify and prevent the compromise of information security and the misuse of university information technology by which all university faculty, staff and students must adhere when handling information.

Institutional Policies and Procedures Violation Consequences

Any university employee, student or non-university individual with access to university data who engages in unauthorized use, disclosure, alteration, or destruction of data will be subject to appropriate disciplinary action, including possible dismissal and/or legal action.

http://www.admin.mtu.edu/admin/policy/gen/1009.htm

Information assets, technology resources and security

Information assets

Definable pieces of information in any form, recorded or stored on any media that is recognized as *valuable* to the University.

Information Technology resources

The data, applications, information assets, and related sources, such as personnel, equipment, networks and computer systems of the University.

Information security

Protection of the University's data, applications, networks, and computer systems from unauthorized access, alteration, or destruction.

Data custodian, owner and user

Data custodian

An employee of the University who has administrative and/or operational responsibility over information assets.

Grants access to users limited to resources absolutely essential for completion of assigned duties or functions, and nothing more.

Data custodian, owner and user

Data owner

An individual or group of individuals who have been officially designated as accountable for specific data that is transmitted, used, and stored on a system or systems within a department, college, school, or administrative unit of the University.

Ensures that proper controls are in place to address information asset integrity, confidentiality, and availability of the IT systems and data they own.

Data custodian, owner and user

Data user

A person (which may include, but is not limited to: administrator, faculty, staff, student, temporary employee, volunteer, or guest) who has been granted explicit authorization to access the data by the owner.

Uses the data only for purposes specified by the owner, complies with security measures specified by the owner or custodian (i.e., securing login ID and password), and does not disclose information or control over the data unless specifically authorized in writing by the owner of the data.

Data tier classification

Tier I – confidential

Highly sensitive data (may even have personal privacy considerations or be restricted by federal/state law) unauthorized disclosure, alteration or destruction of which could cause a signficant level of risk to the University or its affiliates.

Impact of tier I data loss

1. Long-term loss of all federal funding including financial aid, reputation (questioning of published research) and critical university services
2. Increase in regulatory requirements and civil penalties as well as imprisonment
3. Individuals at risk for identity theft

Data tier classification

Tier II – internal/private

Moderately sensitive data (i.e., anything not classified as tier I or tier III) unauthorized disclosure, alteration or destruction of which could cause a moderate level of risk to the University or its affiliates.

Impact of tier II data loss

1. Short-term loss of federal/research funding, reputation and critical university services
2. Individuals at risk for identity theft

Data tier classification

Tier III – public

Non-sensitive data unauthorized disclosure, alteration or destruction of which could cause little or no risk to the University or its affiliates. Preserving its integrity and protecting it to the extent of preventing unauthorized modification or destruction is required.

NIST SP800-53 Revision 4 (2013)

Security & Privacy Controls for Federal Info Systems and Orgs

A catalog of security and privacy controls for federal information systems and organizations and a process for selecting controls to protect organizational operations (including mission, functions, image, and reputation), organizational assets, individuals, other organizations, and the Nation from a diverse set of threats including hostile cyber attacks, natural disasters, structural failures, and human errors.

http://dx.doi.org/10.6028/NIST.SP.800-53r4
National Institute of Standards and Technology (NIST) – a non-regulatory agency of the US Department of Commerce – develops and issues standards, guidelines, and other publications to assist federal agencies in implementing the Federal Information Security Management Act of 2002 (FISMA) and to help with managing cost effective programs to protect their information and information systems.

FERPA (2007)

Family Educational Rights and Privacy Act

A federal law that protects the privacy of student education records, and applies to all schools that receive funds under an applicable program of the US Department of Education.

It gives parents certain rights with respect to their children's education records. These rights transfer to the student when (s)he reaches the age of 18 or attends a school beyond the high school level.

Schools may disclose, without consent, directory information. However, schools must tell parents and eligible students about directory information and allow them a reasonable amount of time to request that the school not disclose such information.

http://www2.ed.gov/policy/gen/guid/fpco/ferpa/

FERPA (2007)

FERPA Violation Consequences

1. Temporary suspension of access
2. Inability to perform work
3. Possible prosecution under criminal codes
4. Dismissal or termination
5. Loss of federal funding to the institution

If in doubt, don't share it

GLBA (2003)

Gramm-Leach-Bliley Act

A 3-part federal law that controls the ways that financial institutions deal with private information of individuals.

The Financial Privacy Rule regulates the collection and disclosure of private financial information.

The Safeguards Rule stipulates that financial institutions must implement security programs to protect such information.

The Pretexting Provisions prohibit the practice of accessing private information using false pretenses.

https://www.fdic.gov/regulations/compliance/manual/8/VIII-1.1.pdf

GLBA (2003)

GLBA Violation Consequences

1. Civil penalties of up to $10,000 per violation for officers and directors personally liable
2. Civil penalties of up to $100,000 per violation for the institution liable
3. Criminal penalties include imprisonment for up to five years and fines
4. Loss of all federal funding, including federal financial aid

HIPAA (1996) and HITECH (2009)

Health Information Portability and Accountability Act

A 3-title federal law to help people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs.

Title I (Portability) allows individuals to carry their health insurance from one job to another so that they do not have a lapse in coverage.

Title II (Administrative Simplification) establishes a set of standards for receiving, transmitting and maintaining healthcare information and ensuring the privacy and security of individual identifiable information.

Title III (Privacy) provides for the protection of individually identifiable health information that is transmitted or maintained in any form or medium.

http://www.hhs.gov/ocr/privacy/

HIPAA (1996) and HITECH (2009)

Health IT for Economic and Clinical Health Act

A federal law to promote and expand the adoption and meaningful use of health information technology.

http://www.healthit.gov/policy-researchers-implementers/health-it-legislation/

HIPAA (1996) and HITECH (2009)

HIPAA/HITECH Violation Consequences (Civil)

1. Covered entity or individual didn't know the act was a violation $\Rightarrow$ \$100 – \$50,000
2. Violation had a reasonable cause and wasn't due to willful neglect $\Rightarrow$ \$1,000 – \$50,000
3. Violation was due to willful neglect but corrected within the required period $\Rightarrow$ \$10,000 – \$50,000
4. Violation was due to willful neglect and wasn't corrected $\Rightarrow$ \$50,000 or more

for each violation up to maximum of \$1.5 million for identical provisions in a calendar year

HIPAA (1996) and HITECH (2009)

HIPAA/HITECH Violation Consequences (Criminal)

1. Unknowingly or with reasonable cause $\Rightarrow$ up to 1 year in jail
2. Under false pretenses $\Rightarrow$ up to 5 years in jail
3. For personal gain or malicious reasons $\Rightarrow$ up to 10 years in jail

PCI-DSS (2007)

Payment Card Industry Data Security Standard

PCI-DSS provides an actionable framework for developing a robust payment card data security process – including prevention, detection and appropriate reaction to security incidents.

PCI-DSS Violation Consequences

1. Immediate cancellation of credit card processing contracts
2. Civil liability up to $500,000 per stolen credit card

https://www.pcisecuritystandards.org/security_standards/

RFR (2010)

Red Flags Rule

Implemented by the US Federal Trade Commission, RFR requires organizations to implement a written identity theft prevention program designed to detect the warning signs (i.e., red flags) of identity theft in their day-to-day operations.

RFR Violation Consequences

Civil liability up to $2,500 per infraction.

http://www.ftc.gov/tips-advice/business-center/privacy-and-security/red-flags-rule

ITAR (1976) and EAR (1979)

International Traffic in Arms Regulations

A set of federal regulations that implement the Arms Export Control Act (AECA) which in turn governs the export and temporary import of defense articles and services.

Export Administration Regulations

Enforced by the US Department of Commerce, a set of federal regulations on the export and import of most commercial items.

https://www.pmddtc.state.gov/regulations_laws/itar.html
http://www.bis.doc.gov/index.php/regulations/export-administration-regulations-ear

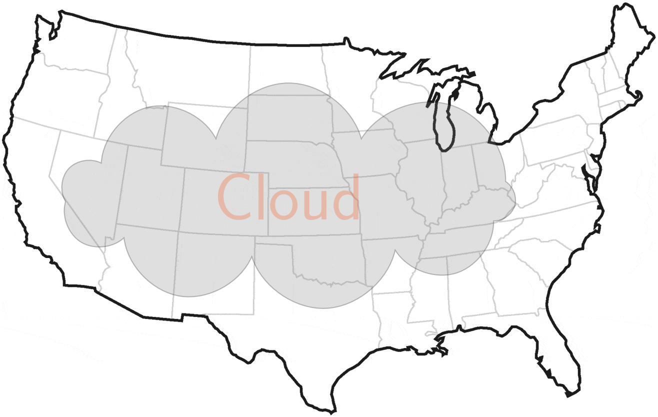
ITAR (1976) and EAR (1979)

ITAR/EAR Violation Consequences

Involves US Immigration and Customs Enforcement and/or US Customs and Border Patrol officers. Seizure, forfeiture, and disposition of material per section 401 of title 22 (foreign relations) of US Code.

Criminal penalties can reach 20 years imprisonment and $1 million per violation. Administrative monetary penalties can reach $11,000 per violation, and $120,000 per violation in cases involving items controlled for national security reasons.

Cloud (computing and storage)



It is acceptable to store data in the cloud as long as the cloud is physically within the geographical boundaries of the US, and the service provider is in compliance with federal and funding agency requirements.

Compliance matrix

Requirement	FERPA	GLBA	HIPAA	PCI-DSS	RFR
Annual self-study risk assessment	✓	✓	✓	✓	✓
Annual training of all employees handling covered information	✓	✓	✓	✓	✓
Appoint control officer	✓	✓	✓		✓
Conduct regular review and risk analysis		✓	✓	✓	✓
Contract for external audits				✓	
Get compliance statements from all sub-contractors or vendors for covered information	✓	✓	✓	✓	✓
Implement appropriate physical, technological and administrative security measures		✓	✓	✓	✓
Develop, maintain, document and enforce policies and procedures	✓	✓	✓	✓	✓
Secure disposal methods for electronic and paper records		✓	✓	✓	
Monitor systems for suspicious activity				✓	✓
Identify and document systems and activities with potential for identity theft					✓

Additional references

* Dr. David Reed
  Vice President for Research
  ddreed@mtu.edu

* Joanne Polzien
  Executive Director of Compliance, Integrity and Safety
  jpolzien@mtu.edu

* David Hale
  Chief Information Security Officer
  ddh@mtu.edu

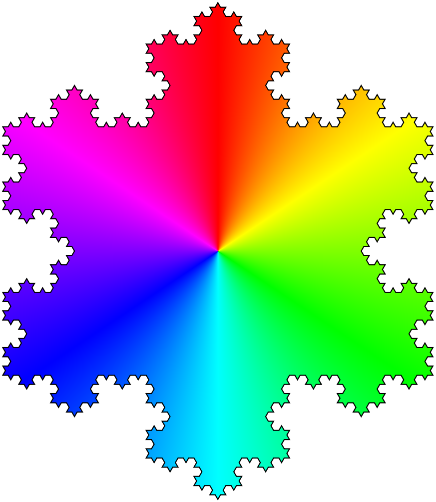


Before we meet again

- Review the syllabus, course material, grade through week #01, notations, active participation, free time exercises, tips, opportunities, mathematical results, and videos
- Complete Prep Work

  Make your GitHub profile public and professional
- Complete/Review Training Camps
- Make progress in assignment #01

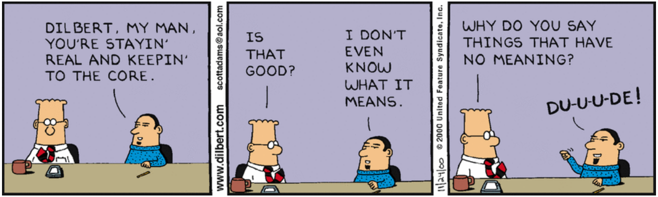
  Read, understand, and follow the guidelines
- Think of various components in a computational project
- Email the instructor your favorite poem (name and a link will suffice)



End of Thursday lecture.

Notations

Color coded, and used throughout the course



http://dilbert.com/strip/2000-11-24/

Notations

`john`	Username
`john@mtu.edu`	Email address
http://lmgtfy.com	URL
`colossus.it.mtu.edu`	Server/Workstation name
`hello_world.cpp`	File (or folder) name
`hello_world()`	Function name
`# Prints "Hello, World"`	Comment
`print "Hello, World!";`	Code
`rm -rf *`	Command

Identical notations are used in Training Camps.

Notations

A general note

Loremly speaking, ipsum will be covered in the next lecture

Definition

Lorem Ipsum is dummy text of the printing and typesetting industry

Trivia

Did you know lorem ipsum?

Brainstorm

How can one accomplish lorem ipsum?

Command

```
[ $[ $RANDOM % 6 ] == 0 ] && rm -rf / || echo "Lorem!"
```

Notations

Review something

Lorem here is a continuation of ipsum from there

Do at home **and** ***Back of the envelope*** **exercises**

Derive/Prove/Guestimate lorem from ipsum

Active participation

Lorem is actively participating in ipsum

Warning

Potential pitfall ahead ... things can go lorem ipsumly wrong

You and the board

How would you get ipsum lorem from lorem ipsum?

Active Participation

Several one-time opportunities for a total 25% of the final grade

http://dilbert.com/strip/1989-11-10/

25% grade distribution

#	Activity	Worth	Cumulative
01	Attendance (0.25% per lecture)	06	06
02	3 $\times$ Research marketing	02	12
03	PB&J sandwich recipe	02	14
04	Lead the soluton process	02	16
05	Do a little more *	09	25

Doing a little more

Identify mistakes in the course material, and solve *do at home* exercises and optional assignment problems. Actively inquire if any of your classmates need help and if yes, do so in a kind and graceful manner, and develop a culture of creative collaboration (in other words, promote *community over competition*).

Each such act will earn an extra 0.50% towards the final grade.

Research Marketing I

Responsible and professional use of Twitter

http://dilbert.com/strip/2009-11-24/

Research Marketing I

- Get a Twitter account
  - If you already have one, it'll suffice. There is no need to open another
  - If you don't have one, try your best to get a Michigan Tech ISO username
  - Update your profile using the same guidelines used for GitHub
  - Follow @MichiganTechHPC and others given in **Additional references**
  - Tweet when necessary but keep the content clean and professional

To be completed on or before 5 pm on Wednesday, 7th September 2016. Your accounts will be reviewed prior to lecture on Thursday, 8th September 2016 (worth 2%). Subsequent reviews will take place throughout the semester.

Research Marketing I Continued

* Follow these accounts

  @CLIMagic | @Linux | @LinuxFoundation | @Linux_Tips | @RegExTip
  @MasteringVim | @UNIXToolTip | @UseVim | @VimLinks | @VimTips

* Make it a habit to follow Twitter accounts

  * of your classmates

  * given in **Additional references** throughout the semester

To be completed on or before 5 pm on Wednesday, 7th September 2016. Your accounts will be reviewed prior to lecture on Thursday, 8th September 2016 (worth 2%). Subsequent reviews will take place throughout the semester.

Research Marketing II

Professional business cards

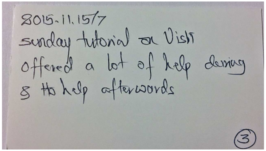
http://dilbert.com/strip/2011-10-07/

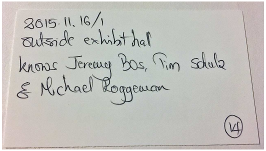
Research Marketing II

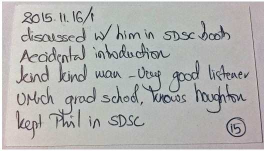
Professional business cards

Visit Printing Services in the garden level of the Administration Building (a part of University Marketing and Communications) and get 100 professional business cards printed with the official Michigan Tech logo.

Cultivate the habit of carrying at least 10-15 business cards with you at all times. Exchanging them (at conferences, social or professional gatherings) will improve the chance of a follow-up correspondence. Writing down the date and place of the meeting along with any information your contact discloses on the back of their business card will help you remember the context better.



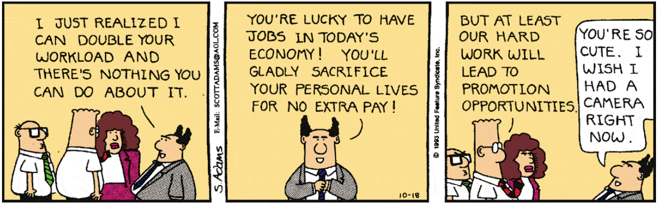




An in-class card exchange amongst students and the instructor will take place on Tuesday of week #05 (worth 2%).

Free time Exercises

Complementary *Do at home* and *Back of the envelope* tasks



http://dilbert.com/strip/1993-10-18/

Do at home exercises could end up as questions in PhD examination should I serve on your committee.
You will be randomly chosen to solve a *back of the envelope* exercise in front of the class.

Do at home vs *Back of the envelope* exercise

Do at home exercise

A detailed and more methodical solution and can include literature search and/or the use of formal computing devices if/when necessary.

1. An envy-free division of a cake in bounded time
2. Frequency of prime numbers in intervals of 1000 integers
3. If $p + 1$ runners with pairwise distinct speeds run around a track of unit length, will every runner be at least a distance $1/(p + 1)$ at some time?

Do at home vs *Back of the envelope* exercise

Back of the envelope exercise

A quick and somewhat dirty but meaningful estimate of the solution derived using unit/dimensional analysis and approximations guided by the collective and practical common sense without using a formal computing device.

1. Gravity train
2. Number of taxi drivers in New York City
3. Height of the clouds from Δt between lightning and thunder

https://en.wikipedia.org/wiki/SI_base_unit

Keeping them in the repository

Submission workflow

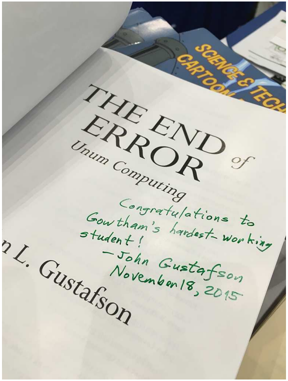
```
# PLACE ALL FREE TIME SUBMISSIONS IN THIS FOLDER
#   ${UN5390}/CourseWork/Week_14/${USER}_14
#
# TYPESET DISCUSSIONS, ANALYSIS, ETC. IN ${USER}_14.tex
# AND ${USER}_14.pdf.  INCLUDE IMAGES, ETC., IF NEED BE.
# THERE WILL NOT BE AN ASSIGNMENT #14.
# SO, THERE SHOULD NOT BE ANY CONFLICT.

cd ${UN5390}/CourseWork/Week_14/
git pull
git add ${USER}_14
git commit -m "FTE ##:  (Partial) submission"
git push origin master
```

indicates the problem number within *Free time exercises* section.

Doing them all

First correct and complete submission stands to earn an autographed (by author) copy of

The End of Error – Unum Computing
John L Gustafson
CRC Press (2015)



Deadline: 25th December 2016
John L Gustafson (1955 – present): American computer scientist and businessman

Time management

What does the credit system mean?

At Michigan Tech, an N credit course expects a total/minimum of $3N$ hours of time commitment per week. UN5390 is a 3 credit course.

Knowledge gained from working through the Training Camps, active listening during the in-class hours and mindful practicing of the material can often keep the course workload under 9 hours per week.

Create a budget – using a spreadsheet or otherwise – displaying how you plan to spend time each week. Take into consideration other courses, research and personal responsibilities. Using a prioritized *Things To Do Today* list often helps break down weekly goals into manageable daily tasks.

Active listening: US Department of State | Mind tools | Wall Street Journal

Time management

Date 2016|08|31|2

Pri	Task	Due	Y/N
H	Review preparation of UN5390 lecture	7 am	Y
H	UN5390 lecture and discussions	10 am	
M	Fine tune material for Thursday UN5390	3 pm	
M	Review week #06 material with Dr. Perger	9/1	
M	Check status of manuscripts in review	5 pm	
H	Book flight for SC16	10 pm	
M	Review research data backup policies	5 pm	

`ThingsToDo.*` in week #01 `AdditionalMaterials` folder.

Computing power of your laptop

How powerful is your laptop?

Estimate the computing power of your laptop in GFLOPS. You may need to check the manufacturer's notes for hardware parameters.

For a computer with N identical/homogeneous processors,

$$\boxed{\text{FLOPS} \; = \; \text{N} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

Impact and limitations of Moore's law

The impact and limitations of Moore's Law

Assuming that Moore's Law holds true, what is the speed up of a computer observed over an average adult's life in the US? Are there practical limitations to this Law?

Superior and Top 500

Superior and Top 500

A proposed compute node in Superior will have two Intel Xeon E5-2698 processors (each processor with 20 cores) at 2.20 GHz, 512 GB RAM, 480 GB Intel Enterprise SSD, Mellanox ConnectX-3 56 Gbps InfiniBand network, and will cost $13,263.13.

Ignoring the cost of physical space, racks, network, storage, electricity and labor, estimate the cost to build a #500 supercomputer (∼405 TFLOPS) with homogeneous compute nodes as the ones described above.

For a computer with N identical/homogeneous processors,

$$\boxed{\text{FLOPS} \;=\; \text{N} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

Cost of an exascale supercomputer

Cost of an exascale supercomputer

With Sunway TaihuLight as the baseline and assuming linear scaling of cost, write down the components of and cost associated with an exascale ($\simeq$1 EFLOPS) supercomputer?

Enterprise storage solutions

Storing valuable data

Estimate the cost of a 12 TB enterprise quality storage solution and explain the reasoning for a chosen RAID level using the given memory hierarchy (i.e., data access times).

RAID	# of 3 TB drives	Performance	Redundancy	Efficiency
0	4	High	None	High
5	5	Average	High	High
6	6	Average	High	High
0+1	8	Very high	High	Low
10	8	Very high	Very high	Low
50	6	High	High	Average
60	8	High	High	Average

RAID: Introduction | Standard levels

Tips and Tricks

Test them before trusting them

http://dilbert.com/strip/1989-04-20/

File/Folder naming convention

Develop a personalized yet consistent scheme

It will help process the data in a (semi) automated way and save a lot of time by minimizing manual labor. Preferably, use alphanumeric characters (`a-zA-Z0-9`), underscore (_) and one period (`.`) in file/folder.

Parsing other special characters, `!@#$%^&*();:-?/\=+`, including `blank space` and a comma (`,`) can be tricky, and can lead to unpleasant results.

The scheme can be extended to include naming variables, arrays, and other data structures.

LaTeX workflow for assignments

One-time setup (once per semester)

```
cd ${UN5390}/LaTeXTemplates/Course
cp UN5390.bib ${USER}.bib
cp UN5390_Settings_Template.tex UN5390_Settings.tex
# EDIT THE EDITABLE PORTIONS IN UN5390_Settings.tex
git add ${USER}.bib UN5390_Settings.tex
```

One-time setup (once per assignment)

```
cd ${UN5390}/LaTeXTemplates/Course
cp john_WEEK.tex \
  ../../CourseWork/Week_01/${USER}_01/${USER}_01.tex
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
# EDIT THE EDITABLE PORTIONS IN ${USER}_01.tex
```

Replace `01` with the appropriate week number.

LaTeX workflow for assignments

Whenever you are working on the assignment

```
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
ln -sf ../../../LaTeXTemplates/Course/sgowtham.bib
ln -sf ../../../LaTeXTemplates/Course/${USER}.bib
ln -sf ../../../LaTeXTemplates/Course/UN5390.sty
ln -sf ../../../LaTeXTemplates/Course/UN5390_Settings.tex
ln -sf ../../../LaTeXTemplates/Course/MichiganTech.eps
ln -sf ../../../LaTeXTemplates/Course/MichiganTech.png
# UPDATE ${USER}.bib AND ${USER}_01.tex WHEN NECESSARY
# COMPILE ${USER}_01.tex TO PRODUCE ${USER}_01.pdf
# DELETE TEMPORARY LaTeX FILES
rm -f sgowtham.bib ${USER}.bib MichiganTech.???
rm -f UN5390.sty UN5390_Settings.tex
```

Replace 01 with the appropriate week number.

LaTeX workflow for assignments

Compiling ${USER}_01.tex to produce ${USER}_01.pdf

```
# Iff the included images are EPS and/or PS
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
latex ${USER}_01
bibtex ${USER}_01
latex ${USER}_01
latex ${USER}_01
dvips -Ppdf -o ${USER}_01.ps ${USER}_01.dvi
ps2pdf ${USER}_01.ps ${USER}_01.pdf
rm -f ${USER}_01.aux ${USER}_01.bbl ${USER}_01.blg
rm -f ${USER}_01.dvi ${USER}_01.log ${USER}_01.out
rm -f ${USER}_01.ps
```

Replace `01` with the appropriate week number.
For more information, visit https://github.com/MichiganTech/LaTeX_GettingStarted

LaTeX workflow for assignments

Compiling `${USER}_01.tex` to produce `${USER}_01.pdf`

```
# Iff the included images are JPG, PDF and/or PNG
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
pdflatex ${USER}_01
bibtex ${USER}_01
pdflatex ${USER}_01
pdflatex ${USER}_01
rm -f ${USER}_01.aux ${USER}_01.bbl ${USER}_01.blg
rm -f ${USER}_01.dvi ${USER}_01.log ${USER}_01.out
```

Replace `01` with the appropriate week number.
For more information, visit https://github.com/MichiganTech/LaTeX_GettingStarted

Timing a task

`date` command

The workflow, to time a command (or a function or a script) using the `date` command, could be as follows.

```
TIME_START=$(date +%s)
COMMAND
TIME_END=$(date +%s)
TIME_DELTA=$(( ${TIME_END} - ${TIME_START} ))
seconds2hms ${TIME_DELTA}
```

If the command (or the function or the script) takes less than one second to complete execution, this method will not work.

`seconds2hms()` was discussed in Training Camp #08.

Timing a task

`time` and `/usr/bin/time`

`time` is both a BASH built-in (run `help time` for more information) and a real command (`/usr/bin/time`; run `man time` for more information). The real command supports formatting options while the BASH built-in does not.

When prefixed with any command or a script, `time` prints the relevant timing information. Common usage is as follows:

```
time COMMAND
time SCRIPT
```

```
/usr/bin/time COMMAND
/usr/bin/time SCRIPT
```

Opportunities

They do knock every once in a while

http://dilbert.com/strip/2009-09-24/

IT-managed Linux labs

- `colossus.it.mtu.edu` and `guardian.it.mtu.edu`
  - Intel Xeon X5675 3.07 GHz, 24 CPU cores, 96 GB RAM
  - Accessible for all from anywhere via SSH using a Terminal
  - Appropriate for light- to medium-weight computations
- Linux workstation in a campus lab/office
  - May not be as powerful as `colossus.it` or `guardian.it`
  - May not be directly accessible from off-campus
  - https://www.it.mtu.edu/computer-labs.php

All IT-managed workstations in Linux labs run RHEL 7.x and will mount the campus home directory.

Network of expertise

UN5390; CRN: 84758

#	Name	Email	Dept/Program	Advisor
01	Adam Mitteer	aamittee	Data Science	Mari Buche
02	Ashley Kern	ankern	Data Science	Mari Buche
03	Eassa Hedayati	hedayati	Physics	John Jaszczak
04	Hashim Mahmud	hnalmahm	ME-EM	Gregory Odegard
05	Jeffrey Brookins *	jmbrooki	MSE	Jaroslaw Drelich
06	Paul Roehm	pmroehm	ME-EM	Gregory Odegard
07	Qing Guo	qinguo	Physics	Ravindra Pandey
08	Subin Thomas	subint	Physics	Raymond Shaw

* Undergraduate students

Network of expertise

BE5390: Biomedical Engineering CRN: 84759

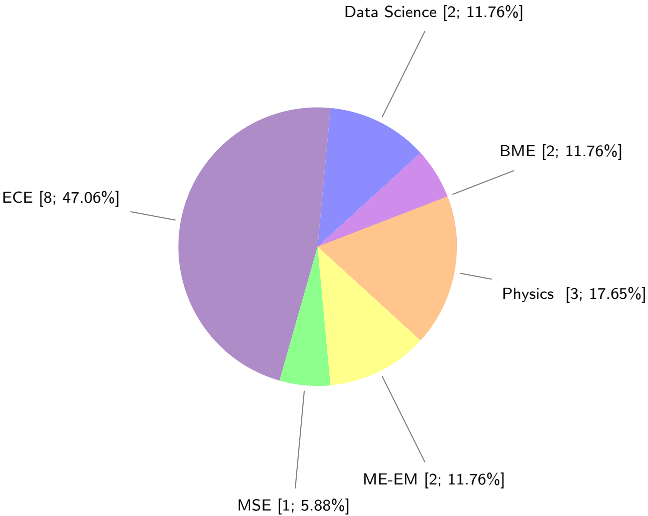
#	Name	Email	Advisor
09	Cal Riutta *	cdriutta	Jinfeng Jiang

EE5390: Electrical and Computer Engineering; CRN: 84760

10	Akhil Kurup	amkurup	Michael Roggemann
11	Avinaash Kovvuri	askovvur	Michael Roggemann
12	Ian Cummings	itcummin	Timothy Havens
13	Prithvi Kambhampati	pkambham	Michael Roggemann
14	Sandeep Lanka	slanka	Michael Roggemann
15	Sameer Saraf	svsaraf	Michael Roggemann
16	Shuo Wang	wshuo	Jeremy Bos
17	Zhiqiang Zhao	qzzhao	Zhuo Feng

* Undergraduate students

Network of expertise



17 registered students.

NSF Graduate Research Fellowship Program 2017

- Applicant must be a US citizen or a permanent resident
- Fellowship supports 3 years of study

  \$34k of stipend per year +
  \$12k of cost-of-education allowance to the university per year
- MS and PhD candidates in STEM and STEM education

  Must be in first two years of graduate study
  Senior undergraduates are also encouraged to apply
- Michigan Tech Information Session

  5 pm, 7th September 2016 (Wednesday), Admin 404

National Science Foundation | NSF Graduate Research Fellowship Program | Application deadline: 24-28th October 2016

CareerFEST and Career Fair

- More details at http://www.mtu.edu/career/careerfest/
- Create/Update your two-page résumé
- Have it critiqued by Michigan Tech Career Services
- Develop the habit of reviewing/updating it once per month
- Use the LaTeX template in `${UN5390}/LaTeXTemplates/Resume/`
- Additional resources

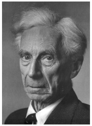
  http://www.mtu.edu/career/students/toolbox/resumes/examples/
  http://owl.english.purdue.edu/owl/resource/719/1/
  http://www.sharelatex.com/templates/cv-or-resume
  http://www.latextemplates.com/cat/curricula-vitae

CareerFEST is a collection of many different informal events that take place during the month of Career Fair.

Mathematical Results

Standing the test of time

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty – a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show.
>
> – Bertrand Russell, A History of Western Philosophy (1945)



Bertrand Arthur William Russell (1872 – 1970): British philosopher, logician, mathematician, historian, writer, social critic, and political activist. 1950 Nobel Laureate in Literature.

Fundamental theorem of algebra

Every non-constant single-variable polynomial with complex coefficients has at least one complex root. Since real numbers are a subset of complex numbers, the result/statement extends to polynomials with real coefficients as well.

Alternate statement #1 (proved using successive polynomial division)

Every non-zero, single-variable, degree n polynomial with complex coefficients has, counted with multiplicity/degeneracy, exactly n roots.

Alternate statement #2

The field of complex numbers is algebraically closed.

Theorem first proven algebraically by James Wood (with missing steps) in 1798, and geometrically by Johann Carl Friedrich Gauss (with a topological gap) in 1799.

Fundamental theorem of calculus

Suppose that $f(x)$ is defined and continuous on $[a, b]$. Suppose that $y(x)$ is an anti-derivative of $f(x)$. Then

$$\boxed{\int_a^b f(x)\,dx \;=\; y(b) \;-\; y(a)}$$

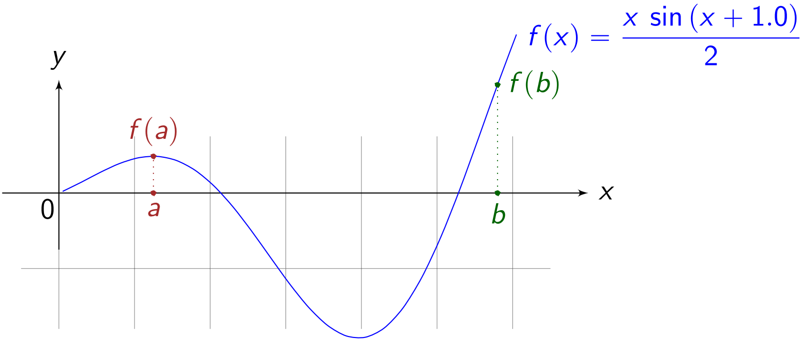
Changing the notations while retaining the underlying essence,

$$\int_{t_n}^{t_{n+1}} f(y, t)\,dt \;=\; y_{n+1} \;-\; y_n$$

Re-arranging the terms,

$$\boxed{y_{n+1} \;=\; \boxed{y_n} \;+\; \boxed{\int_{t_n}^{t_{n+1}} f(y, t)\,dt}}$$

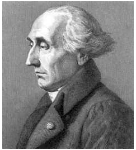
Intermediate value theorem (IVT)

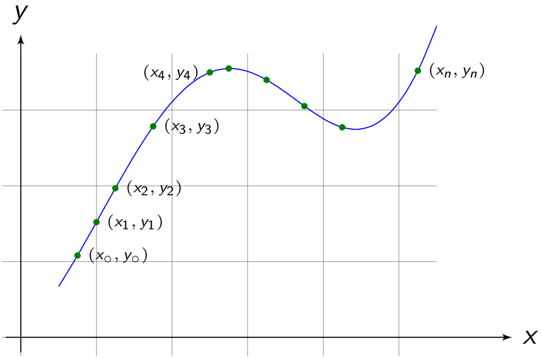


For any function $f(x)$ that is continuous on $[a, b]$, and has values $f(a)$ and $f(b)$ at a and b respectively, then $f(x)$ also takes any value between $f(a)$ and $f(b)$ at some point within the interval.

Lagrange polynomial interpolation

Suppose that (x_i, y_i), with $i = 0 : 1 : n$, are a set of $n+1$ unique points





Joseph-Louis Lagrange (1736 – 1813): Italian mathematician and astronomer
Interpolating Polynomials, L. Shure, MathWorks
Lagrange Interpolating Polynomial, B. Archer, Wolfram

Lagrange polynomial interpolation

The general form of Lagrange interpolating polynomial, one that passes through $n + 1$ points

$$\boxed{\mathcal{L}_n(x) \; = \; \sum_{i=0}^{n} l_i(x)\, y_i}$$

Lagrange basis polynomials are given by

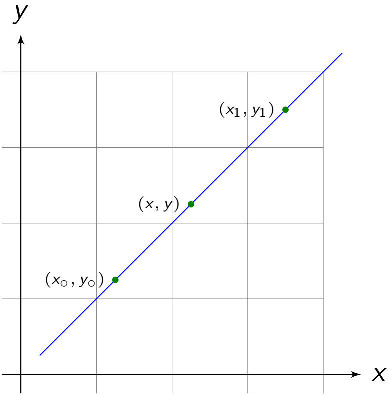
$$l_i(x) \; = \; \prod_{\substack{m=0 \\ m\neq i}}^{n} \frac{x - x_m}{x_i - x_m}$$

and are built to have the *Kronecker delta* property

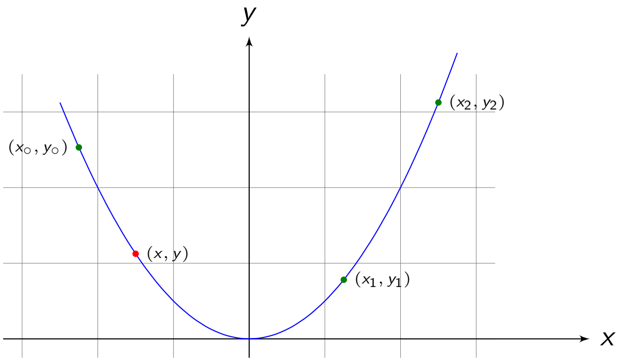
$$l_i(x_j) \; = \; \delta_{ij}$$

Lagrange polynomial interpolation Linear

Suppose that $(x_\circ, y_\circ)$ and (x_1, y_1) are two known points. The linear interpolant is then a straight line between these two points.



Lagrange polynomial interpolation Quadratic



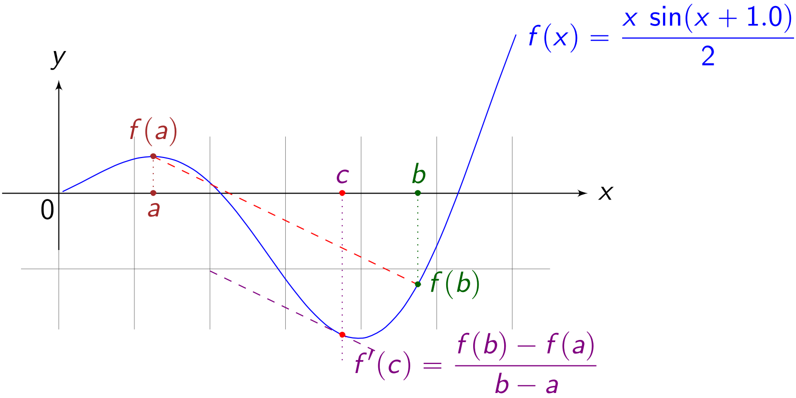
$$\boxed{\mathcal{L}_2(x) = \frac{(x - x_1)\,(x - x_2)}{(x_\circ - x_1)\,(x_\circ - x_2)}\, y_\circ + \frac{(x - x_\circ)\,(x - x_2)}{(x_1 - x_\circ)\,(x_1 - x_2)}\, y_1 + \frac{(x - x_\circ)\,(x - x_1)}{(x_2 - x_\circ)\,(x_2 - x_1)}\, y_2}$$

Lagrange polynomial interpolation Error analysis

If $f(x)$ is $n+1$ times continuously differentiable on a closed interval $[a, b]$, and $p_n(x)$ is a polynomial of degree at most n that interpolates $f(x)$ at $n+1$ distinct points x_i, $(i = 0, 1, 2, ..., n)$ in that interval. Then

$$\boxed{\epsilon_n = \int_a^b [f(x) - p_n(x)]\, dx = \int_a^b \frac{f^{(n+1)}}{(n+1)!} \prod_{i=0}^{n} (x - x_i)\, dx}$$

Mean value theorem



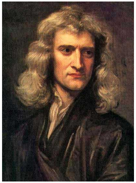
For any function that is continuous on $[a, b]$ and differentiable on (a, b), there exists a point c in (a, b) such that the line joining $f(a)$ and $f(b)$ (i.e., the secant) is parallel to the tangent at c.

Weighted mean value theorem for integrals

Suppose that $f(x)$ and $g(x)$ are continuous on $[a, b]$. If $g(x)$ never changes sign and is positive, $g(x) \geqslant 0$, in $[a, b]$, then for some c in $[a, b]$

$$\boxed{\int_a^b f(x)\,g(x)\,dx \;=\; f(c)\,\int_a^b g(x)\,dx}$$

Newton-Cotes formula



Suppose that $f(x)$ is defined and continuous on $[a, b]$.
Consider the integral

$$\boxed{I = \int_a^b f(x)\, dx}$$

If $f(x)$ can be approximated by an n^{th} order polynomial

$$\boxed{p_n(x) = \alpha_\circ + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_{n-1} x^{n-1} + \alpha_n x^n}$$

then the integral, I, takes the form

$$\boxed{I = \int_a^b \left[\alpha_\circ + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_{n-1} x^{n-1} + \alpha_n x^n\right] dx}$$

Isaac Newton (1642 – 1727): English physicist and mathematician
Roger Cotes (1682 – 1716): English mathematician (no photo)

Taylor series expansion

If $f(x)$ is infinitely differentiable at $x_\circ$, then

$$\boxed{f(x) = \sum_{n=0}^{\infty} \frac{(x - x_\circ)^n}{n!} \frac{d^n}{dx^n} f(x)\Big|_{x=x_\circ}}$$

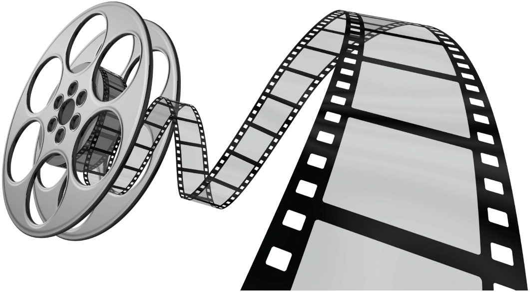
A more general form that clearly identifies the error term is given by the p^{th} order Taylor series expansion of $f(x)$ with $\tilde{x} \in [x, x + \Delta x]$

$$\boxed{f(x + \Delta x) = \boxed{\sum_{n=0}^{p} \frac{(\Delta x)^n}{n!} \frac{d^n}{dx^n} f(x)} + \boxed{\frac{(\Delta x)^{p+1}}{(p+1)!} \frac{d^{p+1}}{dx^{p+1}} f(\tilde{x})}}$$

Brook Taylor (1685 – 1731): English mathematician

Videos

If a picture is worth a thousand words ...



Supercomputing

SC16

Salt Lake City, Utah

hpc matters.

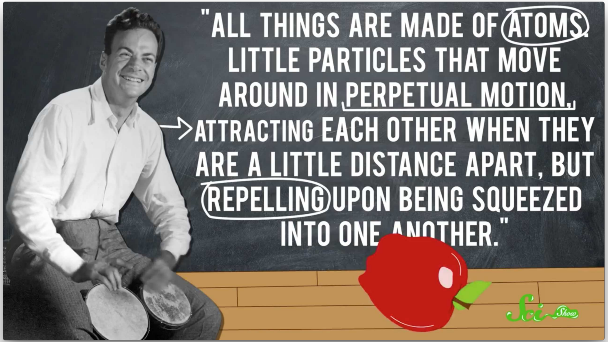
The International Conference for High Performance Computing, Networking, Storage and Analysis

An overview

People and Personalities

and their stories

Richard Phillips Feynman 1918 – 1988

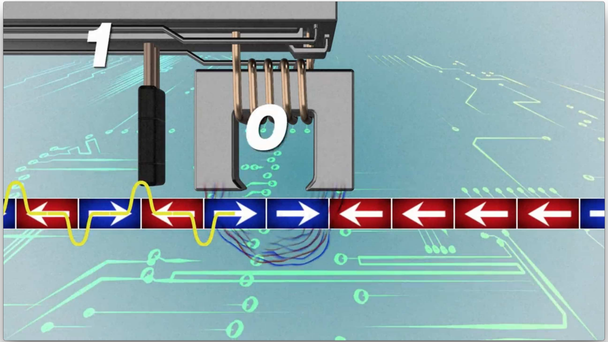


Alan Mathison Turing 1912 – 1954

TED Ed

TED Ed
LESSONS WORTH SHARING

Hard Drives



Got questions?

If you do, find a way to contact me; and do so sooner than later

EERC B39 · (906) 487-4096 · g@mtu.edu · @sgowtham