"Digits" does not refer only to your 10 fingers…

Digitization & Processing

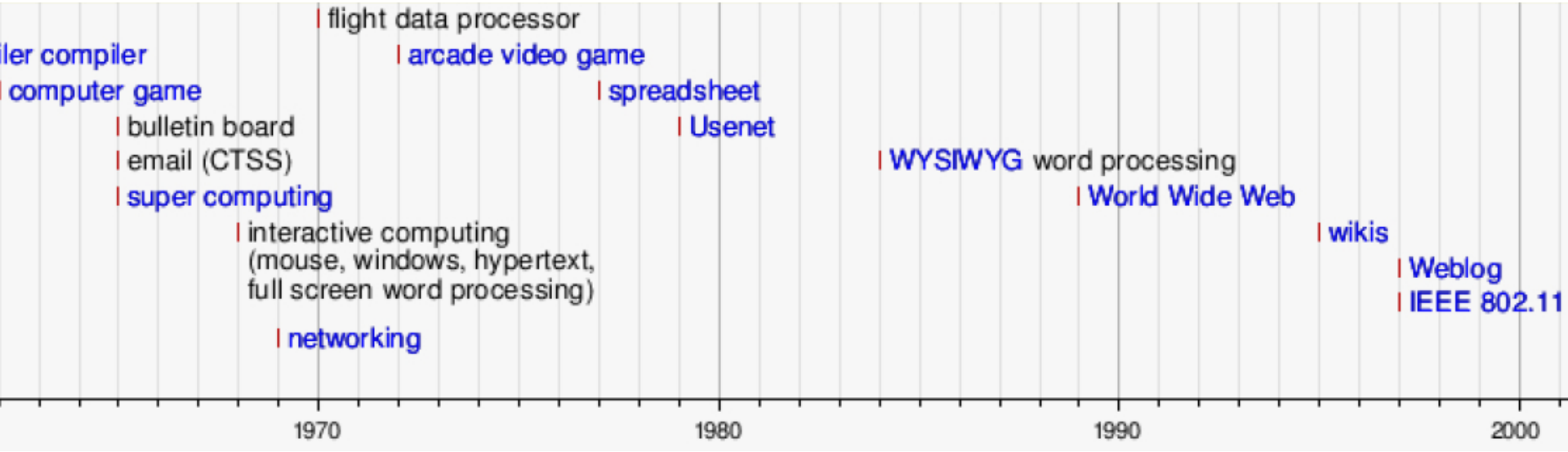
Used with permission from
Lawrence Snyder
University of Washington, Seattle

A Short History of Digital Info

- One goal of CS Principles is to understand how computers and digital information are "game changers," how they *create* opportunities
- I will do that by highlighting the progress of "data processing" over the last 120 years or so (it's very incomplete)
  - Digitization, computers, ICs, transistors, PCs, Internet, and WWW are key
  - Focus on advances since …

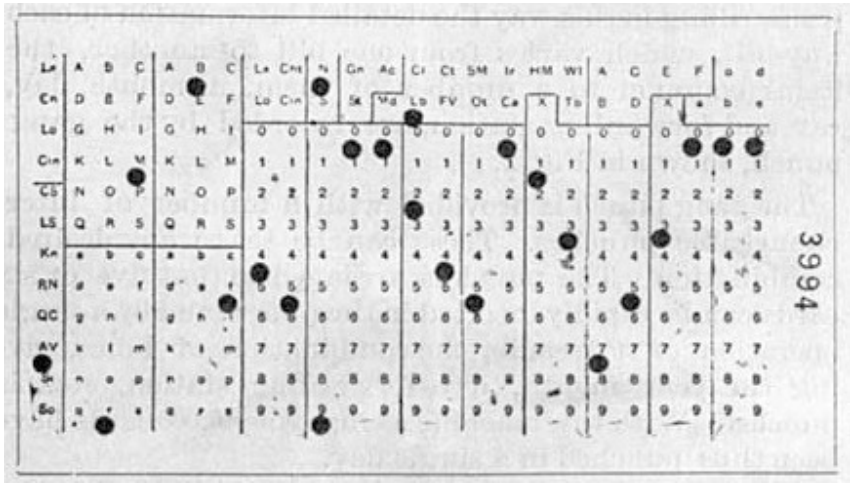
Wikipedia Timeline

en.wikipedia.org/wiki/Timeline_of_computing

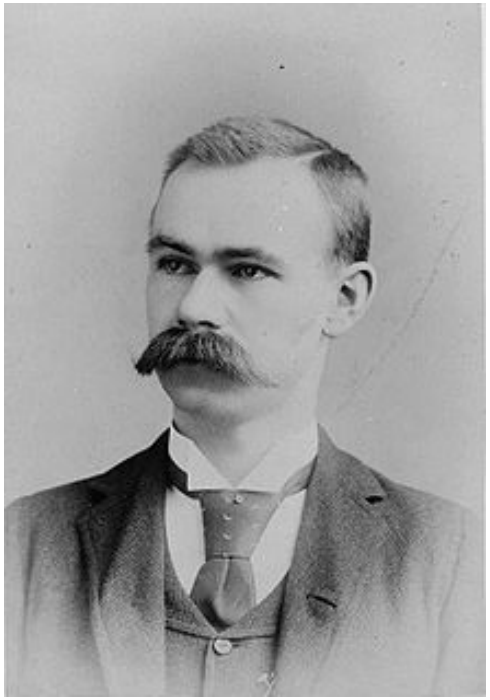


The Problem with Writing …

- Only **people** can read it … [Though recently, *some* progress in handwriting analysis has occurred; limited use.]
- First serious advance in digitization: punch cards
- Herman Hollerith develops idea for 1890 census

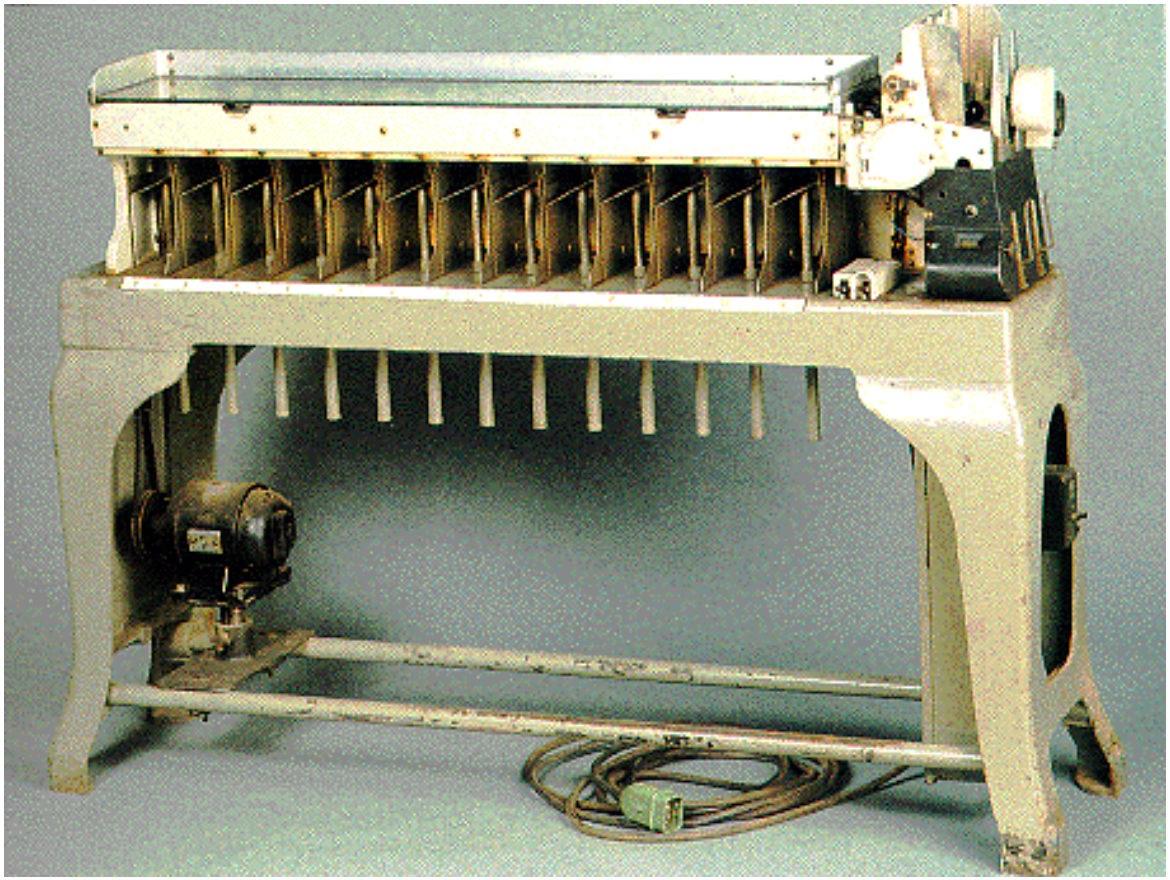


Hollerith Card, Courtesy IBM



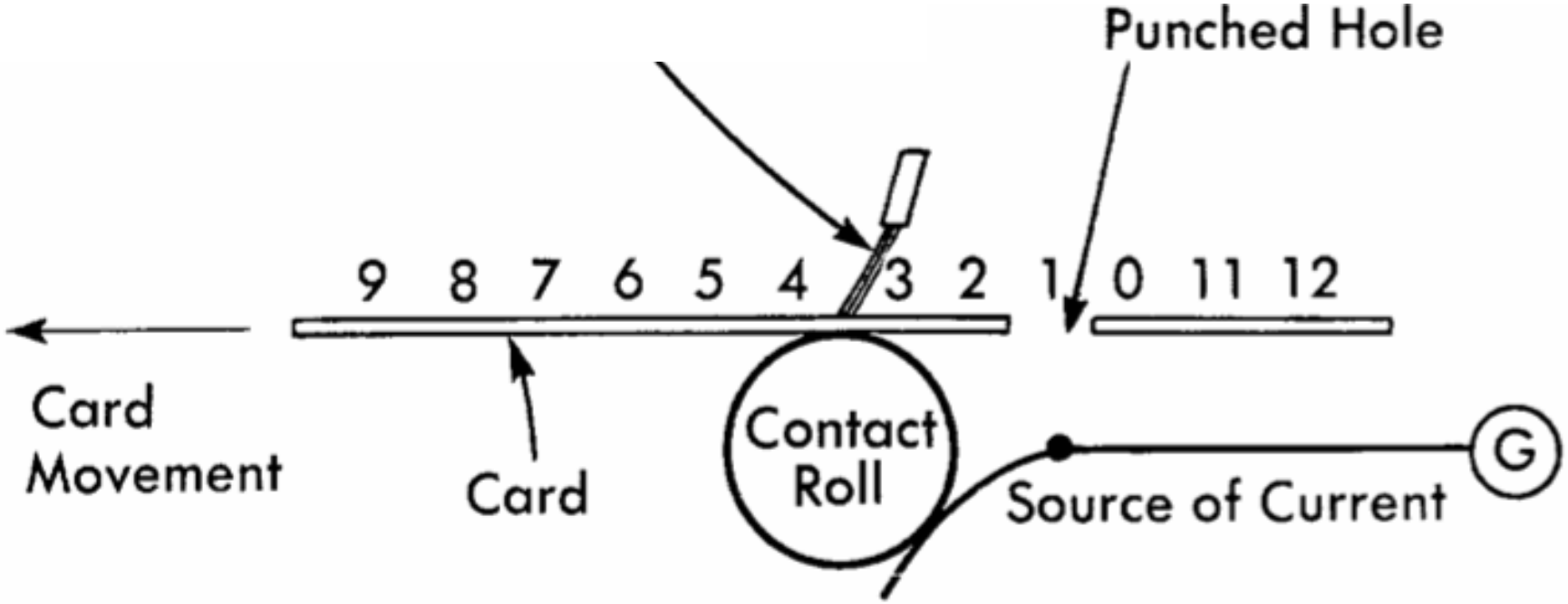
Machines Process Digital Data

- Mechanical methods – sensing a hole in a card or not – allows machines to help w/work



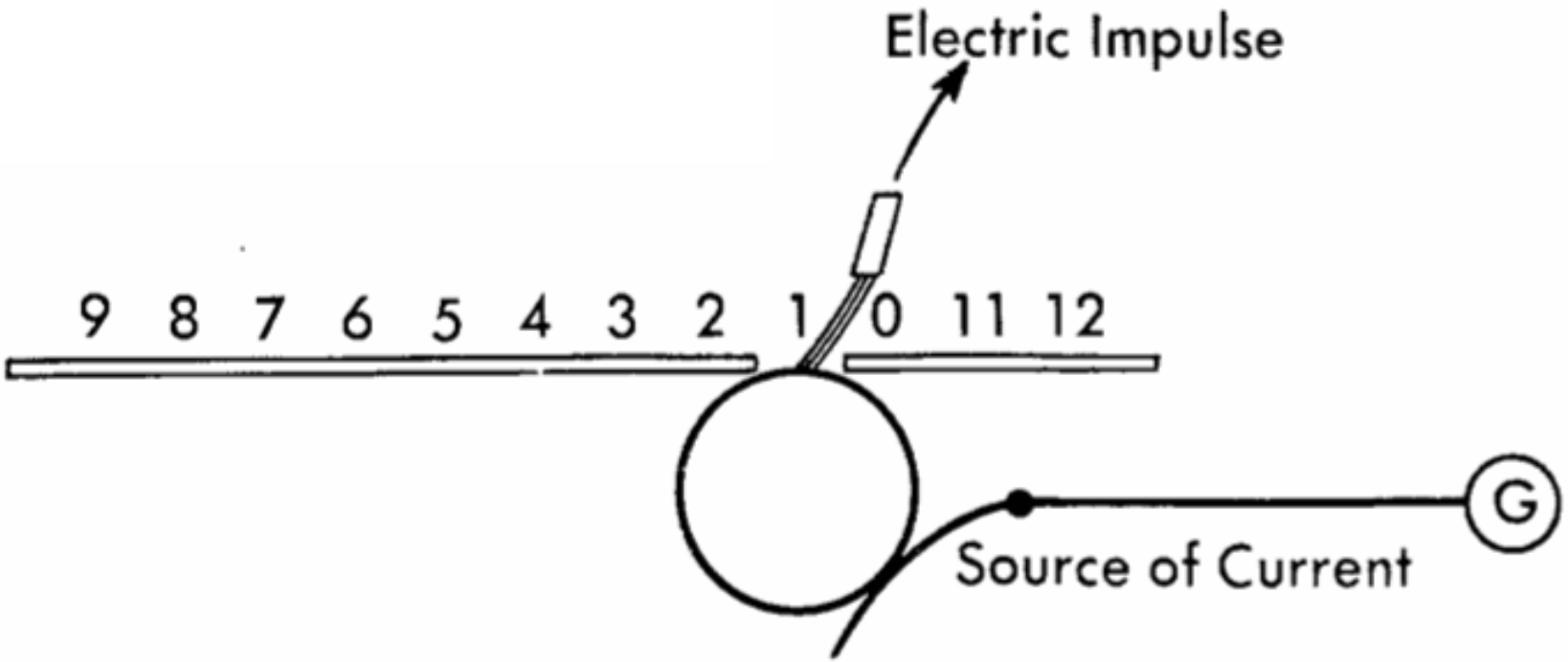
No Computer Needed To Process Data

- A mechanical machine can "read" a card with … a "metal brush"



Sensing Punch Allows Some Action

- When the circuit closes, some mechanical action can happen

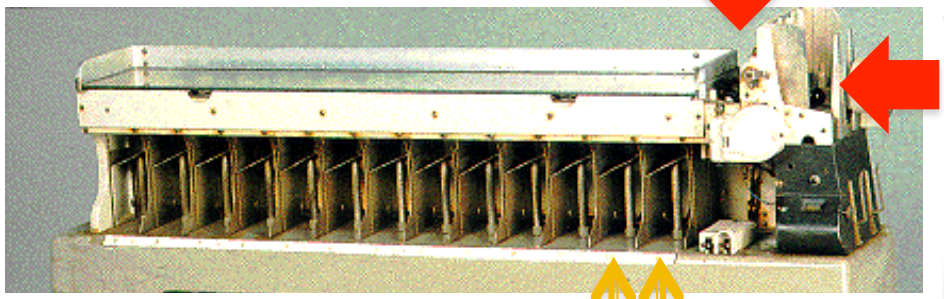


Computing w/o Computers

- Suppose Hollerith coded men as 0, women a 1

How many men and women in the population?

card counter



census data

Machine Reads Cards,
  Puts women in this slot
  Puts men in this slot
… producing 2 piles
Run each pile through again
  just to count them -- done

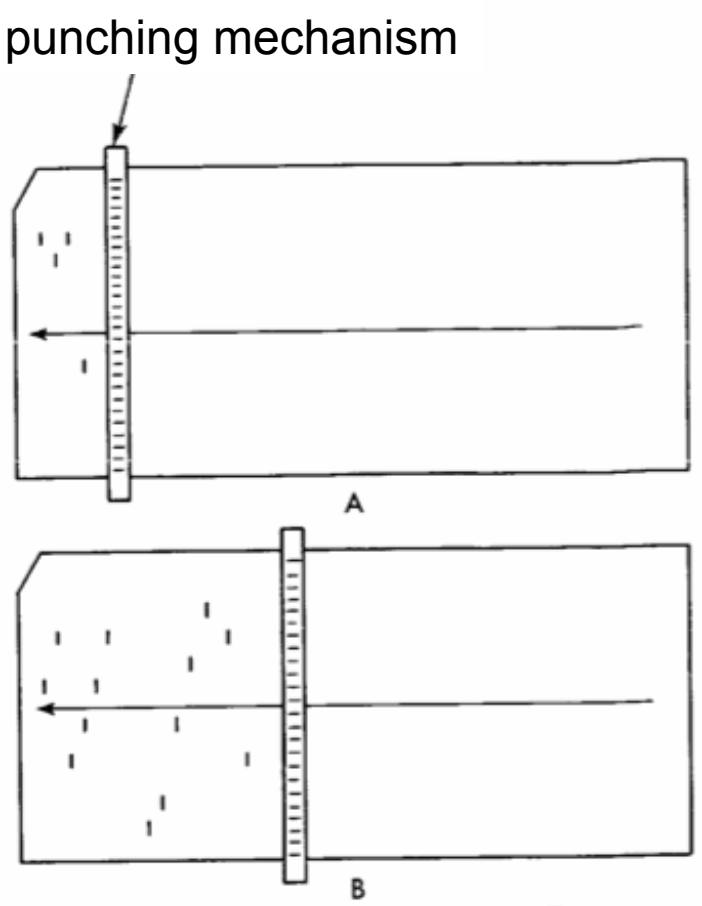
Meanwhile, w/o Digital Data

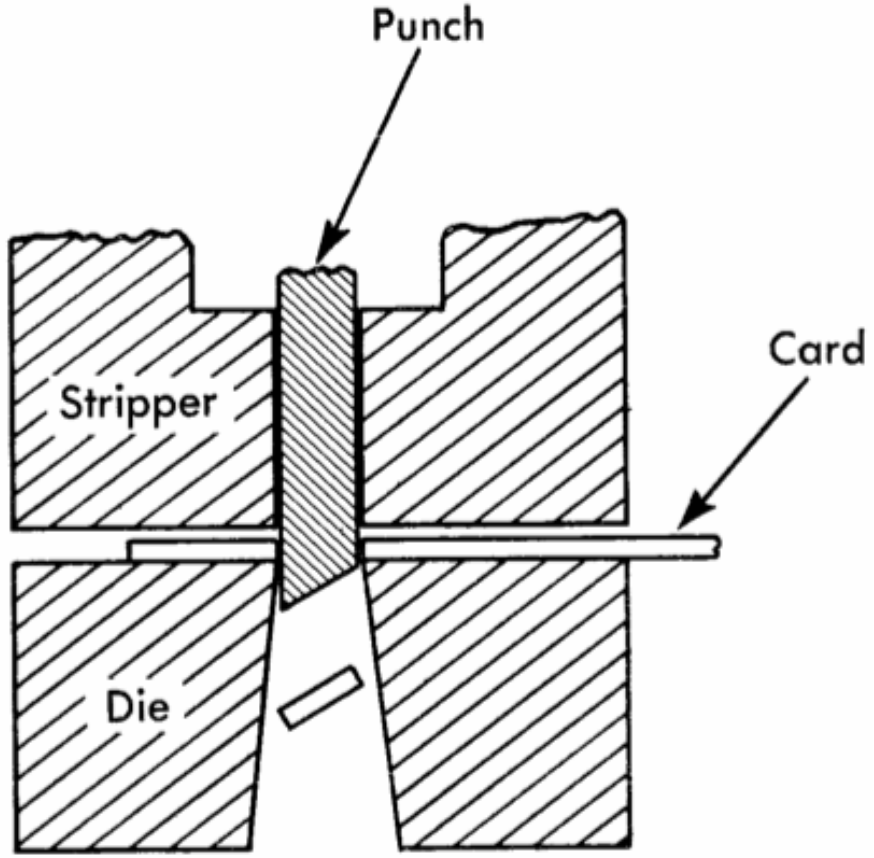
- Poor Kermit must go through census sheets, counting (and probably making mistakes)

The message: "Digitizing" makes information discrete, it's either there (1) or not (0), and a machine can determine that fact using mechanical or electronic means. Once data is digital, it is just a matter for engineers to build more capable machines

Writing As Important As Reading

- After processing based on reading cards, a machine can "save its work" by punching cards





Staying Digital

Next Big Things … Very Big!

- Electronic computers came after WWII

By Mid 20th Century ~ 1960

- Large and medium-size companies used card based digital data; mechanical processing
- Computers began to replace mechanical b/c a computer's "processing instructions" (program) could be easily changed, & they perform more complex operations – flexibility
- Computers, memory much more expensive –

Message: Computers take the task specification (program) and digital data as inputs, making them very versatile machines; one machine does it all! Programming becomes critical technology.

Key to Modern "Computers"

The key difference between the early tabulating machines and modern computers is

A. mechanical vs electrical

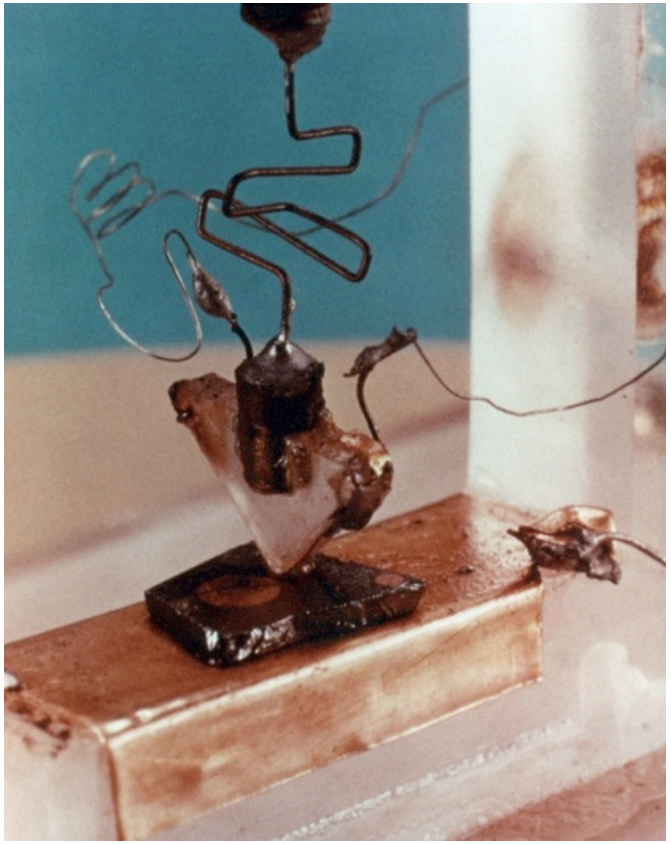
B. non-programmable vs programmable

C. decimal vs binary

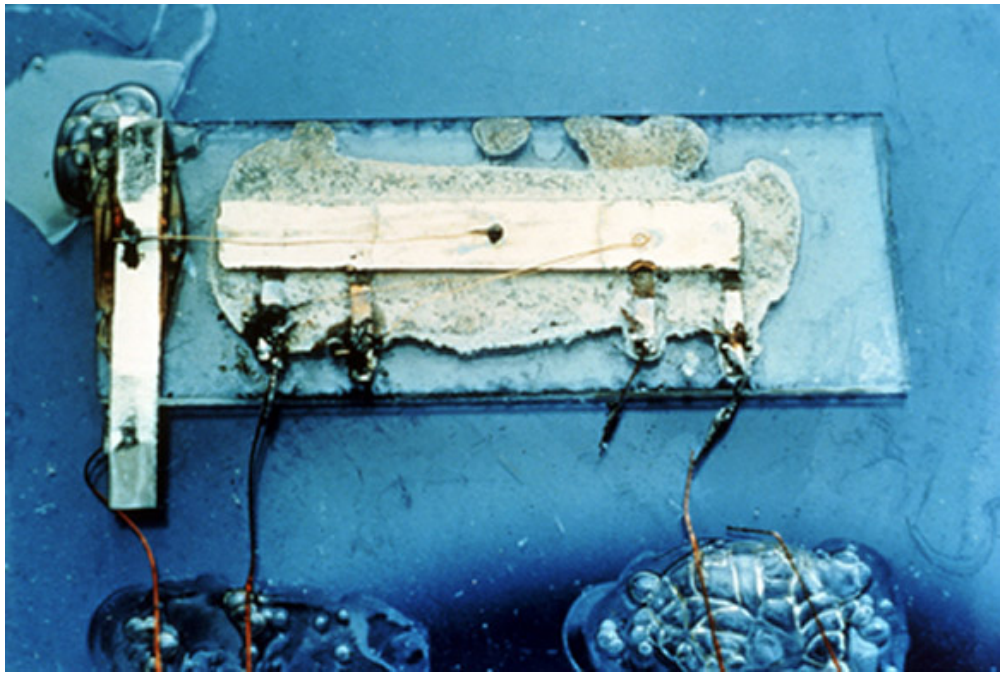
D. all of the above

Next Big Things: Integrated Circuits

- Transistors – solid state switching
- Integrated Circuit – all circuit parts fabbed at once from similar



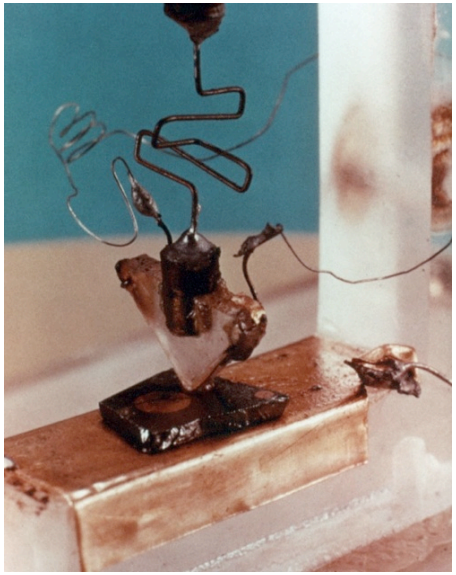
1st transistor

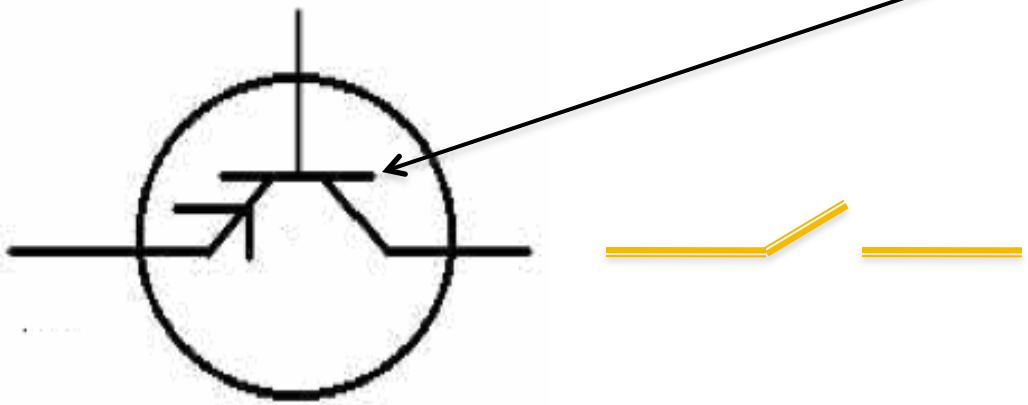


1st integrated circuit

Solid State Electronics

- A transistor is a switch: If the gate (black bar) is neutral, charge cannot pass; if gate is charged, the wires are connected

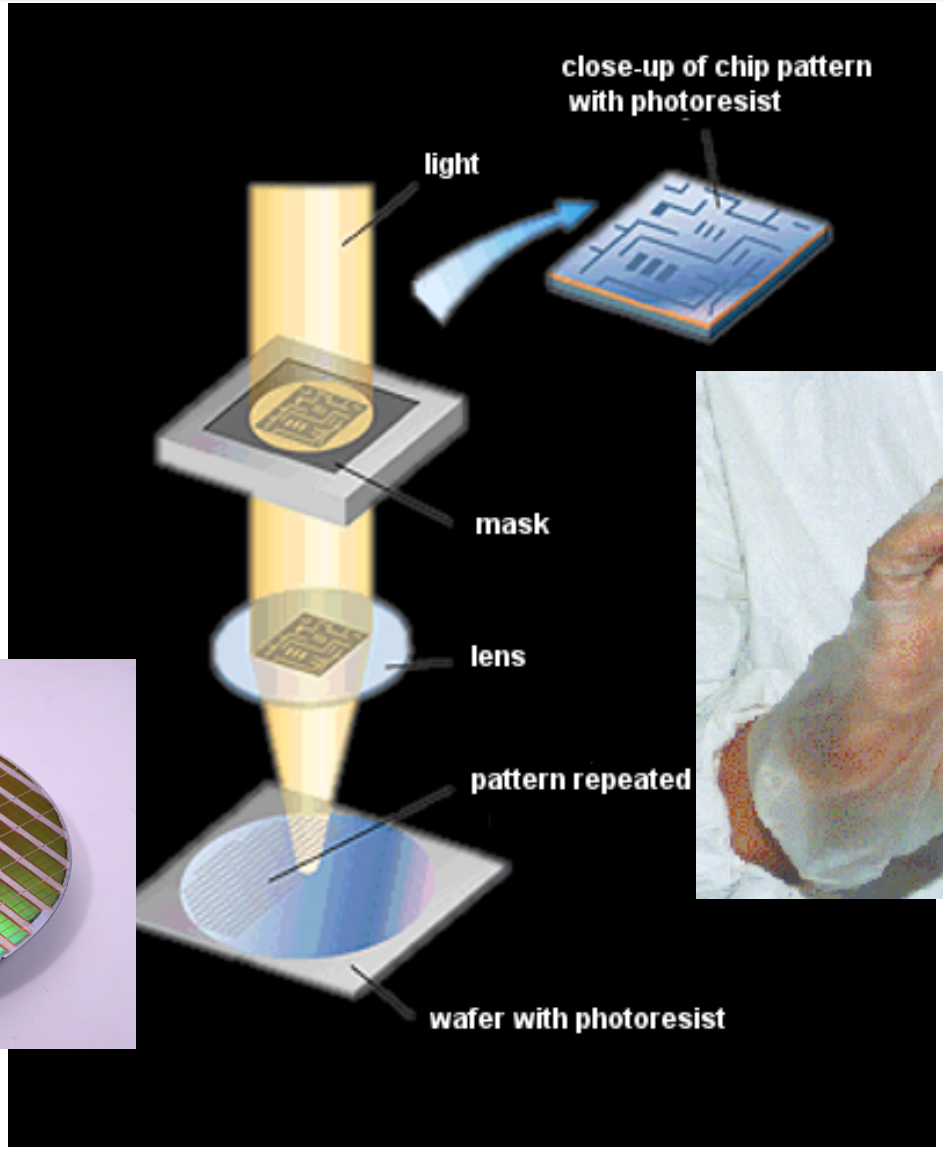


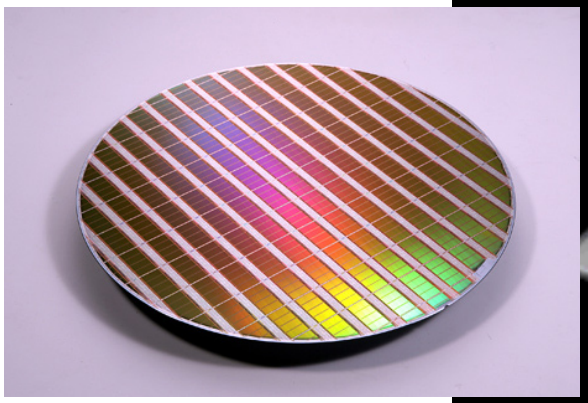


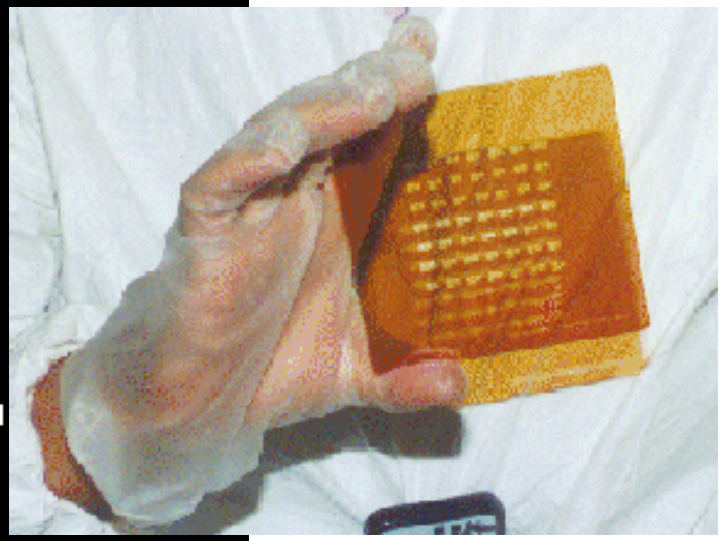
Solid State Electronics

- Transistors are smart, but "wiring them up" with other parts was labor intensive
- **Integrated circuits** – transistors + resistors + capacitors – are created together in one long recipe – small, cheap, reliable
- Key fabrication process is *photolithography* – the transistors are "printed" on the silicon!

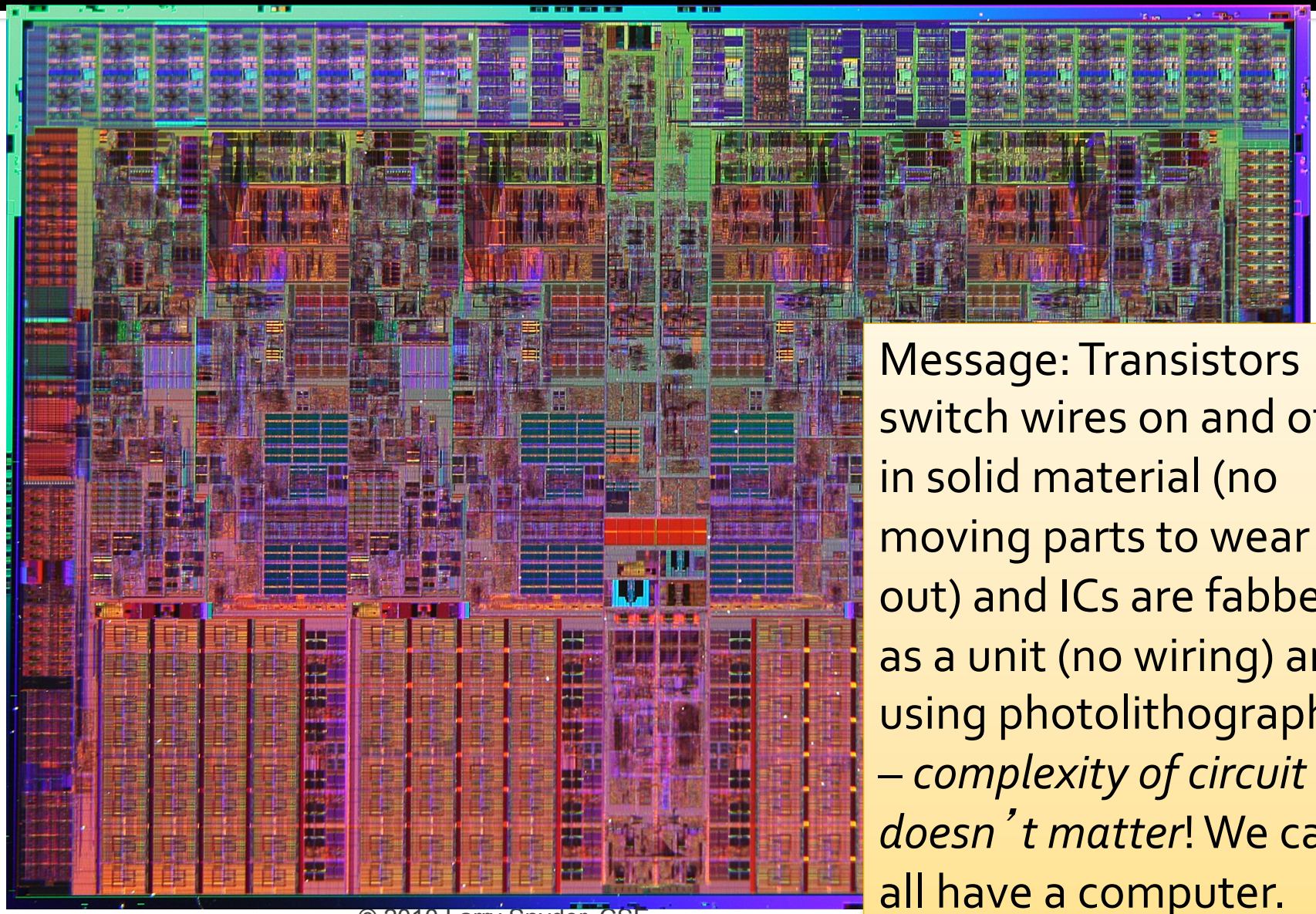
Photolithography







Integrated Circuits



Message: Transistors switch wires on and off in solid material (no moving parts to wear out) and ICs are fabbed as a unit (no wiring) and using photolithography – *complexity of circuit doesn't matter*! We can all have a computer.

Transistors

A transistor is most like

A. A hole (or lack of) in a punched card

B. A light switch

C. An entire punched card with multiple holes

D. A whole bank of light switches

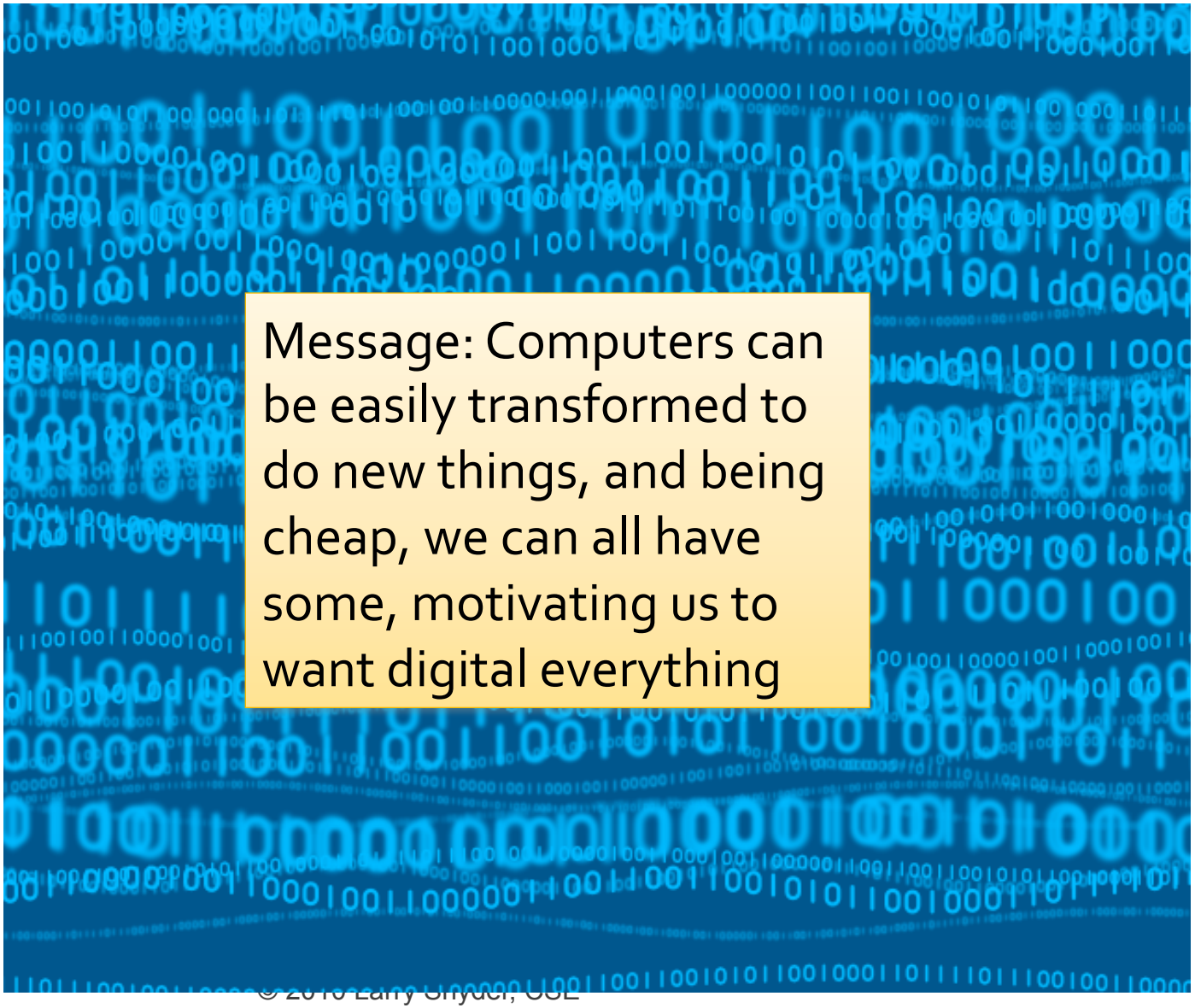
Next Big Thing: Personal Computers

- Ken Olsen, Founder of Digital Equipment, "There is no reason for any individual to have a computer in their home [1977]"

Computing Comes To Everyone

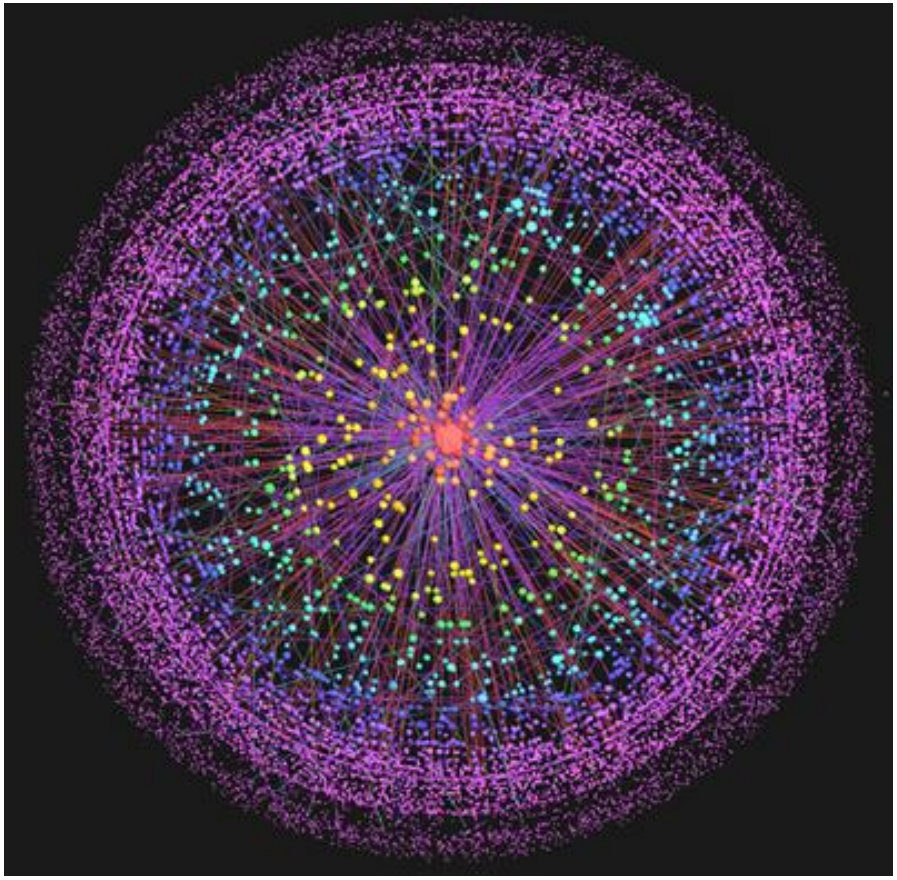
- Regular folks – not just government, military, scientists, banks and companies – could now apply computers to their interests
- Created a demand for digital data: news, pics, audio, video, books, etc., causing old technologies to digitize rapidly. Now it matters to everyone if a machine can "read" it
- From about 1985 most "new" information has been digital
- Quickly, people acquired enormous amounts of information

Digital Rocks



Next Big Thing: Internet

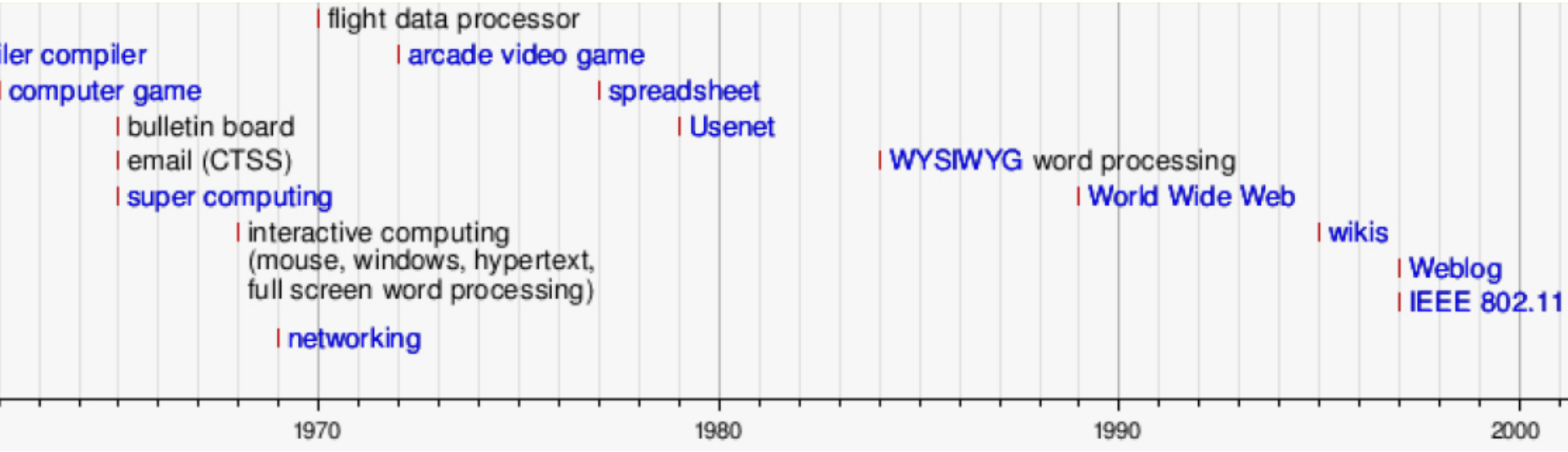
- Invented in 1969, it took almost 20 years to get out of the lab and into public consciousness



"On the Internet, nobody knows you're a dog."

Wikipedia Timeline

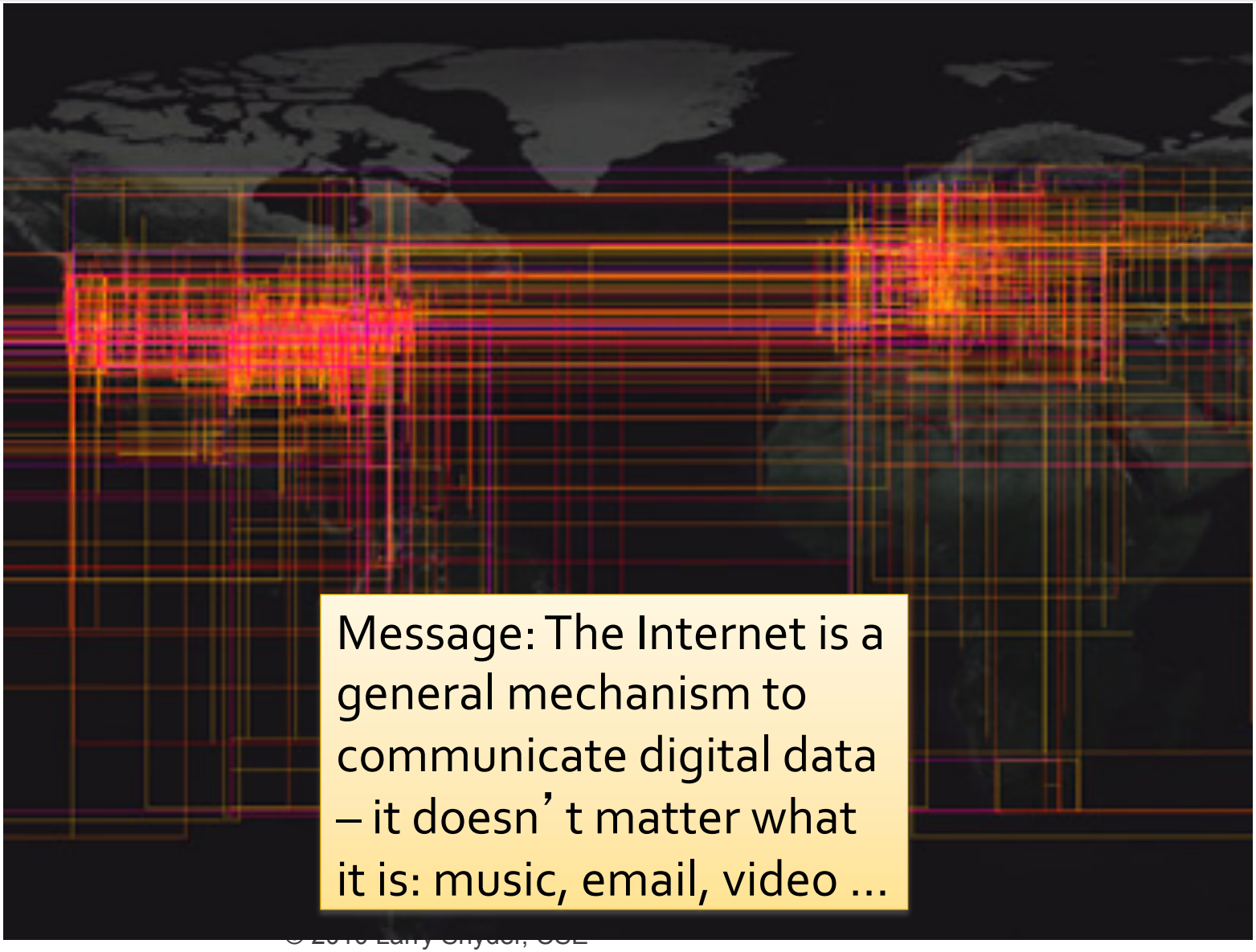
en.wikipedia.org/wiki/Timeline_of_computing



Connecting Up

- Computers are useful; connected computers are awesome
- If *n* computers are connected, adding one more gives *n* new connections!
- Communication with friends or businesses all over the world became easy and casual – some people even found out about time zones
- Digital media allows people to share each other's information at no cost

Connectivity to Change the World



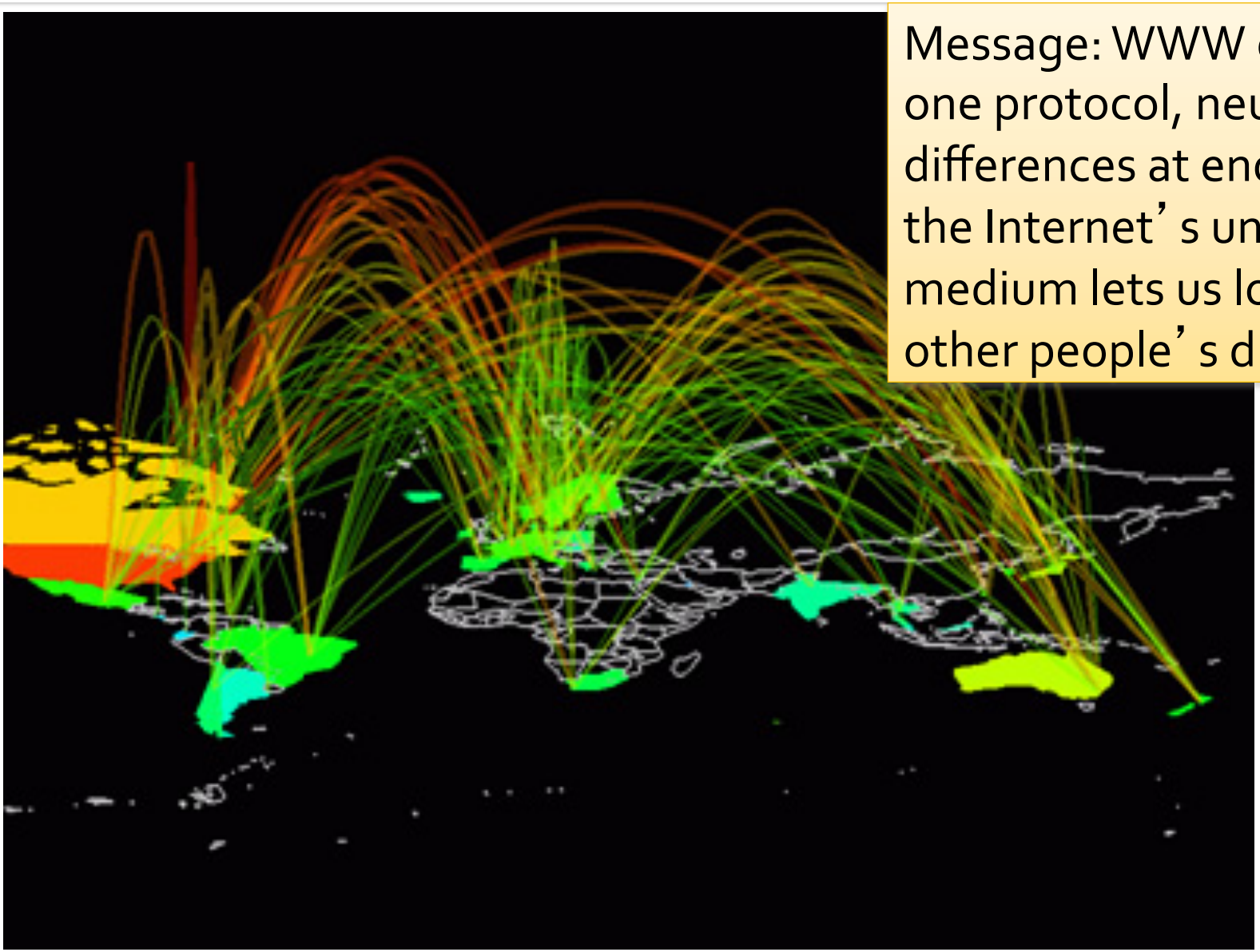
Next Big Thing: WWW + http

- Today, all computers "speak" a common language: hyper-text transfer protocol

WWW Is The Servers + The Data

- Two phenomena make the WWW brilliant
  - All computers use one standard protocol (http) meaning for once all of the world's people – who don't speak one language – have a surrogate that does
  - Publishing and accessing information is completely decentralized – generally, no one limits what you put out or go after

Seeing Other People's Digital Info



Message: WWW exploits one protocol, neutralizing differences at endpoints; the Internet's universal medium lets us look at other people's digital info

So, It All Works Because of Digital

- Key principle of digital encoding: Physically, information is the presence or absence of a phenomenon at a given place and time!
- Card example:
  - Phenomenon – hole in the card
  - Present – detected by brush making elec contact
  - Absent – brush insulated from electrical source
  - Place – there are several on the card; devices can know the positions
  - Time – hole is permanent representation of info

A General Idea

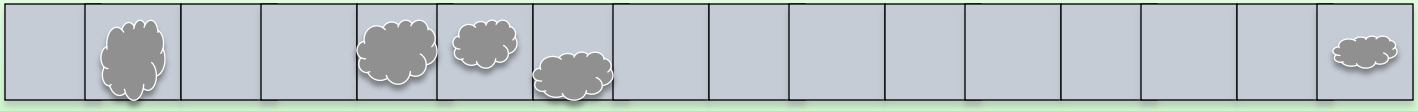
- Digital Information: Detecting the presence or absence of a phenomenon at a specific place and time: PandA
- Phenomena: light, magnetism, charge, mass, color, current, …
- Detecting depends on phenomenon – but the result must be discrete: it was detected or not; there is no option for "sorta there"
- Place and time apply, but usually default to "obvious" values; not so important to us

Digital Discussion

- Alternatives to detecting the hole in a card

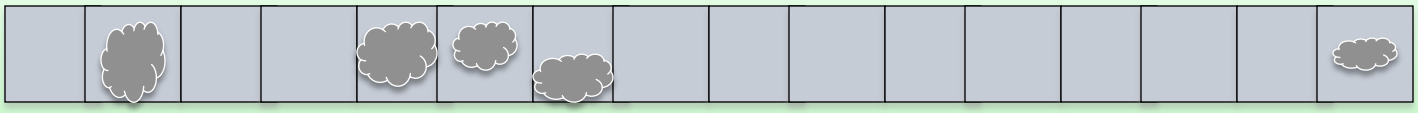
Digital Discussion

- Alternatives to detecting the hole in a card
- Sidewalk Memory – squares and rocks

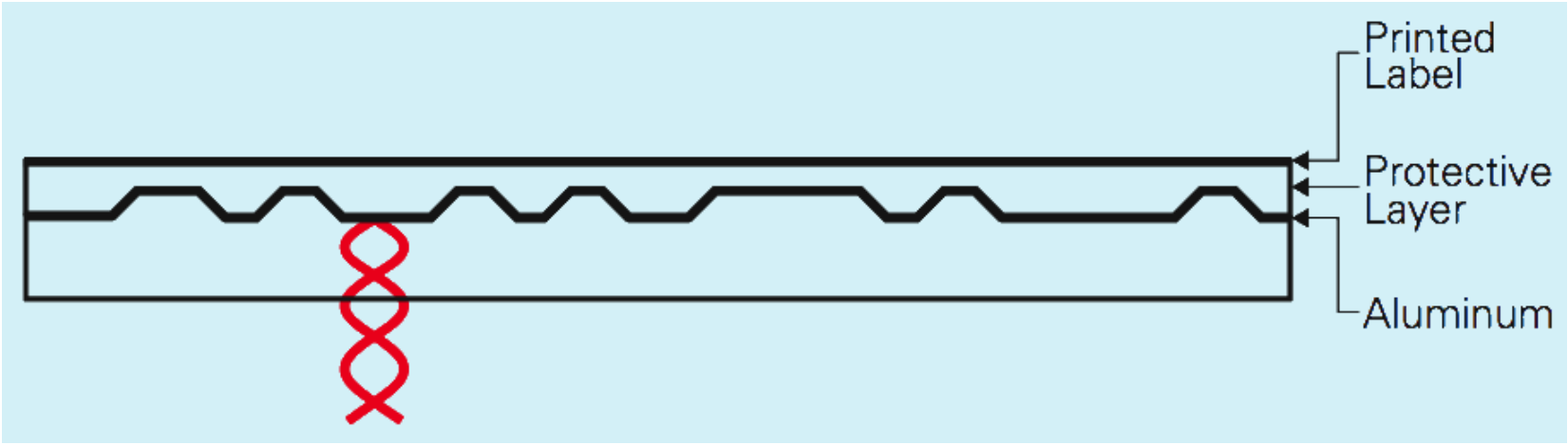


Digital Discussion

- Alternatives to detecting the hole in a card
- Sidewalk Memory – squares and rocks

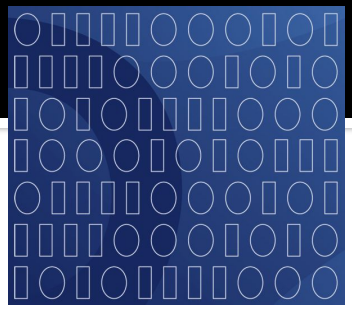


- Other phenomena … CD ROM how it works:



Bits





- PandA is a *binary representation* because it uses 2 patterns

  Bit -- it's a contraction for "binary digit"

  -- a position in space/time capable of being set and detected in 2 patterns

Bytes: Standard encodings of meaning

- A byte is eight bits treated as a unit
  - Adopted by IBM in 1960s
  - A standard measure ever since
  - Bytes encode the Latin alphabet using ASCII -- the American Standard Code for Information Interchange

0101 0101
0101 0111

ASCII

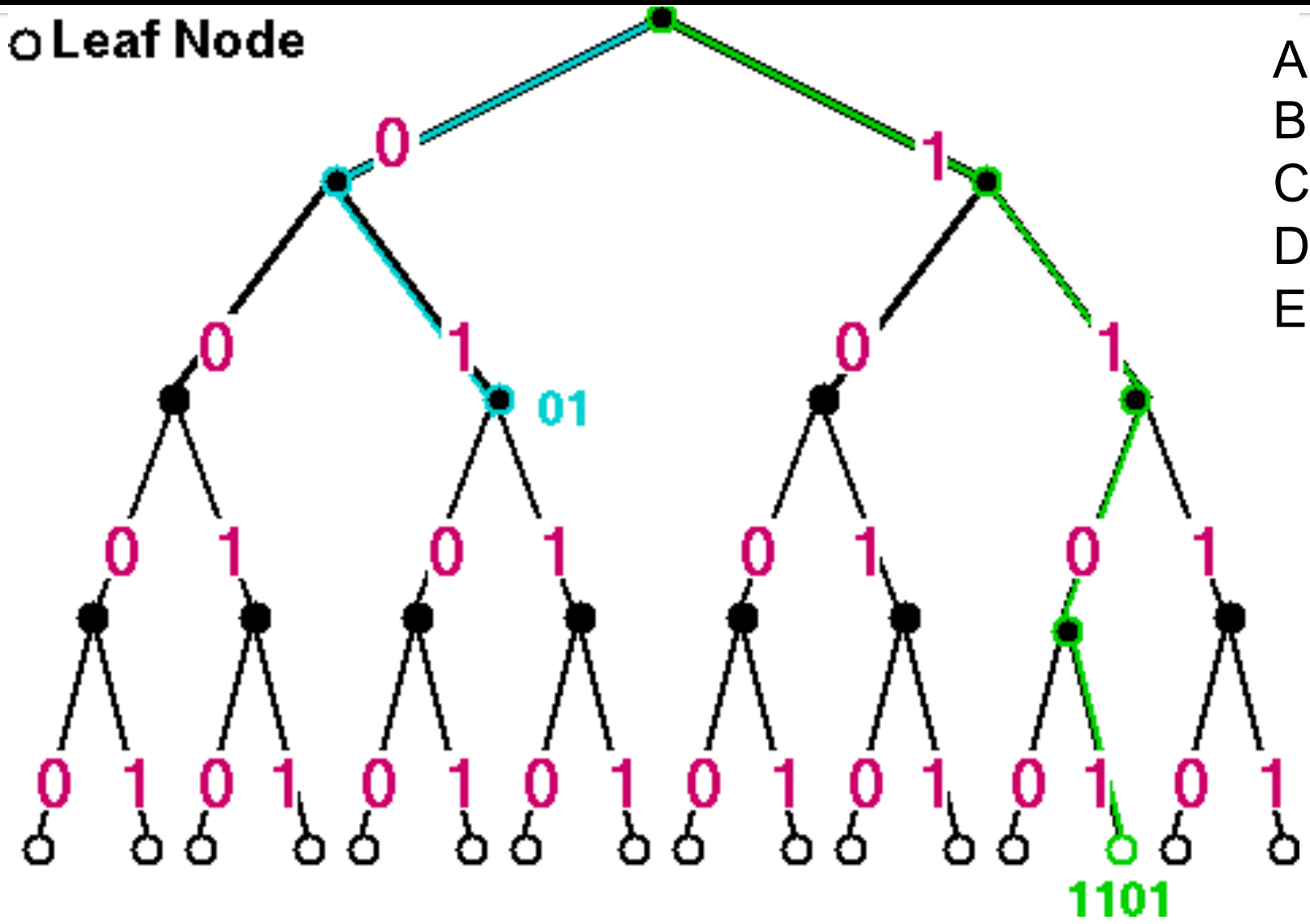
0100 0011
0100 0001
0101 0100

C
A
T

ASCII	0 0 0 0	0 0 0 1	0 0 1 0	0 0 1 1	0 1 0 0	0 1 0 1	0 1 1 0	0 1 1 1	1 0 0 0	1 0 0 1	1 0 1 0	1 0 1 1	1 1 0 0	1 1 0 1	1 1 1 0	1 1 1 1
0000	$^{N}_{U}$	$^{S}_{H}$	$^{S}_{X}$	$^{E}_{X}$	$^{E}_{T}$	$^{E}_{Q}$	$^{A}_{K}$	$^{B}_{L}$	$^{B}_{S}$	$^{H}_{T}$	$^{L}_{F}$	$^{Y}_{T}$	$^{F}_{F}$	$^{C}_{R}$	$^{S}_{0}$	$^{S}_{I}$
0001	$^{D}_{L}$	$^{D}_{1}$	$^{D}_{2}$	$^{D}_{3}$	$^{D}_{4}$	$^{N}_{K}$	$^{S}_{Y}$	$^{E}_{\Sigma}$	$^{C}_{N}$	$^{E}_{M}$	$^{S}_{B}$	$^{E}_{C}$	$^{F}_{S}$	$^{G}_{S}$	$^{R}_{S}$	$^{U}_{S}$
0010		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
0011	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
0110	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	$^{D}_{T}$
1000	$^{8}_{0}$	$^{8}_{1}$	$^{8}_{2}$	$^{8}_{3}$	$^{I}_{N}$	$^{N}_{L}$	$^{S}_{S}$	$^{E}_{S}$	$^{H}_{S}$	$^{H}_{J}$	$^{Y}_{S}$	$^{P}_{D}$	$^{P}_{V}$	$^{R}_{I}$	$^{S}_{2}$	$^{S}_{3}$
1001	$^{D}_{C}$	$^{P}_{1}$	$^{P}_{Z}$	$^{S}_{E}$	$^{C}_{C}$	$^{M}_{M}$	$^{S}_{P}$	$^{E}_{P}$	$^{Q}_{8}$	$^{Q}_{Q}$	$^{Q}_{A}$	$^{C}_{S}$	$^{S}_{T}$	$^{O}_{S}$	$^{P}_{M}$	$^{A}_{P}$
1010	$^{A}_{O}$	¡	¢	£	♀	¥	¦	§	¨	©	♂	«	¬	-	®	¯
1011	°	±	^{2}	^{3}	´	µ	¶	·	¸	^{1}	º	»	¼	½	¾	¿
1100	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
1101	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	β
1110	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
1111	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

0100 0111|0110 1111|0010 0000| 0101 0011|0110 1100|0111 0101|0110 0111|0111 0011|

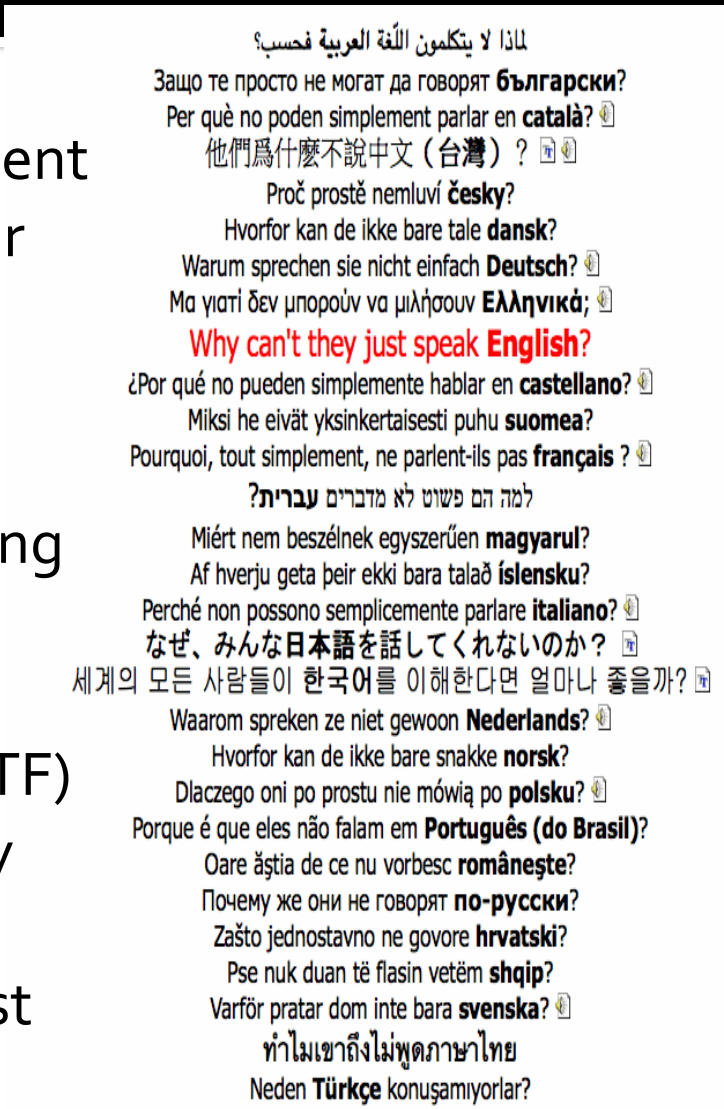
With 8 places how many different letters?



A. 32
B. 64
C. 128
D. 256
E. 512

UTF-8: All the alphabets in the world

- Uniform Transformation Format: a variable-width encoding that can represent every character in the Unicode Character set
- 1,112,064 of them!!!
- http://en.wikipedia.org/wiki/UTF-8
- UTF-8 is the dominant character encoding for the World-Wide Web, accounting for more than half of all Web pages.
- The Internet Engineering Task Force (IETF) requires all Internet protocols to identify the encoding used for character data
- The supported character encodings must include UTF-8.



Recap: Next Big Thing

- Electronic Computers
- Integrated Circuits
- Personal Computers
- Internet
- WWW + http
- ???