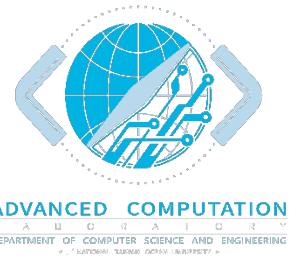


Computer programming and Data Science

William Hsu
Advanced Computation Laboratory
Department of Computer Science and Engineering
Department of Environmental Biology and Fisheries Science
National Taiwan Ocean University

COURSE INFO

- Course name: Computer programming and Data Science (程式設計與資料處理)
  - Credit hours: 2
  - Course ID: B31012SX
  - Class hour: Thr 10:20AM~12:10PM
  - Course website: http://www.deepsea9.taipei/wwyhsu/?page_id=1732
  - Lecturer: William Hsu
  - Office hours: Thursday all day
  - Lab location: CSE R405
  - Email: wwyhsu@ntou.edu.tw
  - Office phone: 6657
- Teaching assistant:
  - 李依柔: li00131022@gmail.com
  - 胡瑞興: howard.rhhu@gmail.com

COURSE INFO

This course accompanies the main course.

- Course name: Computer Programming and Data Science Lab (程式設計與資料處理實習)
  - Credit hours: 1 (2 hours)
  - Course ID: B31013GE
  - Class hour: Thr 1:10PM~3:00PM
  - Course website: http://www.deepsea9.taipei/wwyhsu/?page_id=1732
  - Lecturer: William Hsu
  - Office hours: Thursday all day
  - Lab location: CSE R405
  - Email: wwyhsu@ntou.edu.tw
  - Office phone: 6657
- Teaching assistant:
  - 李依柔: li00131022@gmail.com
  - 胡瑞興: howard.rhhu@gmail.com

Who am I?

許為元 (William W.-Y. Hsu)

資訊工程學系

環境生物與漁業科學系

先進計算實驗室(Advanced Computation Laboratory)

巨量資料科學, 雲端系統, 衛星遙測, 生物資源評估, 系統程式, 財務工程演算法

APCS 團隊

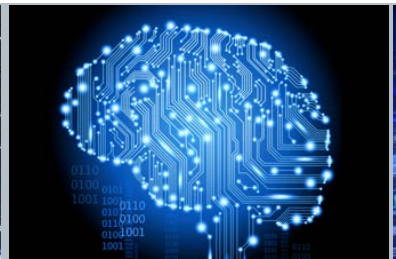
IOI國際資訊奧林匹亞國家教練團

資工系競賽程式團隊教練

Introduction

The basics of computer science



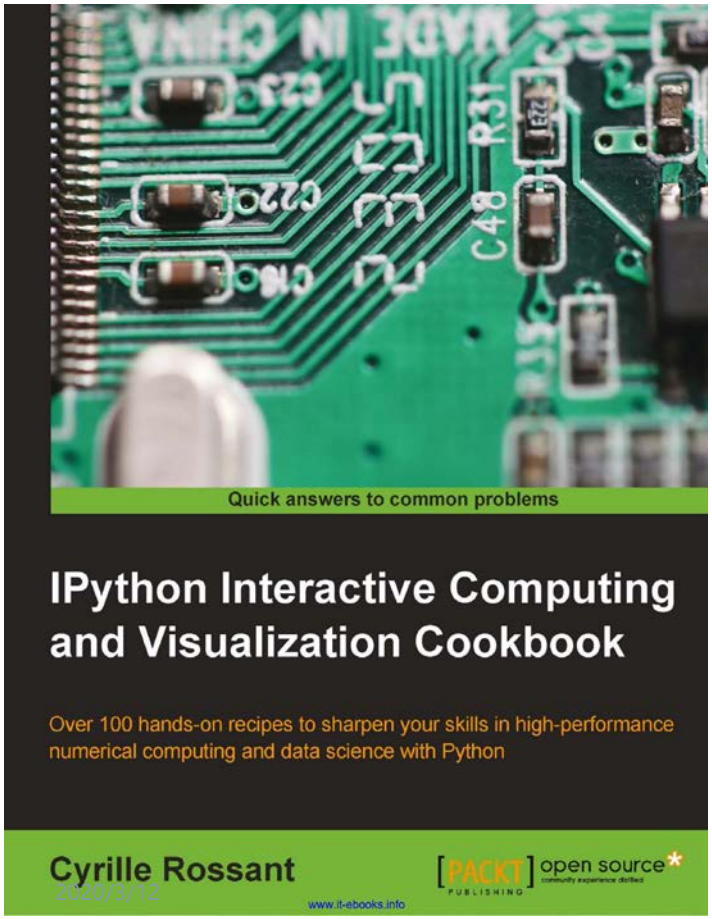



本學期課程摘要 (預計)

- Week 01: 計算機概論總綱
- Week 02: 思維運算
- Week 03: 電腦基本組成, Python簡介
- Week 04: 作業系統導論, Python: Variables and Types
- Week 05: Python: String formatting, Basic String Operations
- Week 06: Python: Conditions
- Week 07: Python: Loops
- Week 08: Python: Functions
- Week 09: 期中考
- Week 10: Python: Dictionaries
- Week 11: Python: 數值運算 Numeric Computing is Fun
- Week 12: Python: 資料科學 Numpy Arrays, Pandas Basics
- Week 13: Python: Statistics, Machine Learning and Data Science
- Week 14: Python: Mathematics, Physics, Chemistry, Biology
- Week 15: Python: Economics and Finance
- Week 16: Python: Earth Science and Geo-Spatial data
- Week 17: Python: Data visualization and plotting
- Week 18: 期末考

參考課本

- IPython Interactive Computing and Visualization Cookbook, Second Edition, Packt Publishing, 2018.
  - Open source!



Tentative Policies

- Class policy:
  - Ask questions anytime!
  - You may leave anytime for WC without asking.
    - It's human rights.
  - Refrain from using any messaging devices.
  - Phones to vibration.
  - Eating and drinking is ALLOWed, however…
    - If anyone leaves trash behind, this policy will be canceled. Penalty will be applied from this point on.
  - Name calling will not be conducted. It is your duty to make up yourself if you skip classes.
    - No need to ask for leave.
    - However, no make up exams will be given! (Unless you have a good reason)
  - Refrain from sleeping in classes.
    - Get sleep in your bed for maximum quality.
    - If it happens too much, I will start to deduct point adjustments of final scores.
  - Refrain from chatting too loud of personal matters.
    - I will remove you from the classroom because you are interfering with the class.

Tentative Policies

- Grading:
  - 60% 作業
    - 部分會在實習內完成
  - 20% 期中考
  - 20% 期末考
  - The university ethics code must be complied.

Tentative Policies

- Grading – Extra credits:
  - Extra credit policy for attending public contests:
    - ITSA (月賽, 中文題為主): 1% +1% per solved problem.
      - ITSA 每年五月另舉辦桂冠挑戰大賽! 5% + 2% per solved problem. (所以計概課程碰不到)
    - CPE: 2% + 1% per solved problem
    - PTC (月賽,以英文題為主): 2% + 2% per solved problem.
    - TOPC: 2%+2% per solved problem.
    - NCPC (教育部大專程式設計競賽, ACM前導賽): 3% + 2% per solved problem.
    - ACM-ICPC Regionals: 5% + 2% per solved problem.
    - ACM-ICPC World finals: 再說吧~

大學生跟高中生很不同!

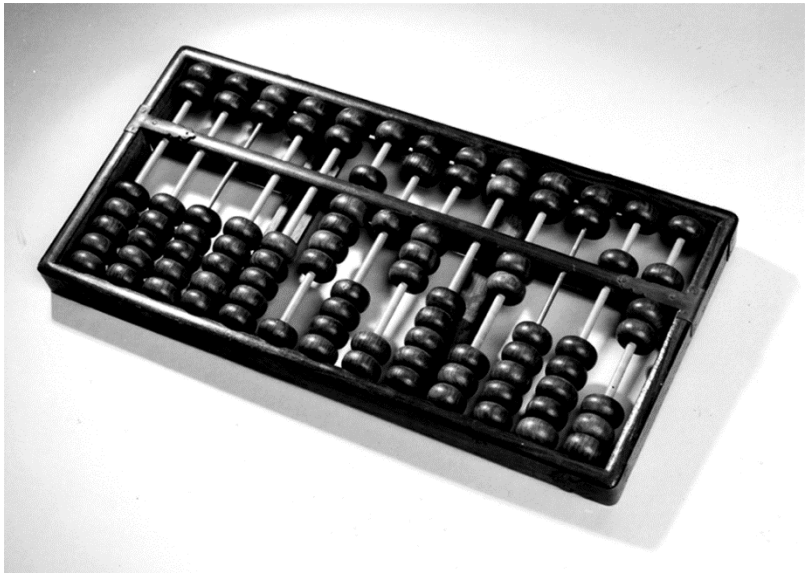
- 我只會三種教法:
  - 1. 台大教法 – 放牛吃草
  - 2. 交大教法 – 硬, 很硬
  - 3. 美國西北大學教法 - 翻轉教室 (地上滾)

History of Algorithms

- The study of algorithms was originally a subject in mathematics.
- Early examples of algorithms
  - Long division algorithm
  - Euclidean Algorithm
- **Gödel's Incompleteness Theorem**: Some problems cannot be solved by algorithms.

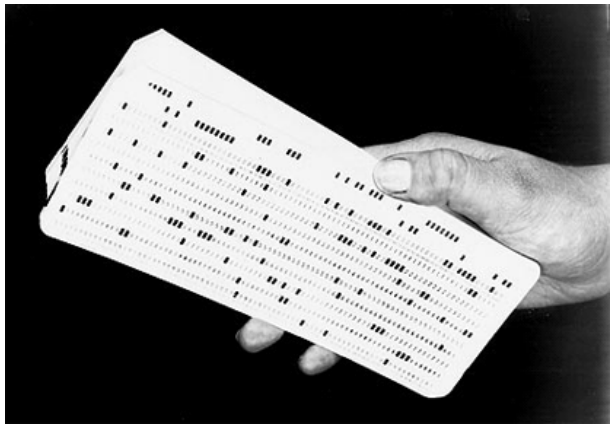
History of Computing

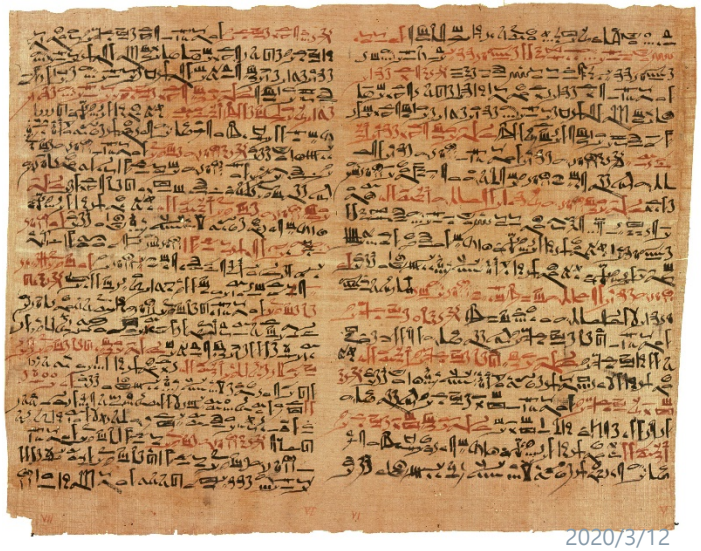
- Early computing devices
  - Abacus: positions of beads represent numbers
  - Gear-based machines (1600s-1800s)
    - Positions of gears represent numbers
    - Blaise Pascal, Wilhelm Leibniz, Charles Babbage



Early Data Storage

- Punched cards
  - First used in Jacquard Loom (1801) to store patterns for weaving cloth
  - Storage of programs in Babbage's Analytical Engine
  - Popular through the 1970's
- Gear positions






bordeno
FLOPPY DISK
5.25" HD.DS

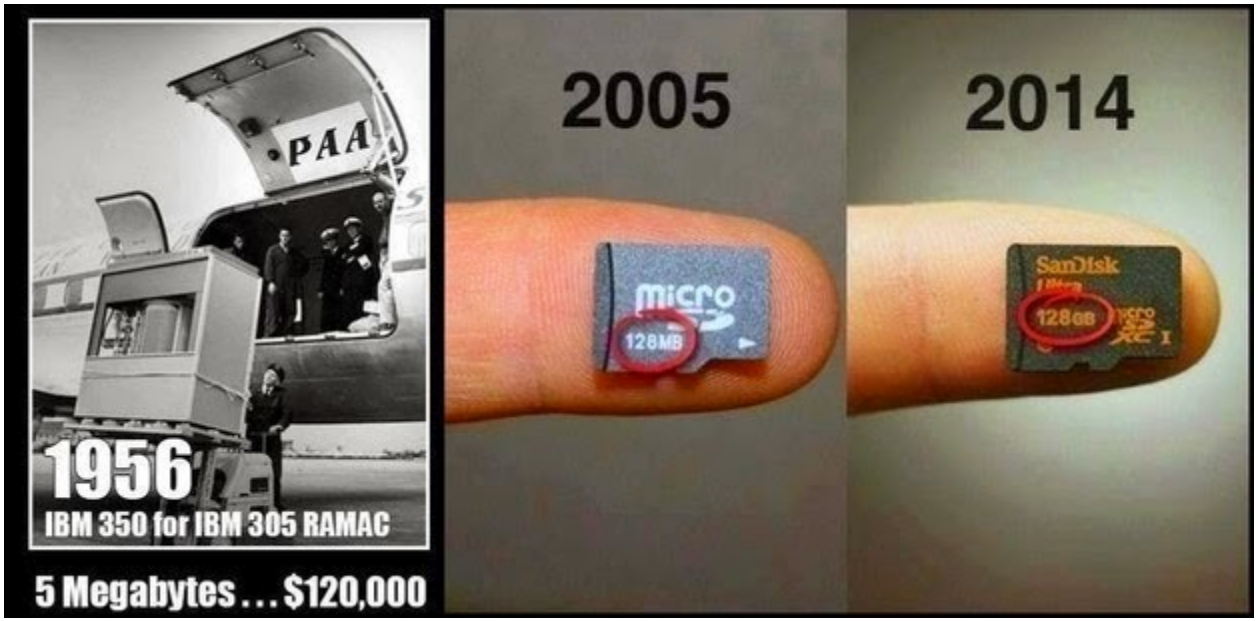

ZIP
100MB
zip


zip
100
NOT FOR INDIVIDUAL RESALE
PC100


Pdo
MO-DISK
640
KONICA MINOLTA


TDK
DC 6250
FUJIFILM
DG6-150


PIONEER



What 256MB looks like in...

1995

2007


Lexar

5 MB hard drive in 1956

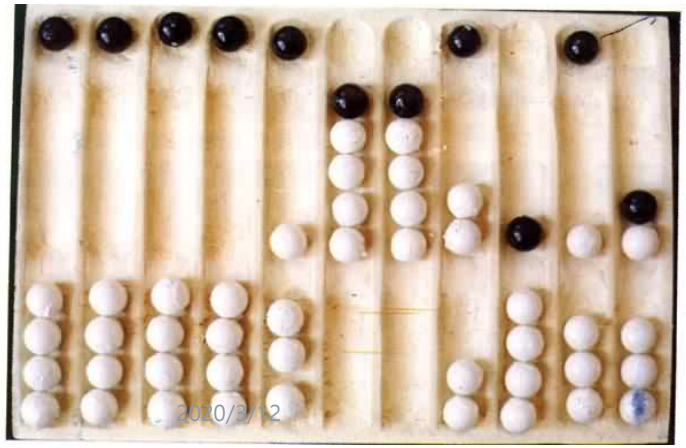
128 GB pen drive in 2017

Early Computers

- Based on mechanical relays
  - 1940: Stibitz at Bell Laboratories
  - 1944: Mark I: Howard Aiken and IBM at Harvard
- Based on vacuum tubes
  - 1937-1941: Atanasoff-Berry at Iowa State
  - 1940s: Colossus: secret German code-breaker
  - 1940s: ENIAC: Mauchly & Eckert at U. of Penn.

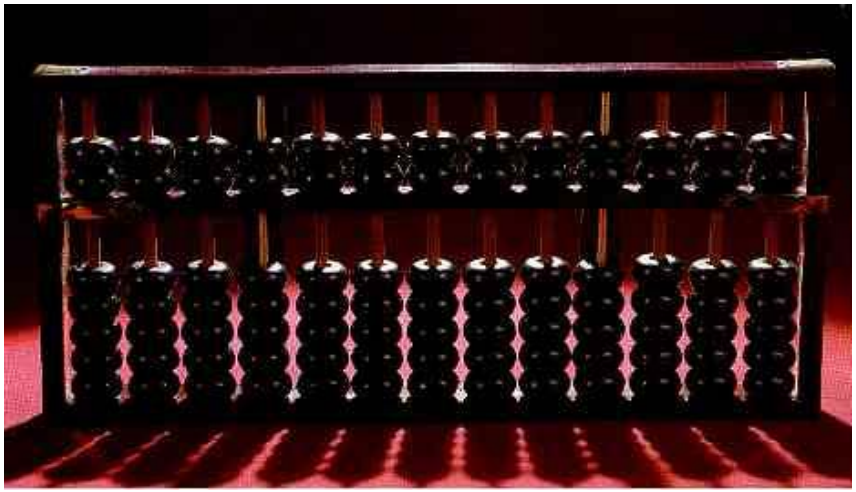
Abacus

- The ***abacus*** was an early aid for mathematical computations.
- Its only value is that it aids the memory of the human performing the calculation.
- A skilled abacus operator can work on addition and subtraction problems at the speed of a person equipped with a hand calculator (multiplication and division are slower).



Abacus

- The abacus is often wrongly attributed to China.
- In fact, the oldest surviving abacus was used in 300 B.C. by the Babylonians.
- The abacus is still in use today, principally in the far east.



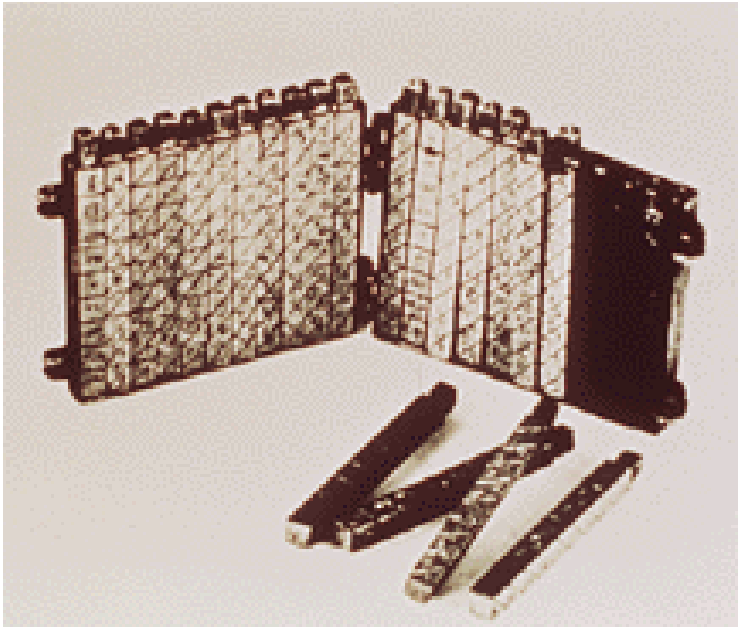
John Napier

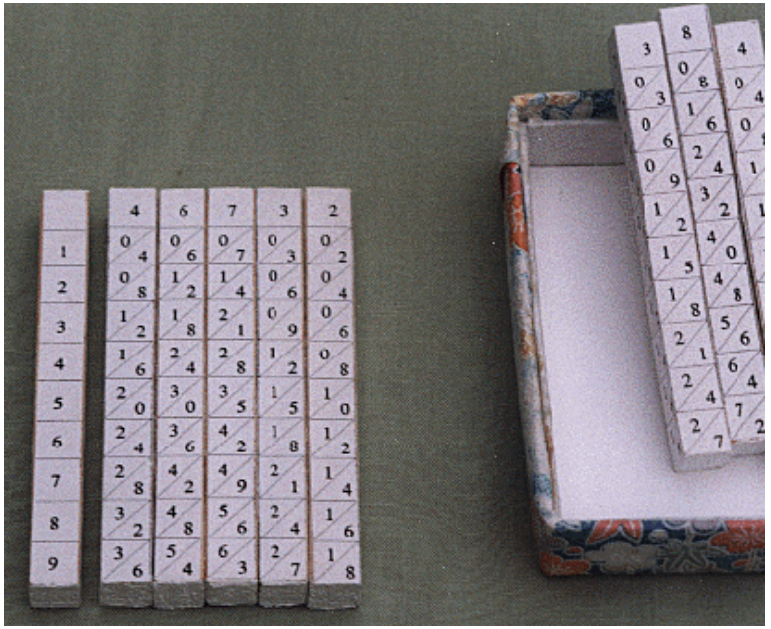
- In 1617 an eccentric (some say mad) Scotsman named John Napier invented ***logarithms***, which are a technology that allows multiplication to be performed via addition

$$\log_2 x = 5$$

Napier’s Bones

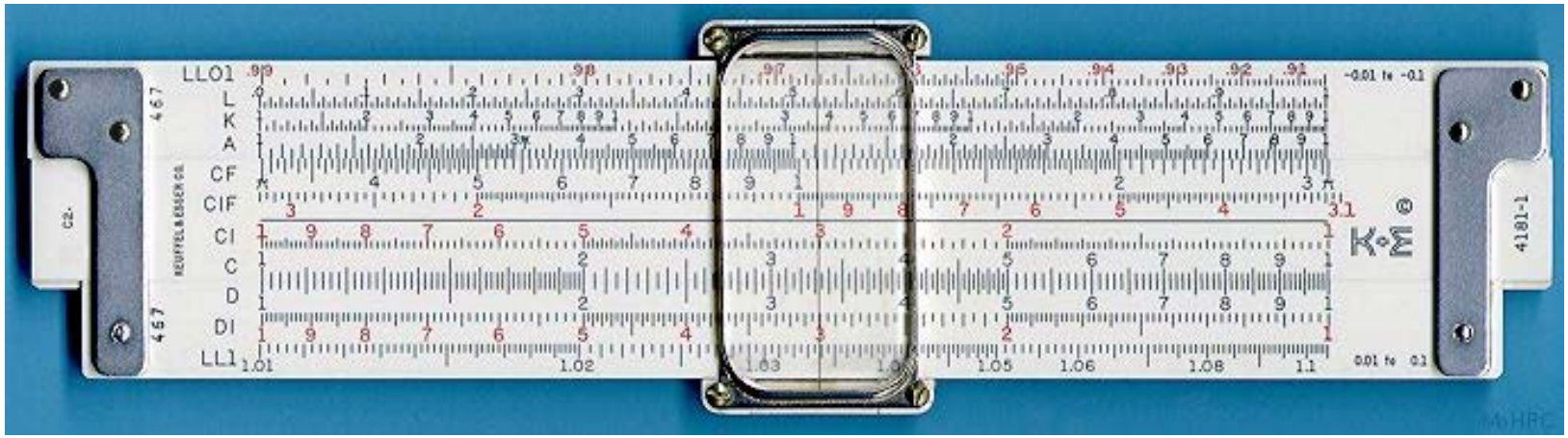
- The magic ingredient is the logarithm of each operand, which was originally obtained from a printed table.
- apier also invented an alternative to tables, where the logarithm values were carved on ivory sticks which are now called ***Napier's Bones***.





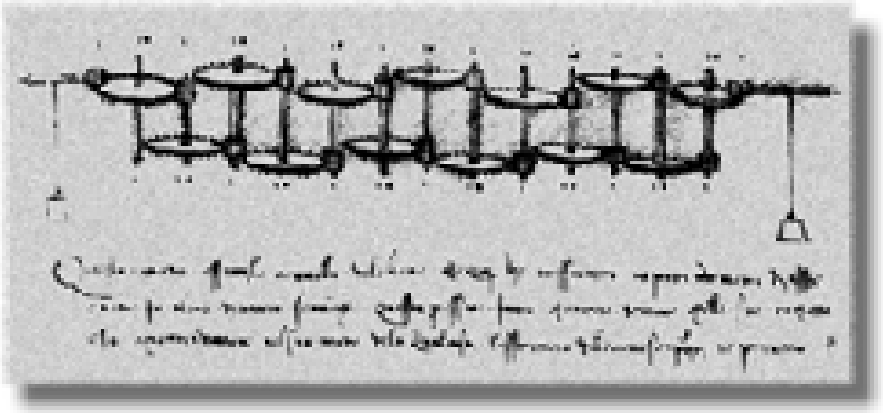
Slide Rule

› Napier's invention led directly to the ***slide rule***, first built in England in 1632 and still in use in the 1960's by the NASA engineers of the Mercury, Gemini, and Apollo programs which landed men on the moon.



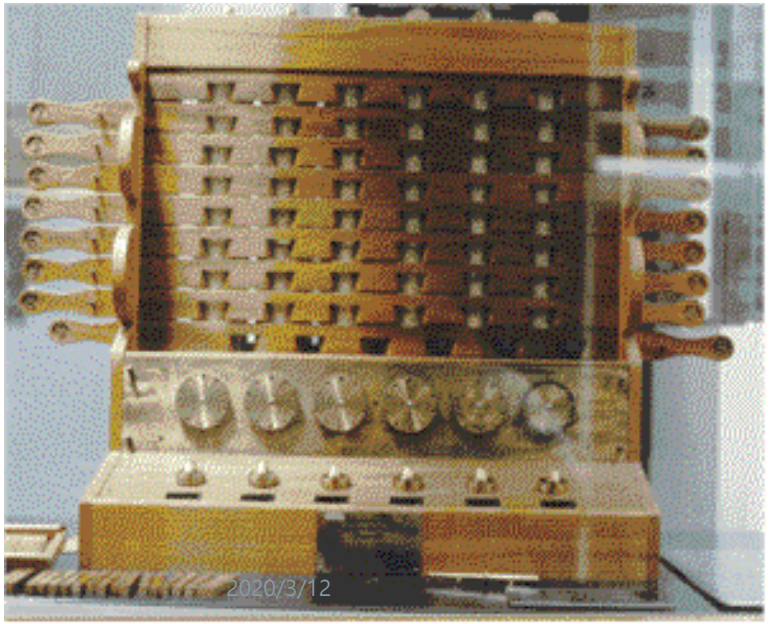
Leonardo da Vinci

› Leonardo da Vinci (1452-1519) made drawings of gear-driven calculating machines but apparently never built any.



Calculating Clock

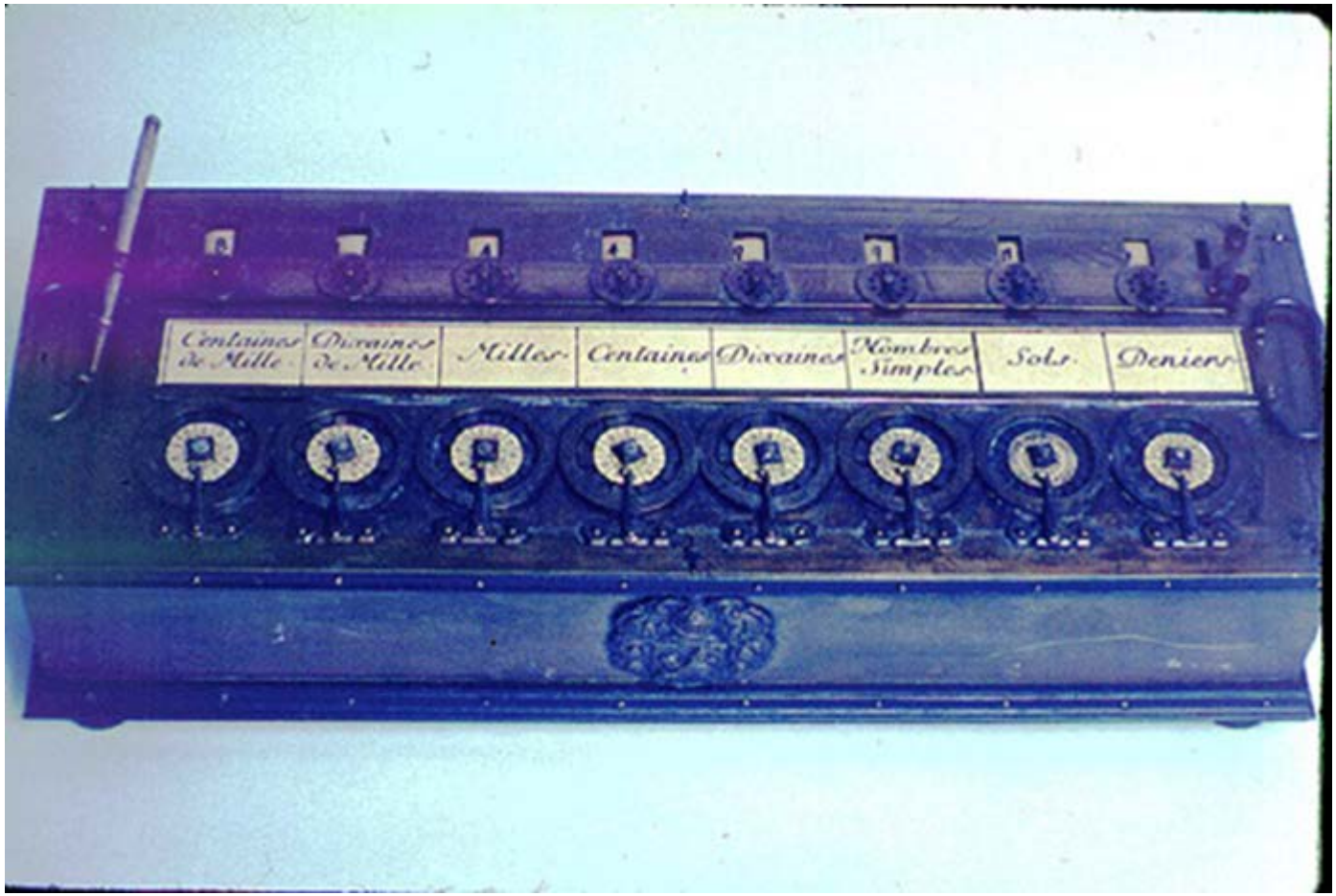
- The first gear-driven calculating machine to actually be built was probably the ***calculating clock***, so named by its inventor, the German professor Wilhelm Schickard in 1623.
- This device got little publicity because Schickard died soon afterward in the bubonic plague.



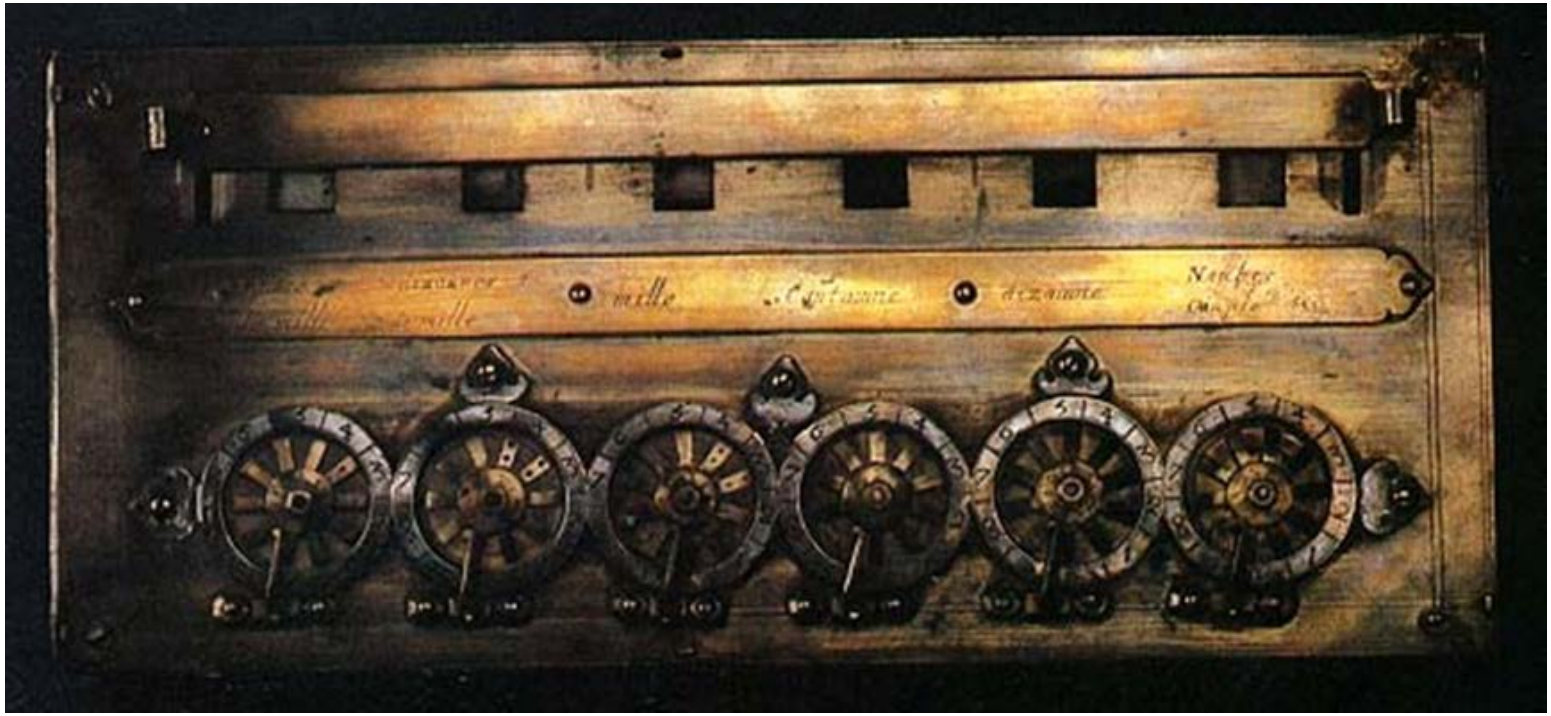
Blaise Pascal

- In 1642 Blaise Pascal, at age 19, invented the ***Pascaline*** as an aid for his father who was a tax collector.
  - Pascal built 50 of this gear-driven one-function calculator (it could only add) but couldn't sell many because of their exorbitant cost and because they really weren't that accurate (at that time it was not possible to fabricate gears with the required precision).
- Up until the present age when car dashboards went digital, the odometer portion of a car's speedometer used the very same mechanism as the Pascaline to increment the next wheel after each full revolution of the prior wheel.

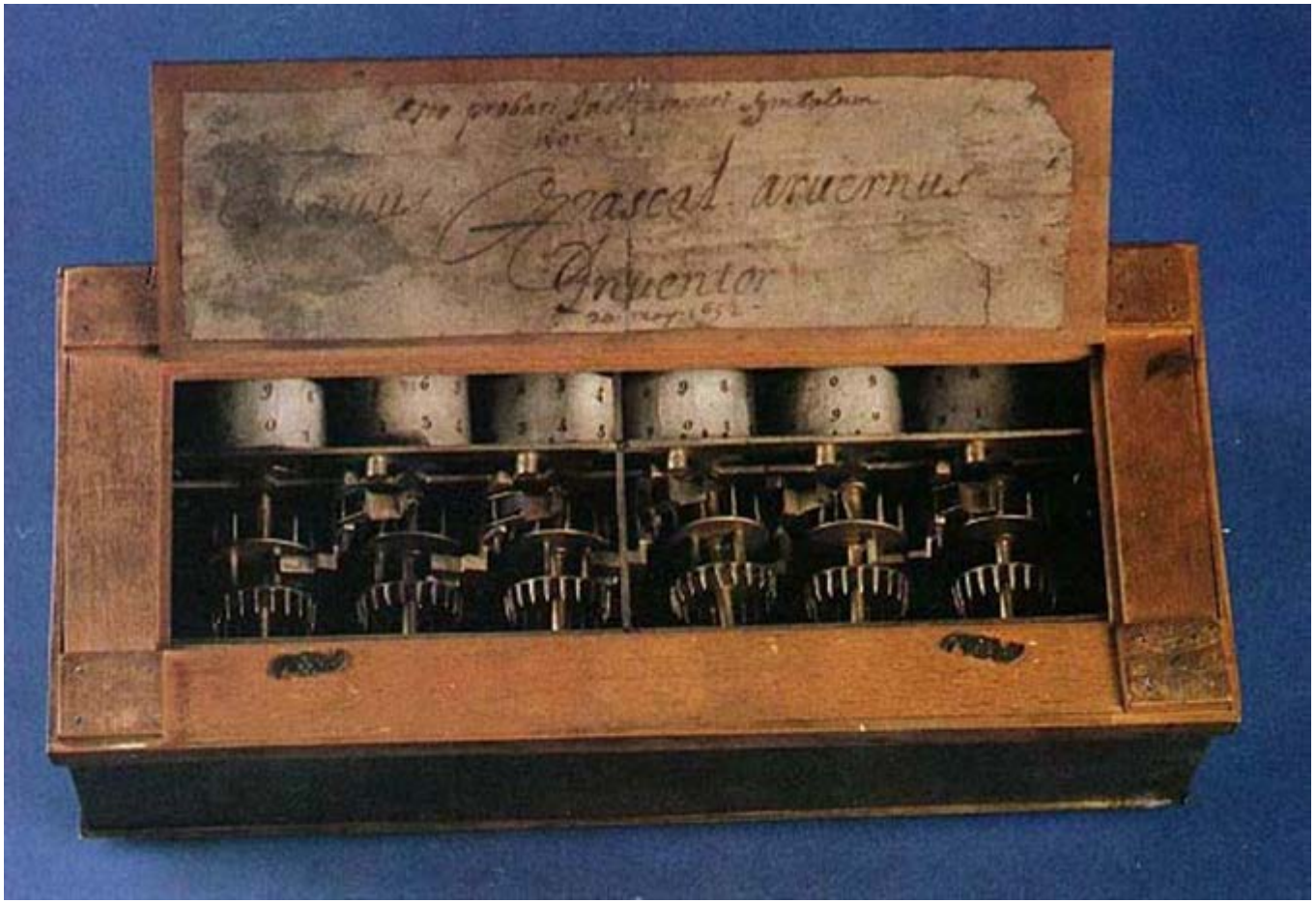
8-digit Pascaline



6-digit Pascaline (Cheaper)



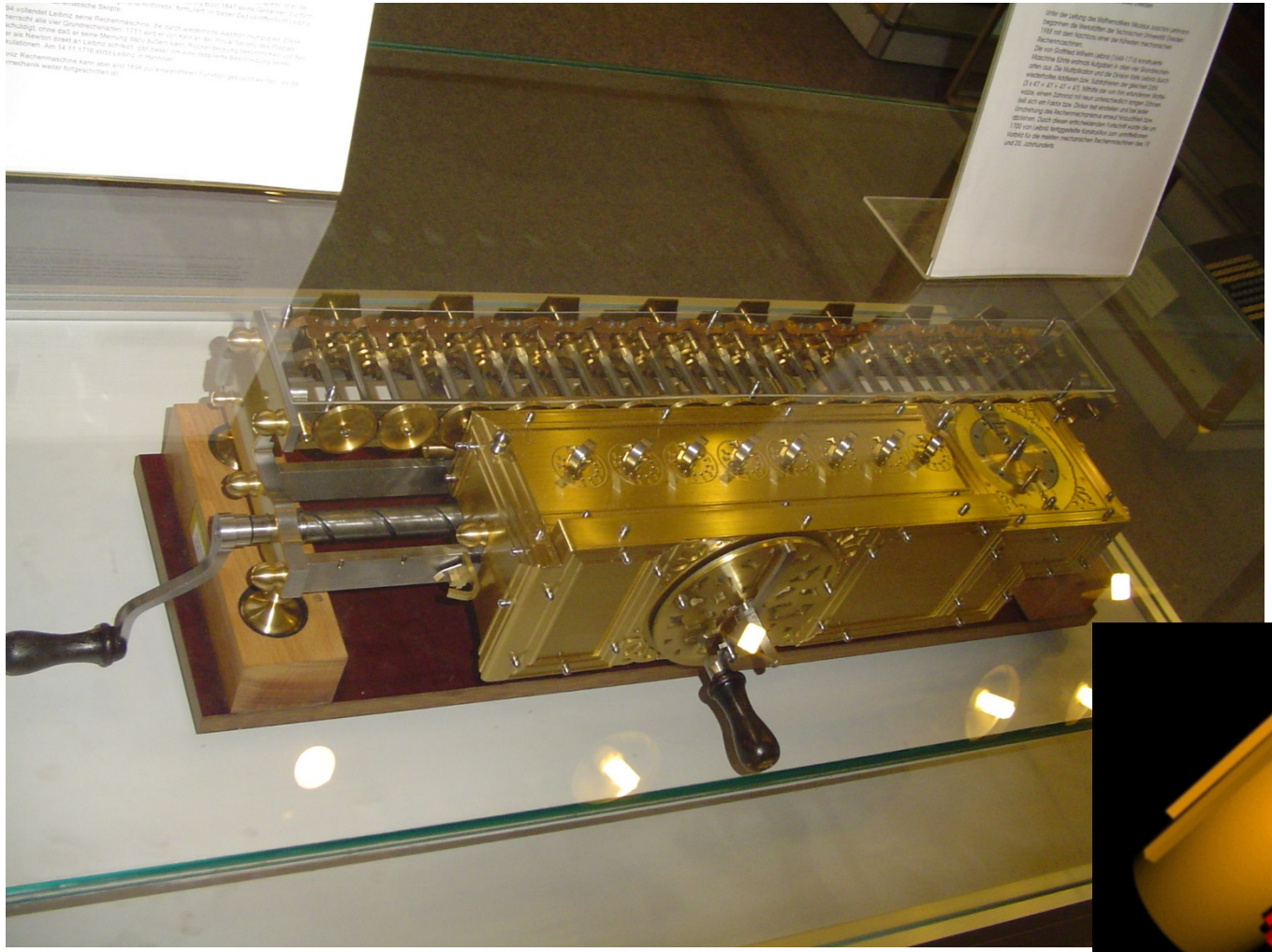
Pascaline Insides

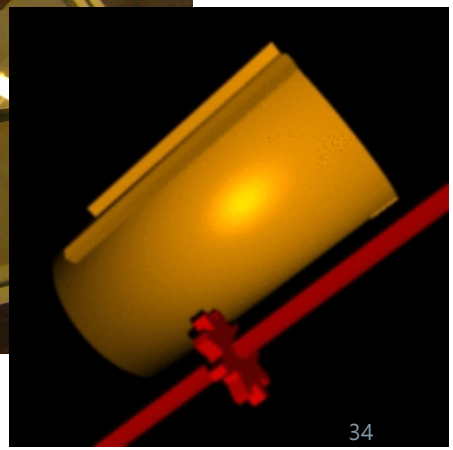


Leibniz

- Just a few years after Pascal, the German Gottfried Wilhelm Leibniz (co-inventor with Newton of calculus) managed to build a four-function (addition, subtraction, multiplication, and division) calculator that he called the ***stepped reckoner*** because, instead of gears, it employed fluted drums having ten flutes arranged around their circumference in a stair-step fashion.
- Although the stepped reckoner employed the decimal number system (each drum had 10 flutes), Leibniz was the first to advocate use of the **binary number system** which is fundamental to the operation of modern computers.
- Leibniz is considered one of the greatest of the philosophers but he died poor and alone.

Stepped Reckoner





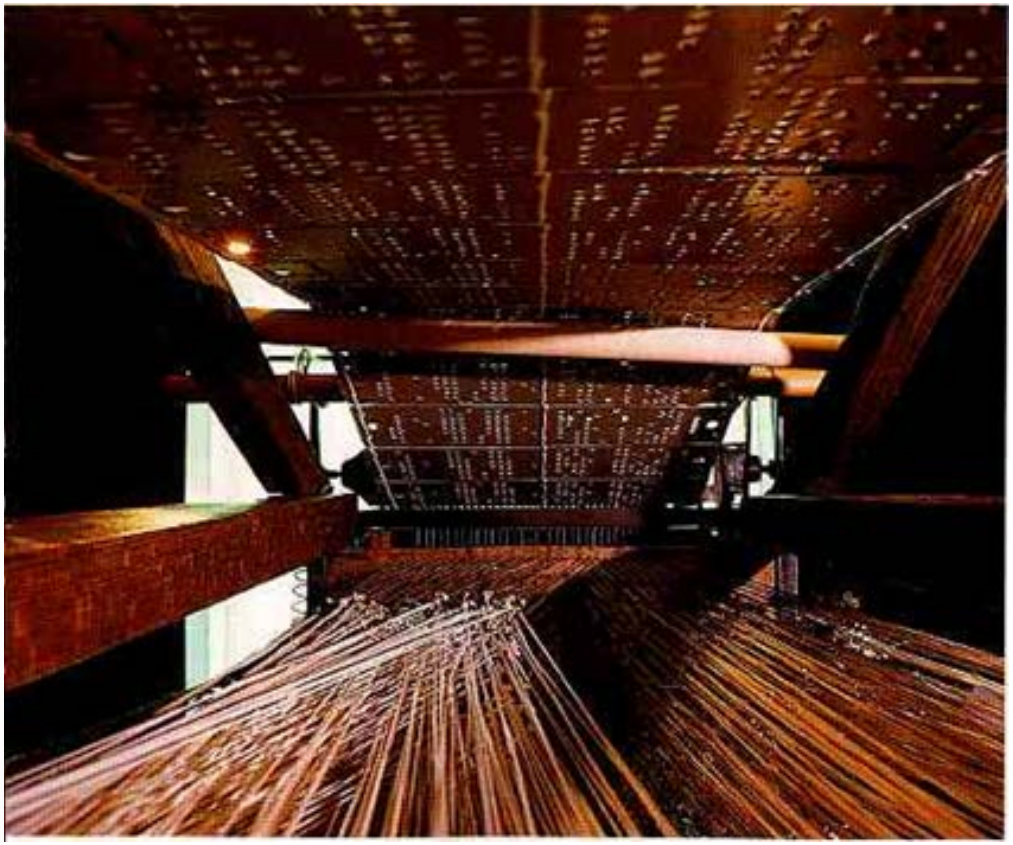
Jacquard

- In 1801 the Frenchman Joseph Marie Jacquard invented a power loom that could base its weave (and hence the design on the fabric) upon a pattern automatically read from punched wooden cards, held together in a long row by rope.
- Descendants of these ***punched cards*** have been in use ever since!

Power Loom

- By selecting particular cards for Jacquard's loom you defined the woven pattern.


[Jacquard-card Making.]



Close up of a Card

Technology -vs- Jobs

- Jacquard's technology was a real boon to mill owners, but put many loom operators out of work.
  - Angry mobs smashed Jacquard looms and once attacked Jacquard himself.
- History is full of examples of labor unrest following technological innovation yet most studies show that, overall, technology has actually increased the number of jobs.

Technology –vs- Jobs

Charle's Babbage

- By 1822 the English mathematician ***Charles Babbage*** was proposing a steam driven calculating machine the size of a room, which he called the ***Difference Engine***.

US Census

- The next breakthrough occurred in America. The U.S. Constitution states that a census should be taken of all U.S. citizens every 10 years in order to determine the representation of the states in Congress.
- While the very first census of 1790 had only required 9 months, by 1880 the U.S. population had grown so much that the count for the 1880 census took 7.5 years.
  - Automation was clearly needed for the next census.
- The census bureau offered a prize for an inventor to help with the 1890 census and this prize was won by Herman Hollerith.

Hollerith Desk

- The ***Hollerith desk***, consisted of:
  - A card reader which sensed the holes in the cards,
  - A gear driven mechanism which could count (similar to Pascal's)
  - A large wall of dial indicators to display the results of the count.

Hollerith Desk

- Hollerith's technique was successful and the 1890 census was completed in only 3 years at a savings of 5 million dollars.

IBM

- Hollerith built a company, the Tabulating Machine Company which, after a few buyouts, eventually became International Business Machines, known today as ***IBM***.

Hollerith's innovation

- By using punch cards, Hollerith created a way to store and retrieve information.
- This was the first type of read and write technology.

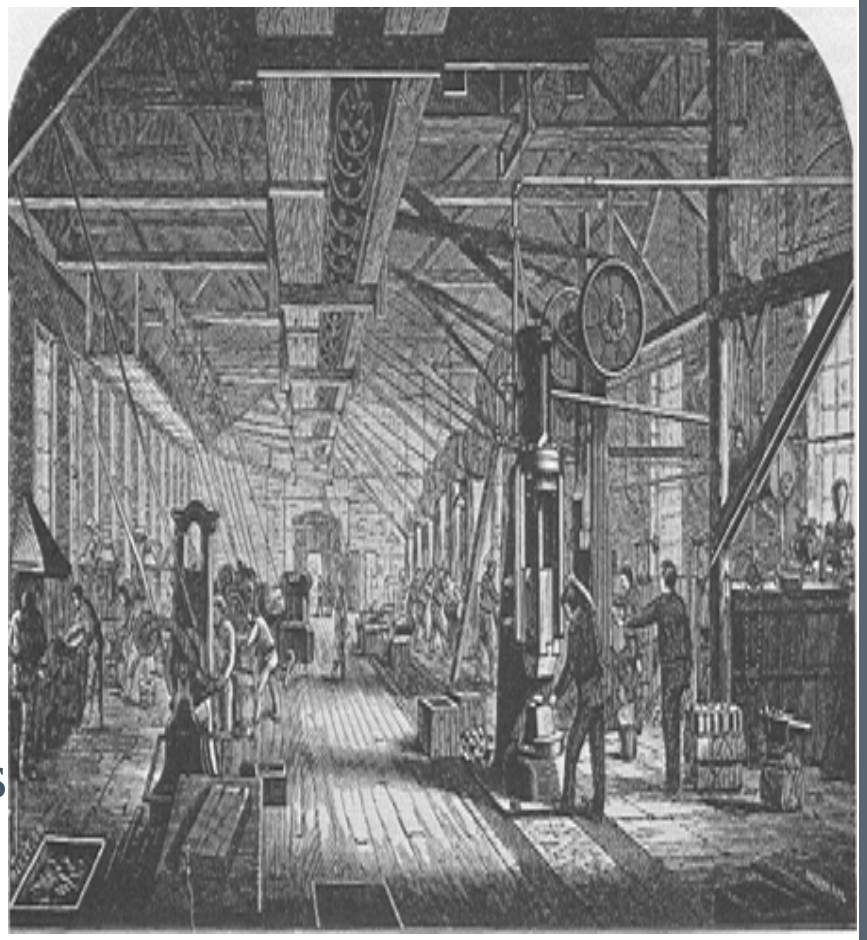
Abb. 7 Titelseite.
Scientific American 63(1890) No. 9.
(August 30, 1890)

US Military

- The U.S. military desired a mechanical calculator more optimized for scientific computation.
- By World War II the U.S. had battleships that could lob shells weighing as much as a small car over distances up to 25 miles.
- Physicists could write the equations that described how atmospheric drag, wind, gravity, muzzle velocity, etc. would determine the trajectory of the shell, but solving such equations was extremely laborious.

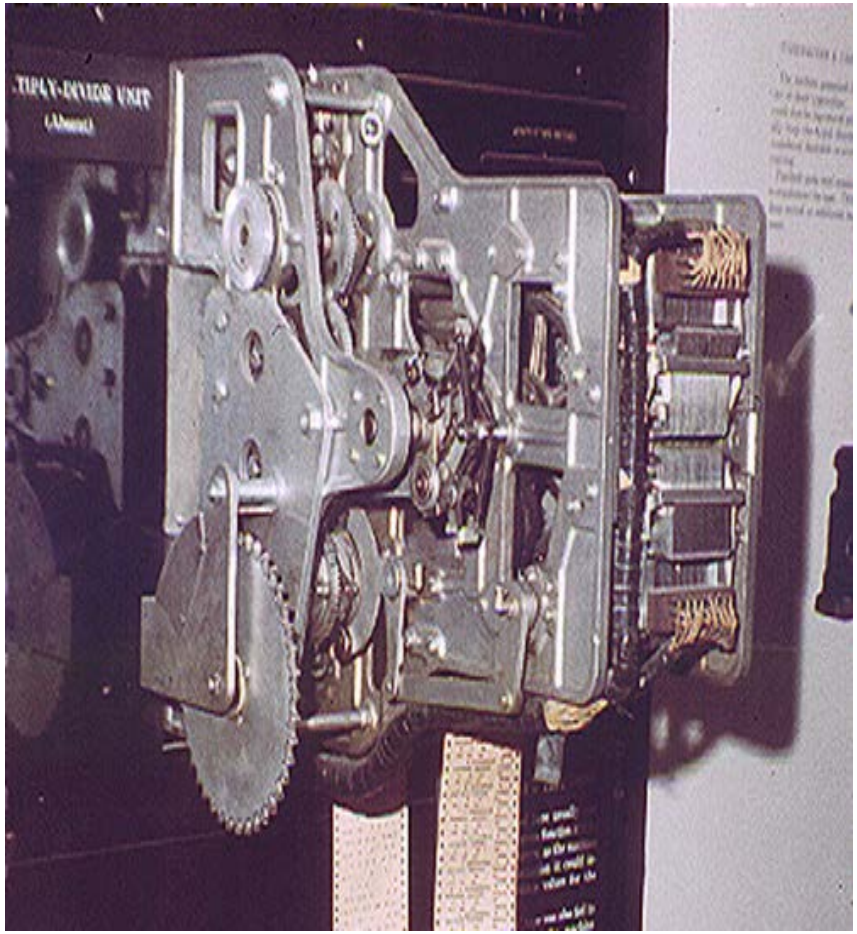
Mark I

- One early success was the Harvard ***Mark I*** computer which was built as a partnership between Harvard and IBM in 1944.
- This was the first programmable digital computer made in the U.S.
- But it was not a purely electronic computer. Instead the Mark I was constructed out of switches relays, rotating shafts, and clutches.



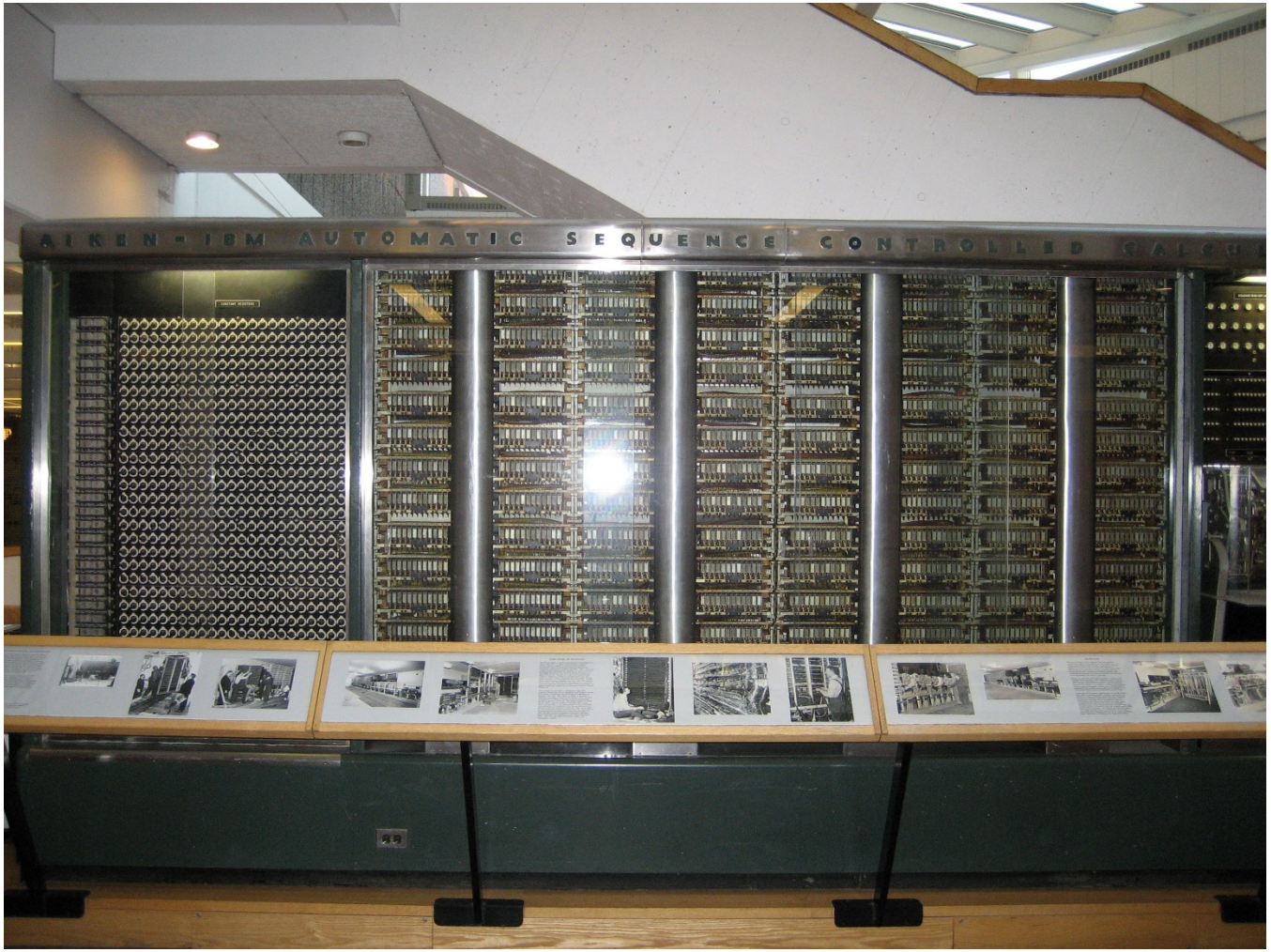
Mark I

- The machine weighed 5 tons, incorporated 500 miles of wire, was 8 feet tall and 51 feet long, and had a 50 ft rotating shaft running its length, turned by a 5 horsepower electric motor.
- The Mark I ran non-stop for 15 years, sounding like a roomful of ladies knitting.



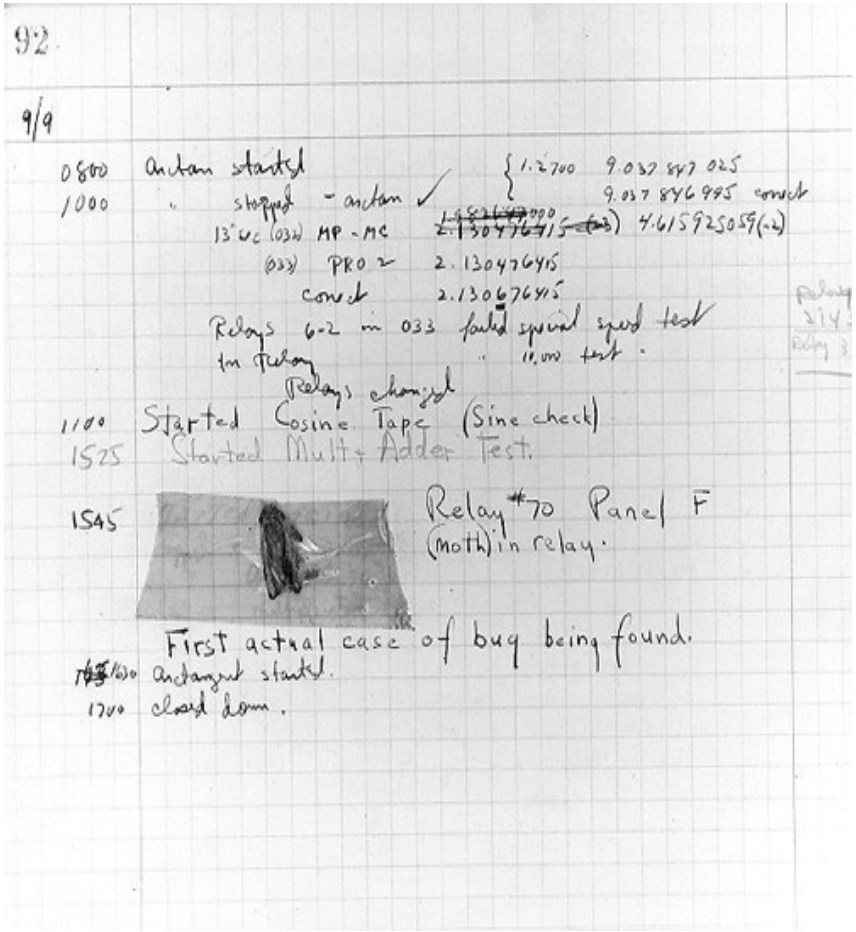
The Mark I computer

Mark I



The First Bug

- One of the primary programmers for the Mark I was a woman, ***Grace Hopper***.
- Hopper found the first computer “bug”: a dead moth that had gotten into the Mark I.
- The word “bug” had been used to describe a defect since at least 1889 but Hopper is credited with coining the word “debugging” to describe the work to eliminate program faults.



Humor

- On a humorous note, the principal designer of the Mark I, ***Howard Aiken*** of Harvard, estimated in 1947 that six electronic digital computers would be sufficient to satisfy the computing needs of the entire United States.

The Future of Computers?

- IBM had commissioned this study to determine whether it should bother developing this new invention into one of its standard products (up until then computers were one-of-a-kind items built by special arrangement).
- Aiken's prediction wasn't actually so bad as there were very few institutions (principally, the government and military) that could afford the cost of what was called a computer in 1947.
- He just didn't foresee the micro-electronics revolution which would allow something like an ***IBM Stretch*** computer of 1959.

First Generation Computers

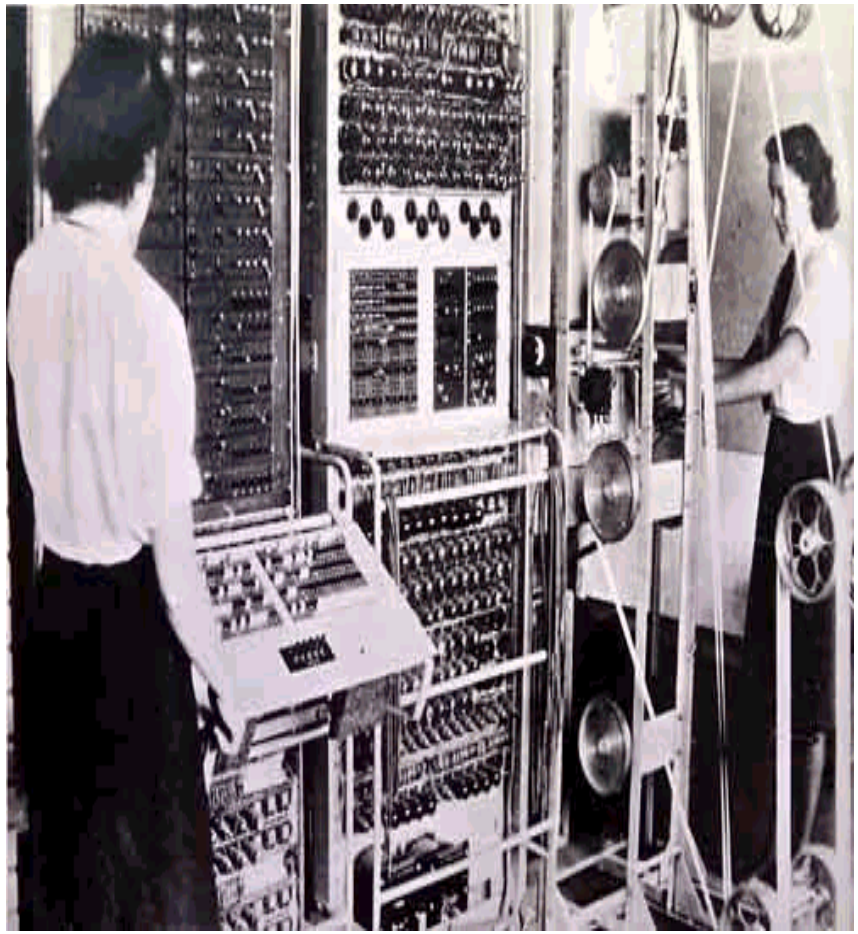
- The first electronic computer was designed at Iowa State between 1939-1942.
- The Atanasoff-Berry Computer used the binary system(1’s and 0’s).
- Contained vacuum tubes and stored numbers for calculations by burning holes in paper.

IBM Stretch - 1959

IBM Stretch - 1959

Colossus

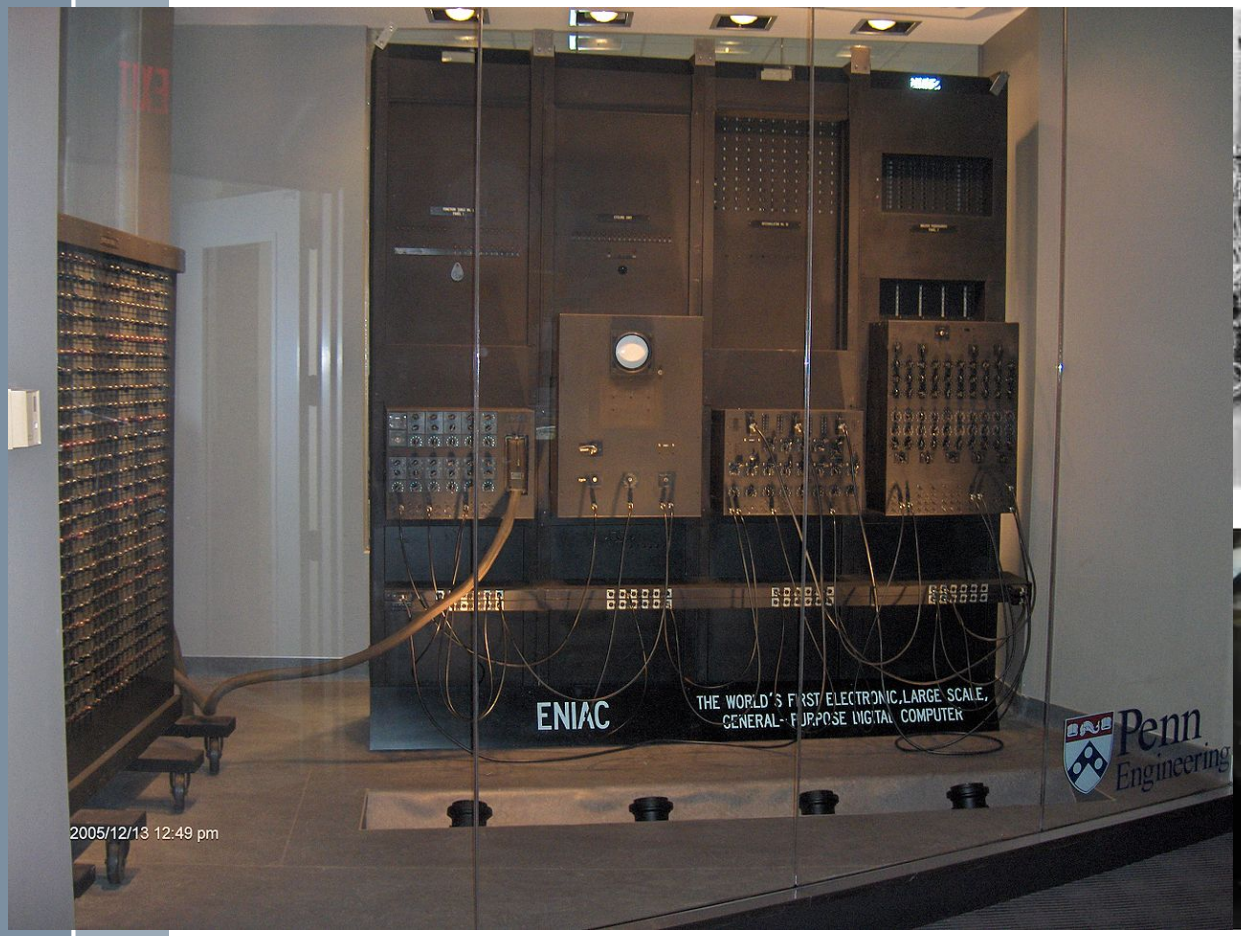
- ***The Colossus***, built during World War II by Britain for the purpose of breaking the cryptographic codes used by Germany.
- Britain led the world in designing and building electronic machines dedicated to code breaking, and was routinely able to read coded Germany radio transmissions.
- Not a general purpose, reprogrammable machine.



History of Microprocessors

- Most important advances in computer technology: 16-bit and 32-bit microprocessors (μp).
- Pioneered by Intel since 1970's and dominated by INTEL since 1980's
  - 4-bit 4004 in 1971.
  - 8-bit 8008 in 1972.
  - 8-bit 8080 and 8085 in 1974.
  - 16-bit 8088 and 8086, brains of famous IBM PC.
  - 32-bit 80286 (1982), 80386 (1985), 80486 (1989), Pentium.
  - (1993), Pentium II (1997), Celeron and Pentium III (1999)and Pentium 4 (2000).
  - 64-bit Itanium (2001).
  - 64-bit Pentium 4 and Xeon (2005).
  - Core iX series (2008-2010).

The ENIAC – Electronic Numerical Integrator and Computer

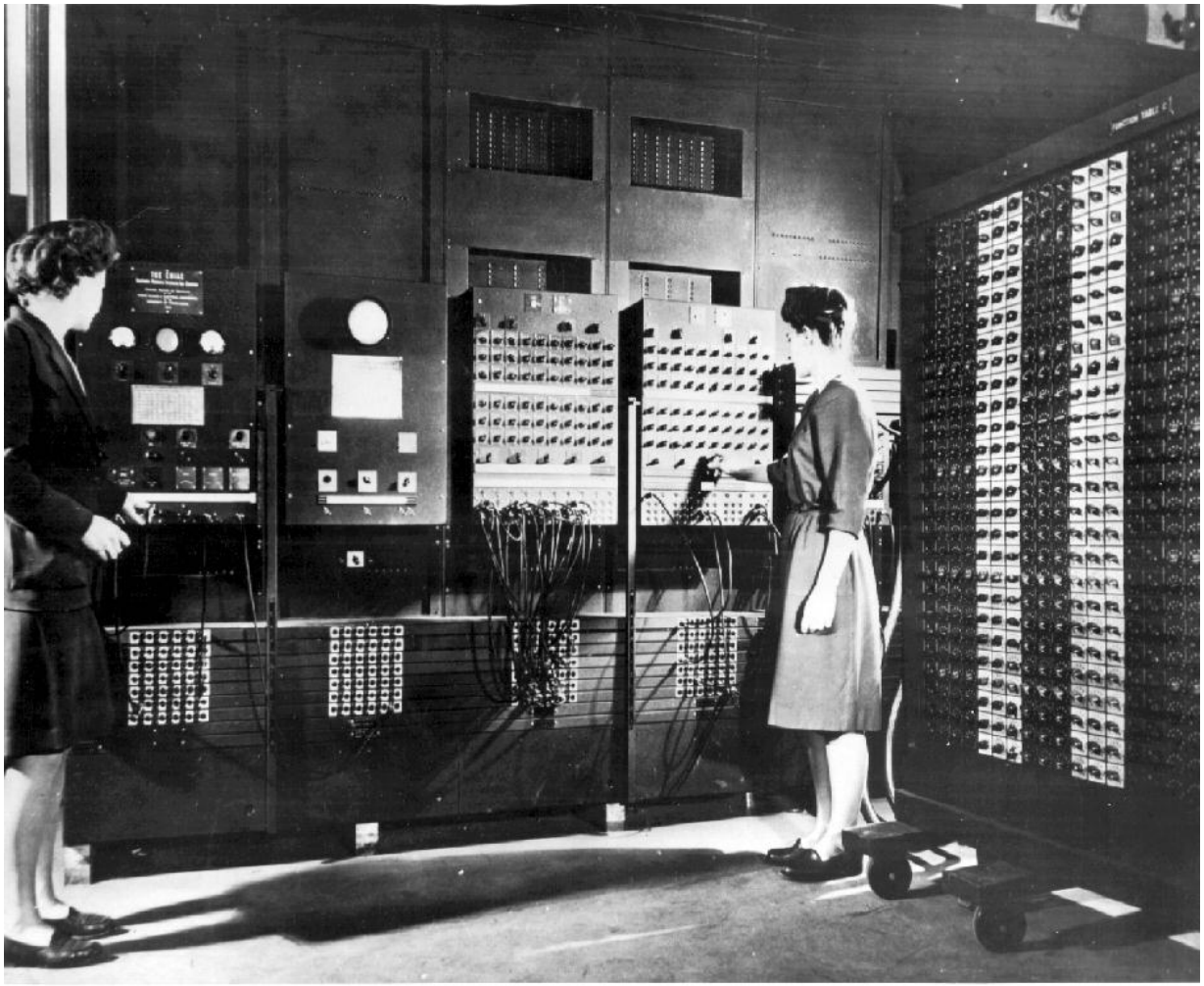


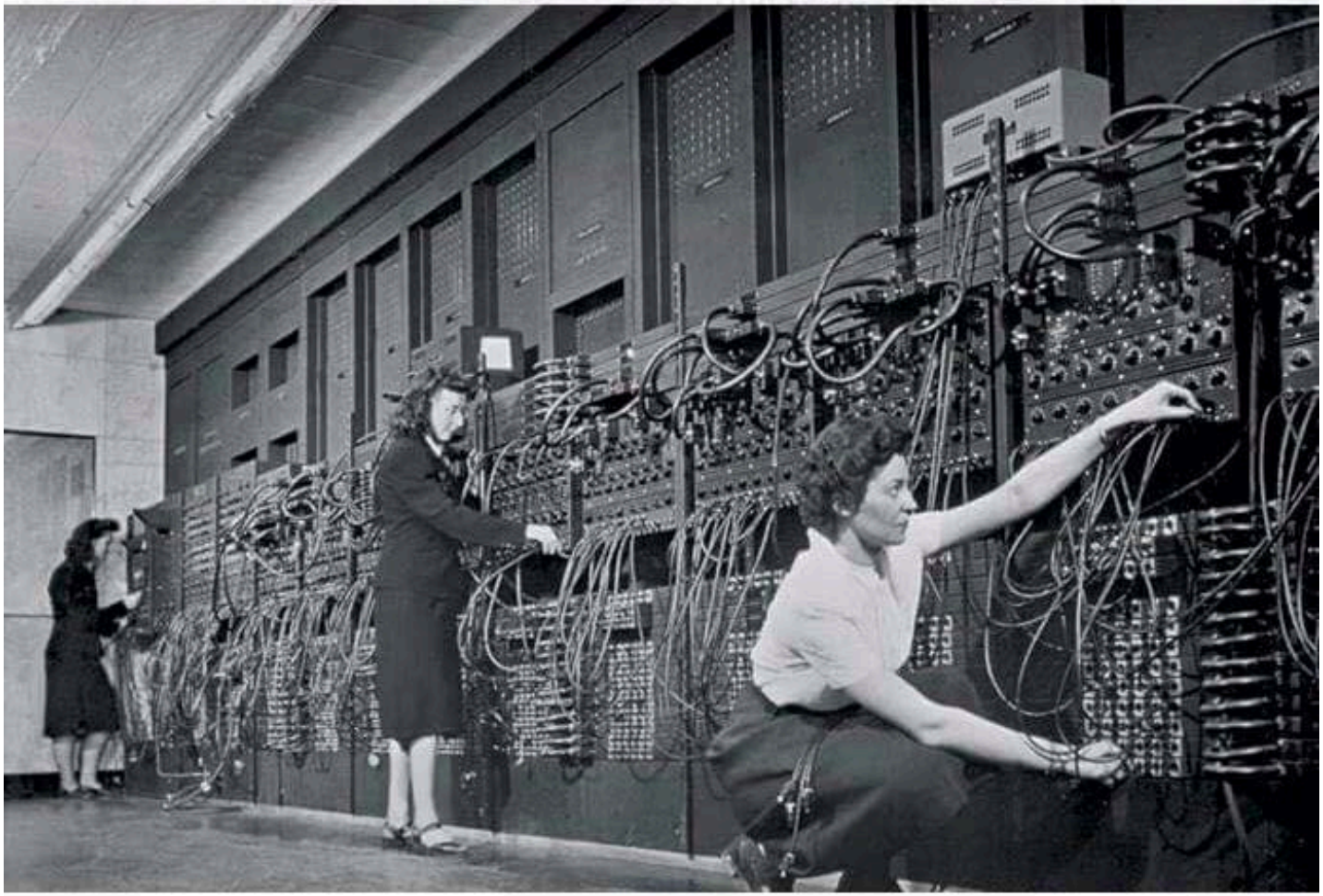
Vacuum tubes

ENIAC

- The title of forefather of today's all-electronic digital computers is usually awarded to ***ENIAC***, which stood for Electronic Numerical Integrator and Calculator.
- ENIAC was built at the University of Pennsylvania between 1943 and 1945 by two professors, ***John Mauchly*** and the 24 year old ***J. Presper Eckert***, who got funding from the war department after promising they could build a machine that would replace all the "computers”
- ENIAC filled a 20 by 40 foot room, weighed 30 tons, and used more than 18,000 vacuum tubes.

“Writing” Programs on the ENAIC

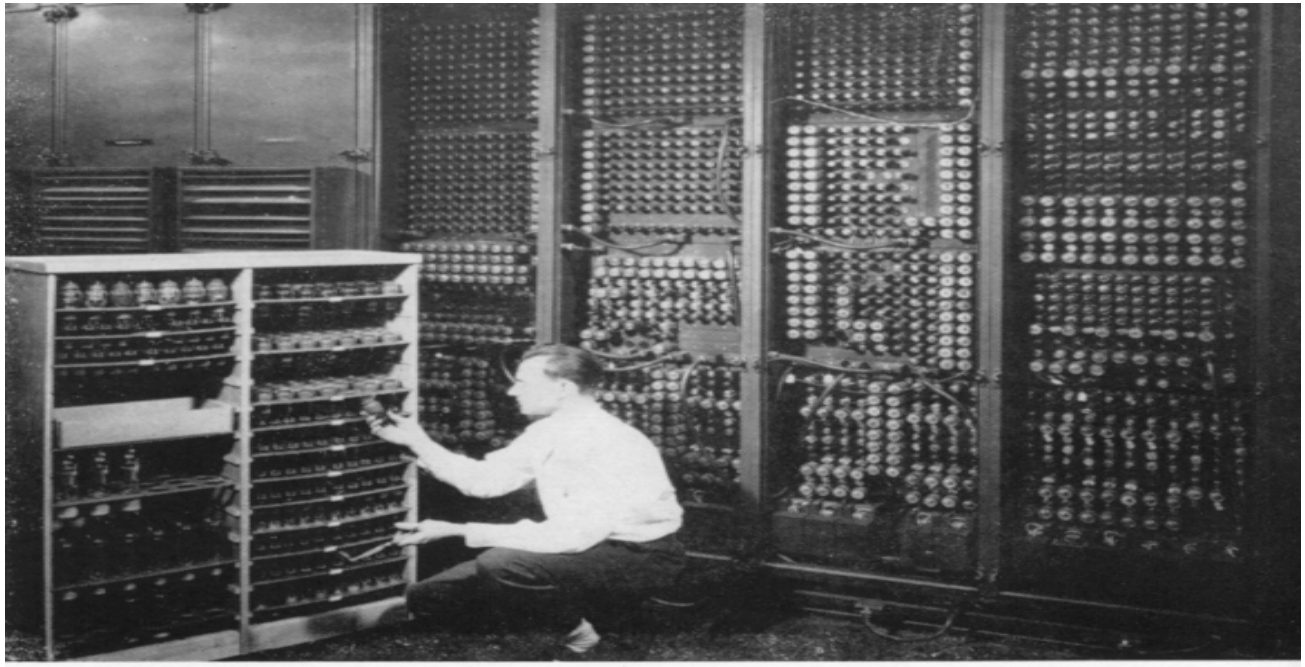




The first computer programmers were women

Problems with the ENIAC

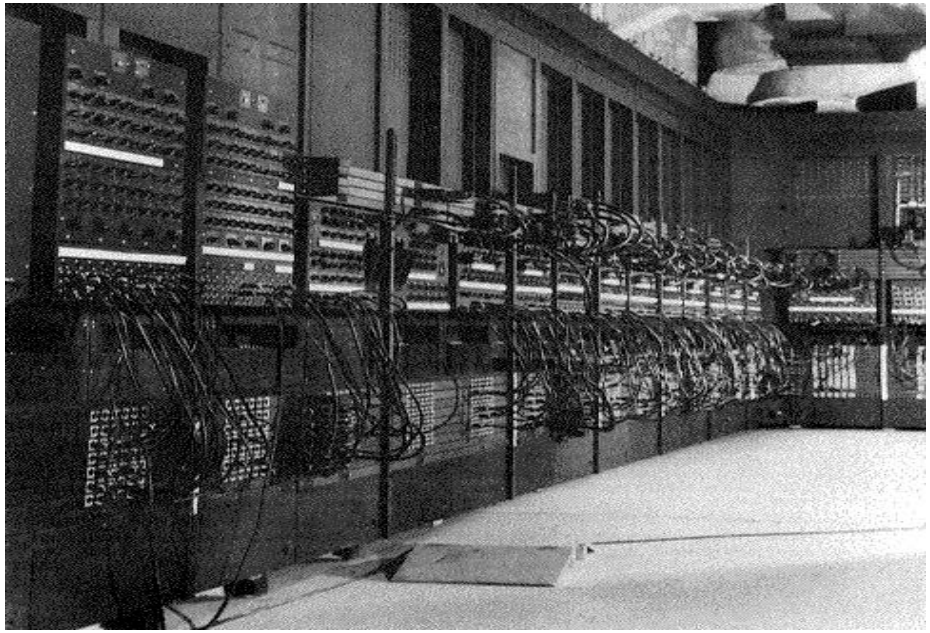
- The ENIAC used 18,000 vacuum tubes to hold a charge.
- Vacuum tubes were so notoriously unreliable that even twenty years later many neighborhood drug stores provided a “tube tester”.



Replacing a bad tube meant checking among ENIAC's 19,000 possibilities.

Von Neumann Architecture

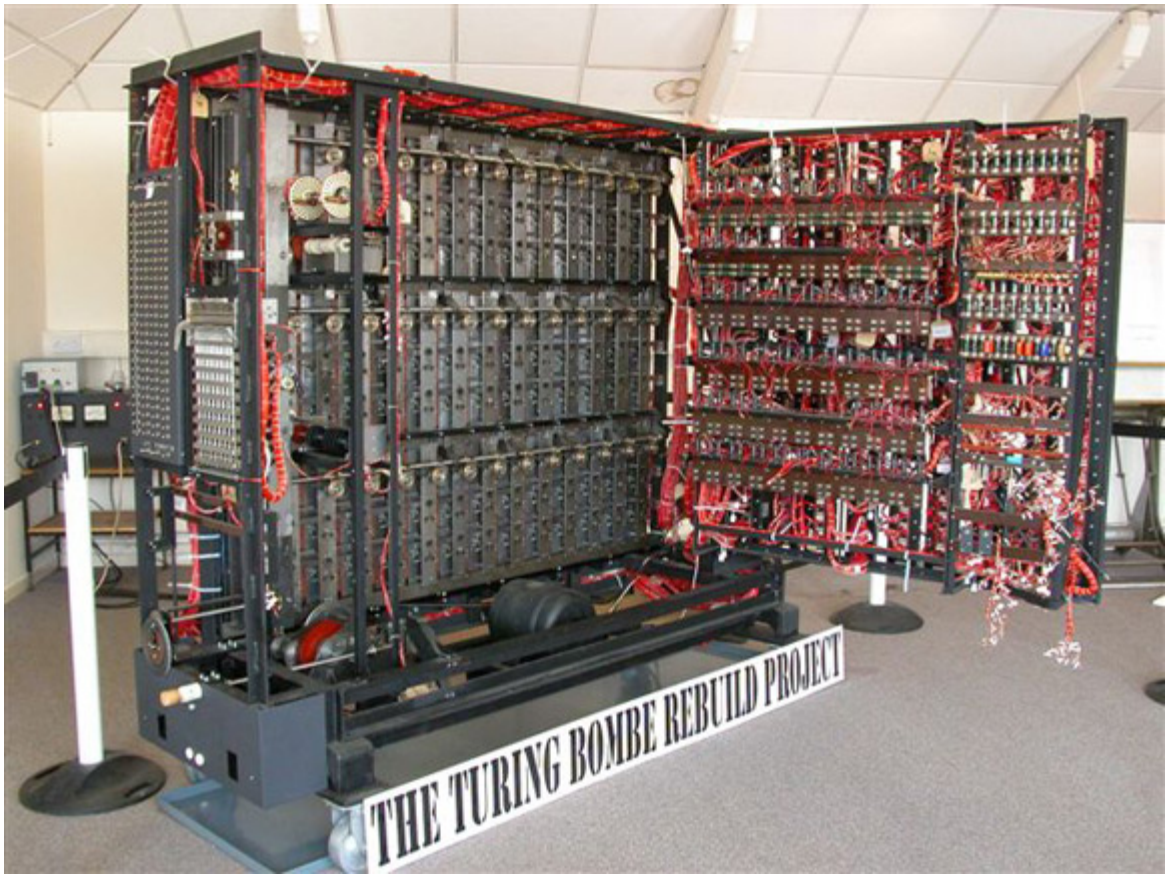
- John Von Neumann came up with the bright idea of using part of the computer's internal memory (called Primary Memory) to "store" the program inside the computer and have the computer go get the instructions from its own memory, just as we do with our human brain.



The Stored Program Computer

- In 1945 John von Neumann presented his idea of a computer that would store computer instructions in a CPU.
- The CPU(Central Processing Unit) consisted of elements that would control the computer electronically.
- The EDVAC, EDSAC and UNIVAC were the first computers to use the stored program concept.
- They used vacuum tubes so they were too expensive and too large for households to own and afford.

Turing Bombe



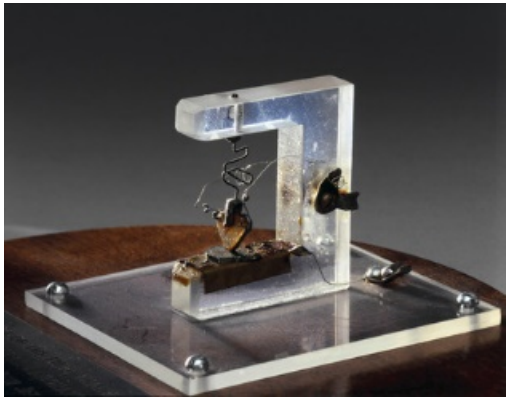
Prices in 1968

› Does not include maintenance and tech support!

Model	Description	Purchase Price	Installation Fee
3011-95	1108 CPU	$566,460	$2,200
7005-72	131 K word Core Memory	$823,500	$2,250
5009-00	FASTRAND^{tm} Controller	$41,680	$600
6010-00	FASTRAND II Storage Unit	$134,400	$1,080
5012-00	FH-432/FH-1782 Drum Controller	$67,360	$600
6016-00	FH-432 Drum (capacity 262,144 words)	$34,640	$480
6015-00	FH-1782 Drum (capacity 2,097,152 words)	$95,680	$540
4009-99	Console (TTY-35)	$29,365	$200

Second Generation Computers

- In 1947, the transistor was invented.
- The transistor made computers smaller, less expensive and increased calculating speeds.
- Second generation computers also saw a new way data was stored.
- Punch cards were replaced with magnetic tapes and reel to reel machines.



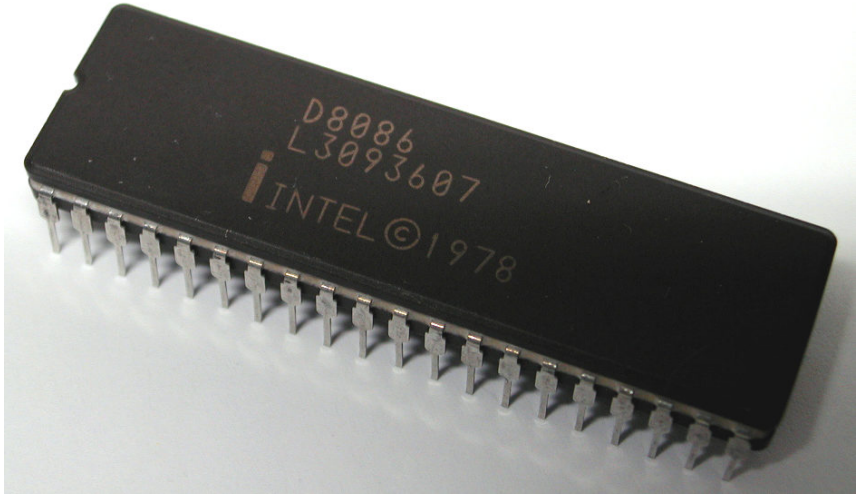
John Bardeen, William Shockley and Walter Brattain

UNIVAC

- The ***UNIVAC*** computer was the first commercial (mass produced) computer.
- In the 50's, UNIVAC (a contraction of "Universal Automatic Computer") was the household word for “computer” just as “Kleenex” is for “tissue”.
- UNIVAC was also the first computer to employ magnetic tape.

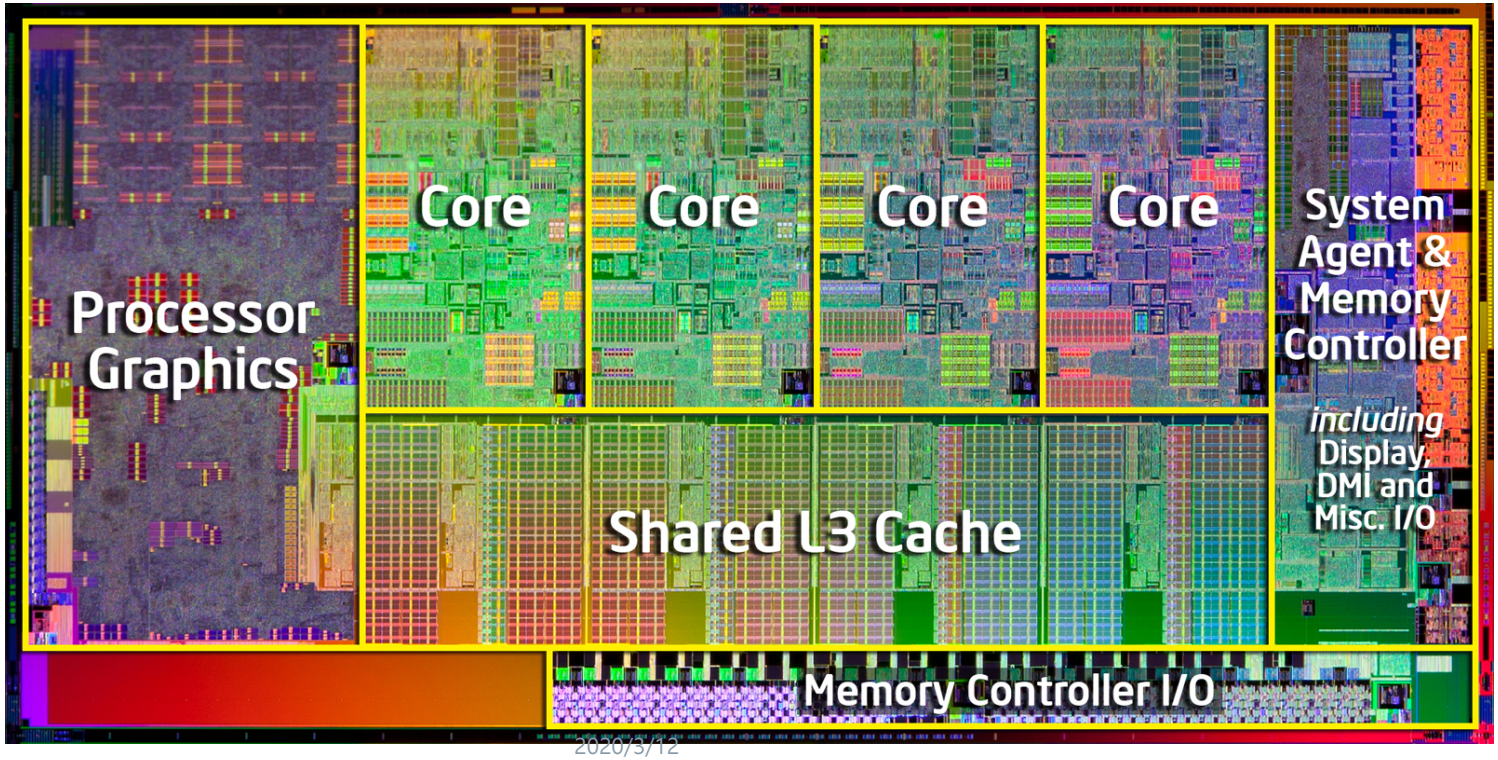
Third Generation Computers

- Transistors were replaced by integrated circuits (IC).
- One IC could replace hundreds of transistors.
- This made computers even smaller and faster.



Fourth Generation Computers

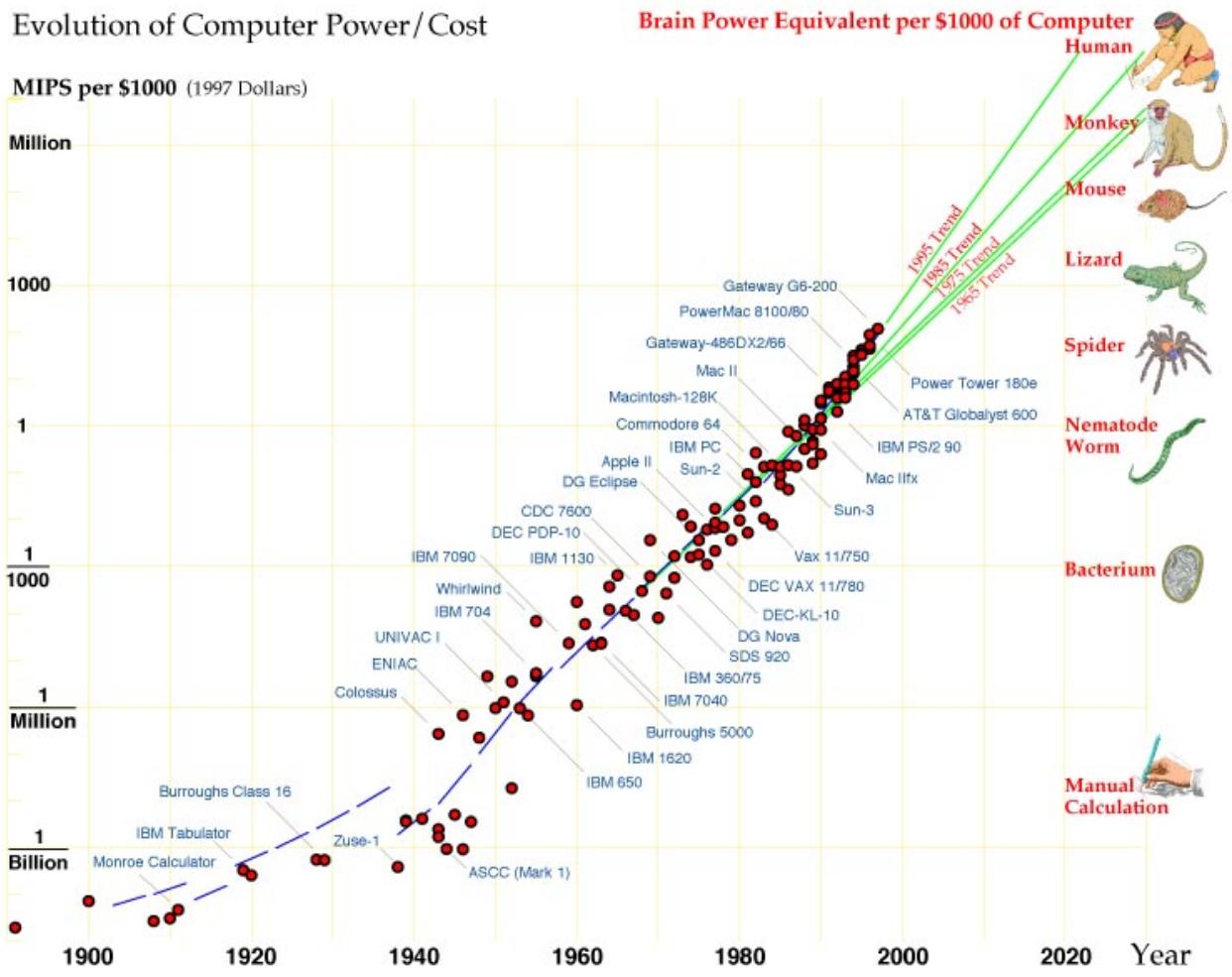
- In 1970 the Intel Corporation invented the Microprocessor: an entire CPU on one chip.
- This led to microcomputers-computers on a desk.



Personal Computers

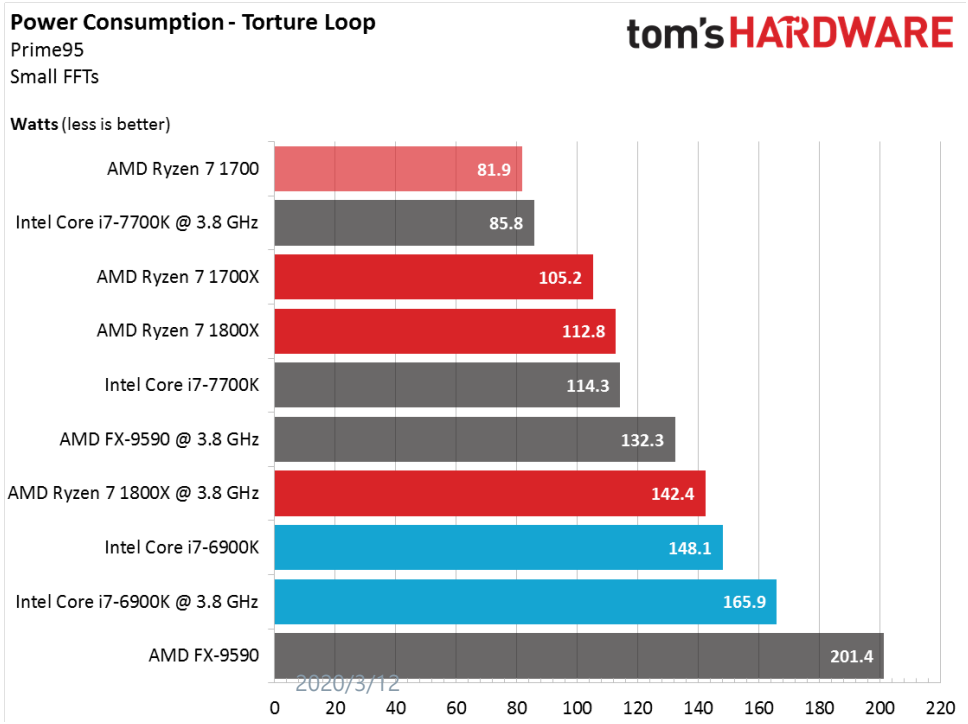
- First used by hobbyists
- IBM introduced the PC in 1981.
  - Accepted by business
  - Became the standard hardware design for most desktop computers
  - Most PCs use software from Microsoft

Evolution of Computer Power vs Cost

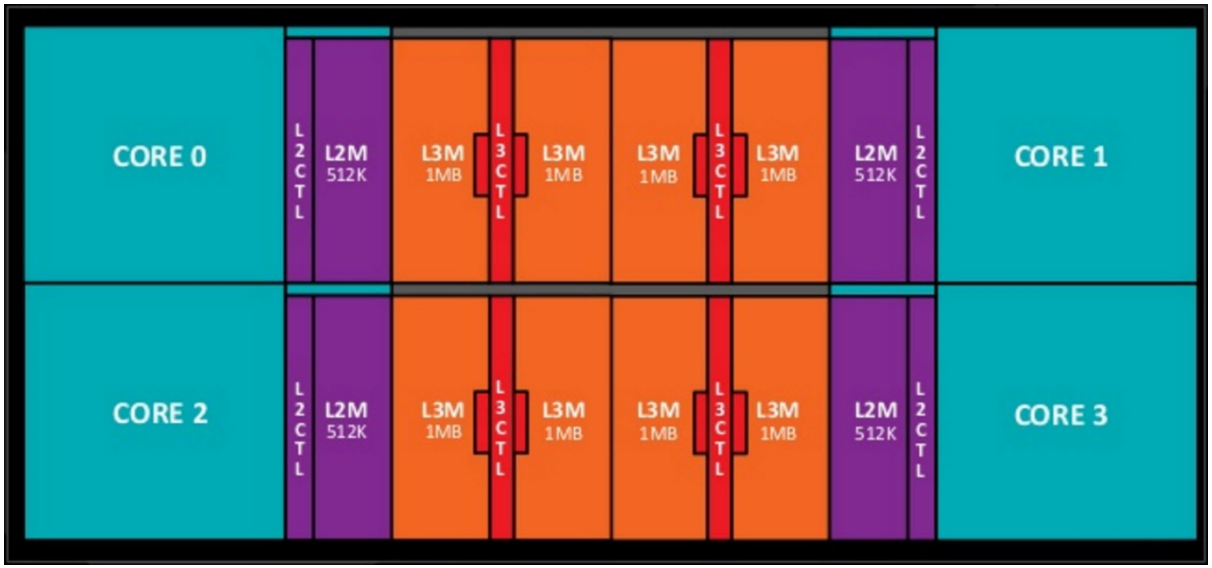


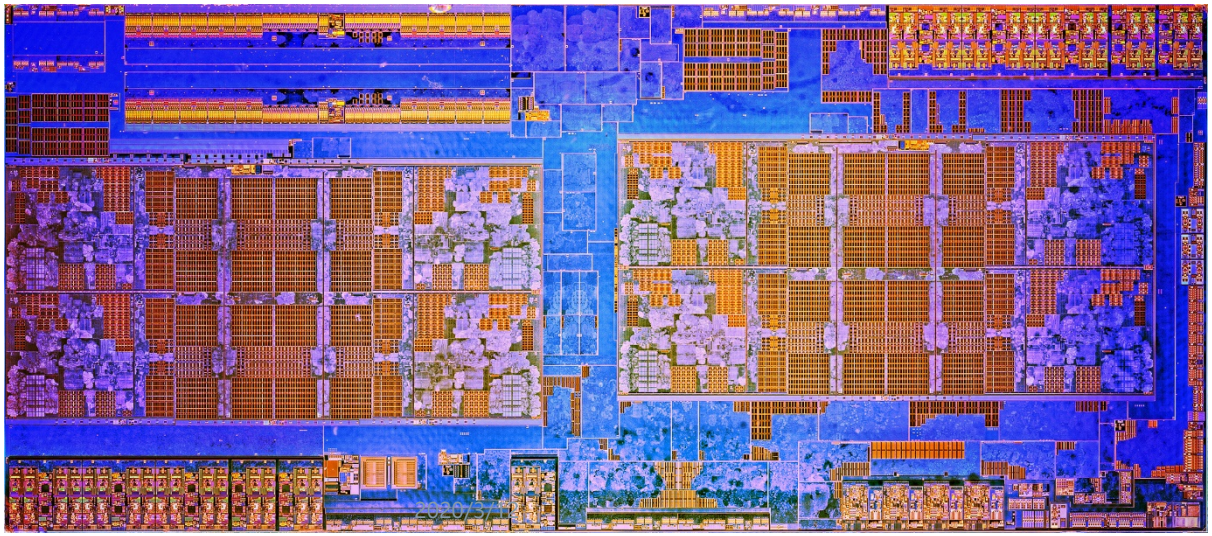
Fifth Generation?

- Multiple interconnected die (AMD)
  - Away from monolithic chip designs.
- Multithreading, thread ripping
- The problem is power consumption



Fifth Generation?





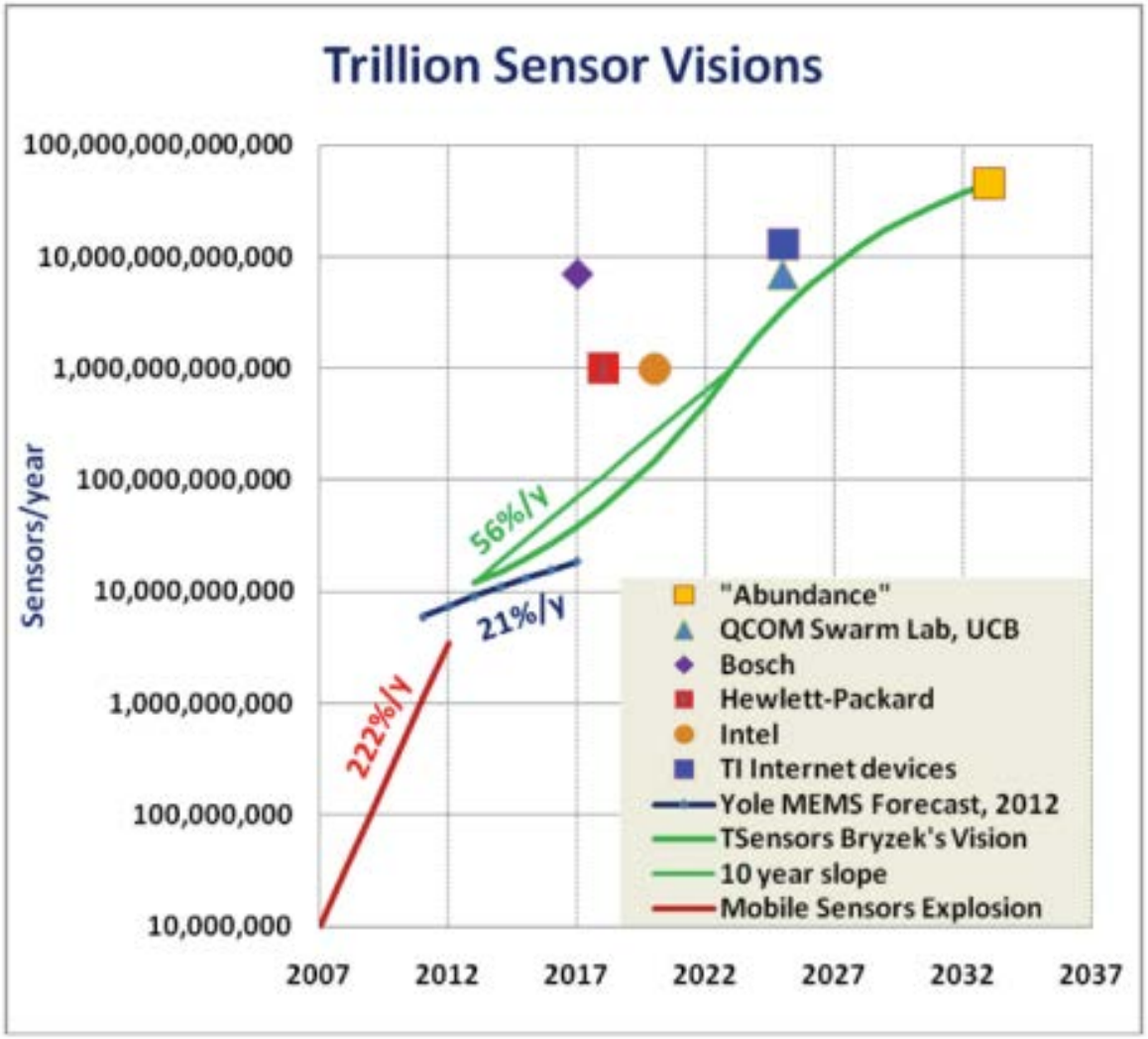
Infinity band (channel)

Super Clusters (PS3)

Into the Millennia

- Internet revolutionized communications
  - World Wide Web
  - Search Engines (Google, Yahoo, and Microsoft)
- Miniaturization of computing machines
  - Embedded (GPS, in automobile engines)
  - Smartphone

Internet of Things



Wearable devices