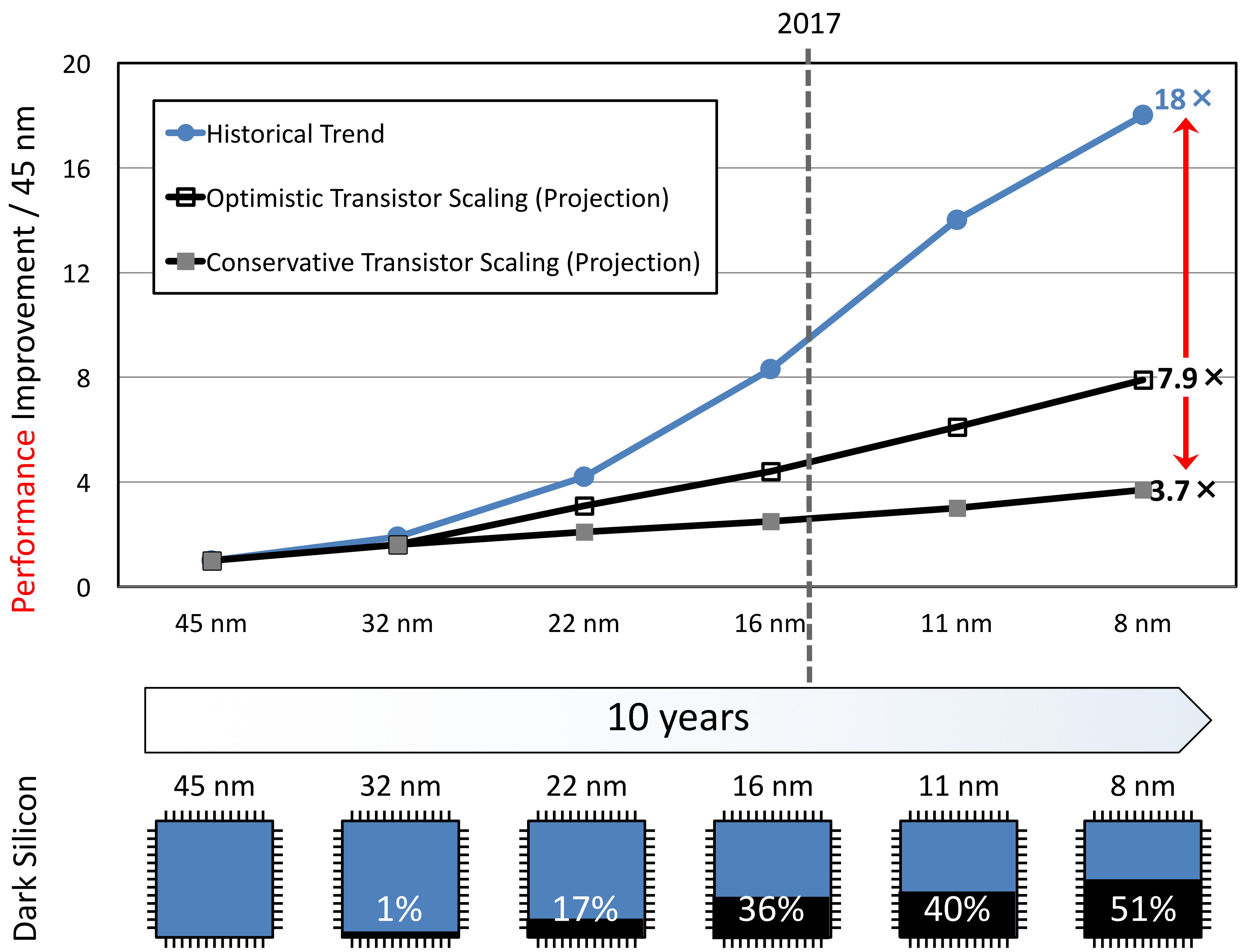
CSE 548: Computer Systems Architecture

Trends

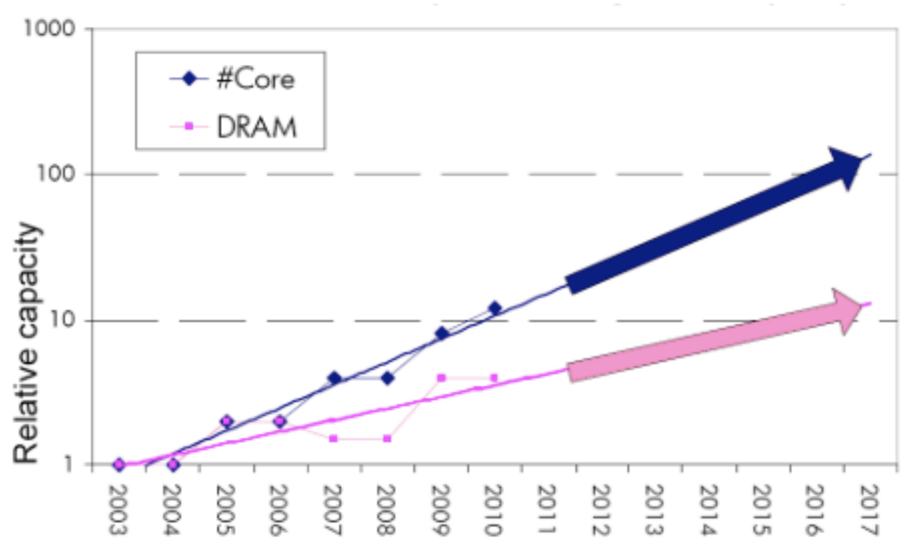
Luis Ceze, Spring 2017



2017
20
16
12
8
4
0
Performance Improvement / 45 nm
Historical Trend
Optimistic Transistor Scaling (Projection)
Conservative Transistor Scaling (Projection)
18 ×
7.9 ×
3.7 ×
45 nm
32 nm
22 nm
16 nm
11 nm
8 nm
10 years
Dark Silicon
45 nm
32 nm
22 nm
16 nm
11 nm
8 nm
1%
17%
36%
40%
51%

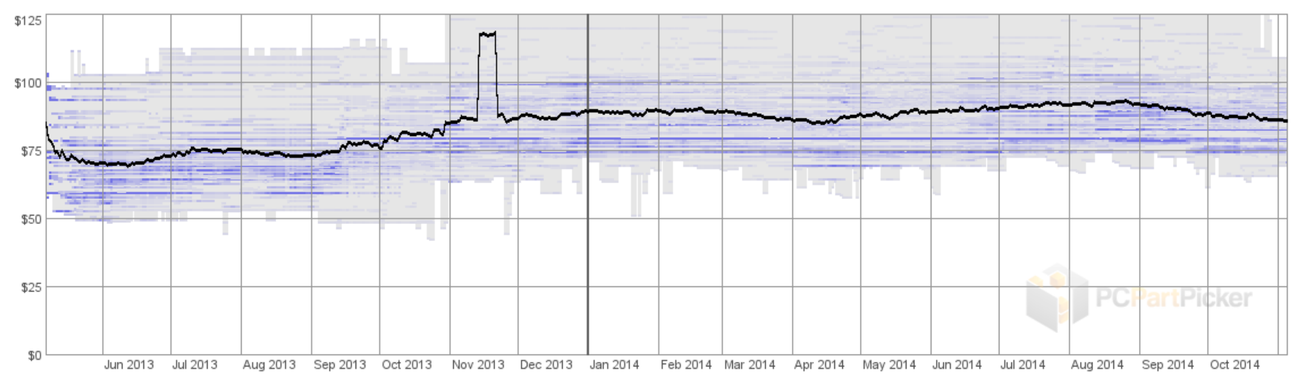
Related trend: DRAM

- Memory capacity per core dropping fast (30%/yr)
- Trends worse for memory bandwidth per core
- ITRS projects for sub 20nm not encouraging



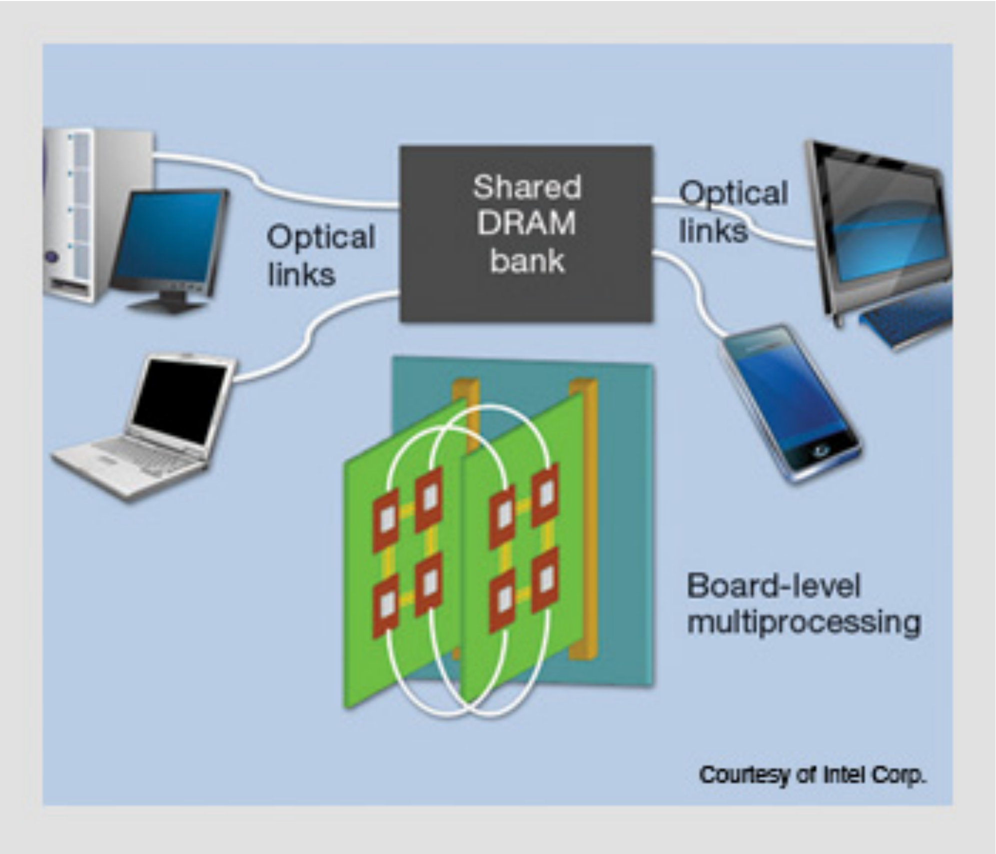
Source: Lim et al., ISCA 2009.

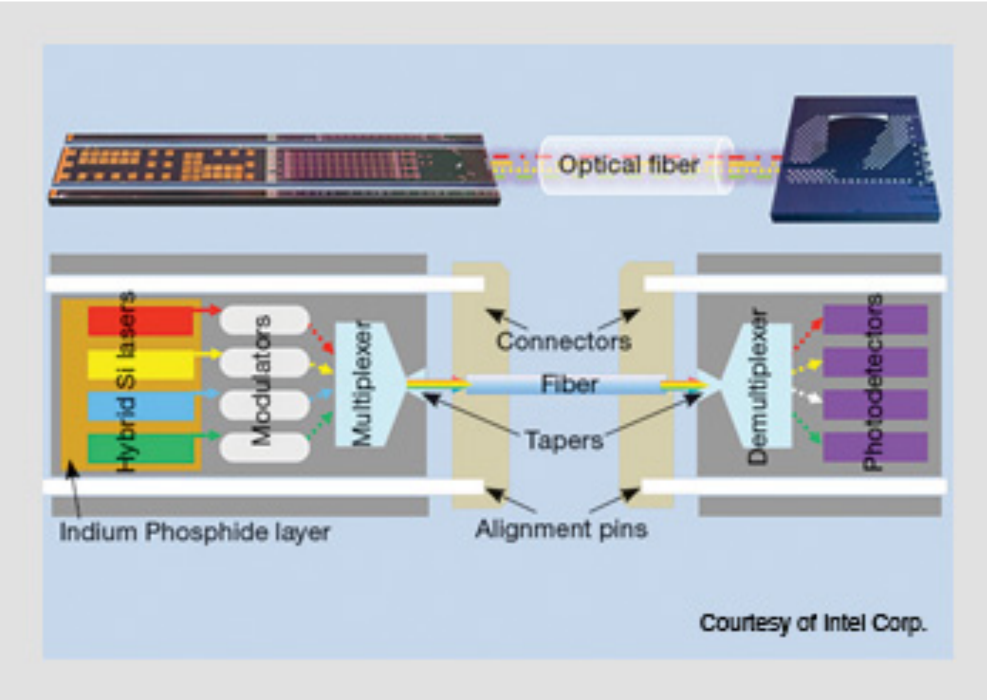
And btw, DRAM price keeping steady for a while (though that might change)…



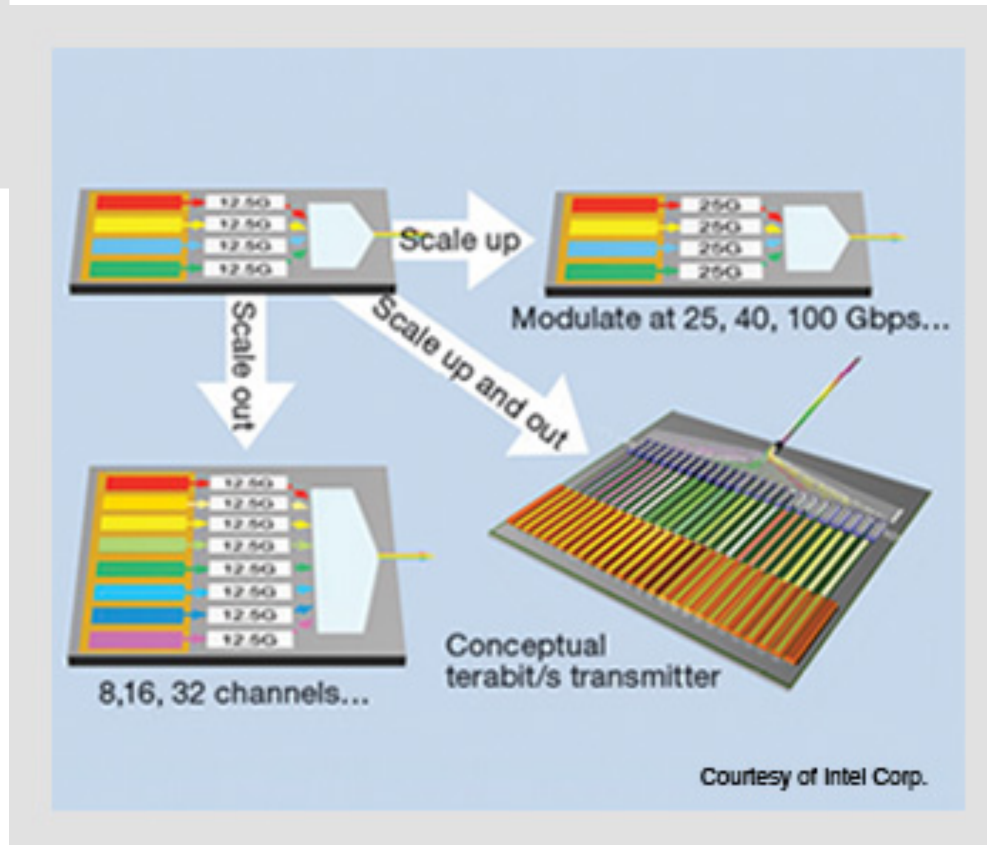
Average RAM Price (USD) Over Last 18 Months (DDR3-1333 240-pin DIMM 2x4GB) -- pcpartpicker.com

Silicon Photonics





Optical fiber
Hybrid Si lasers
Modulators
Multiplexer
Connectors
Fiber
Tapers
Demultiplexer
Photodetectors
Indium Phosphide layer
Alignment pins
Courtesy of Intel Corp.



12.5G
Scale up
25G
Modulate at 25, 40, 100 Gbps...
Scale out
Scale up and out
8,16, 32 channels...
Conceptual
terabit/s transmitter
Courtesy of Intel Corp.

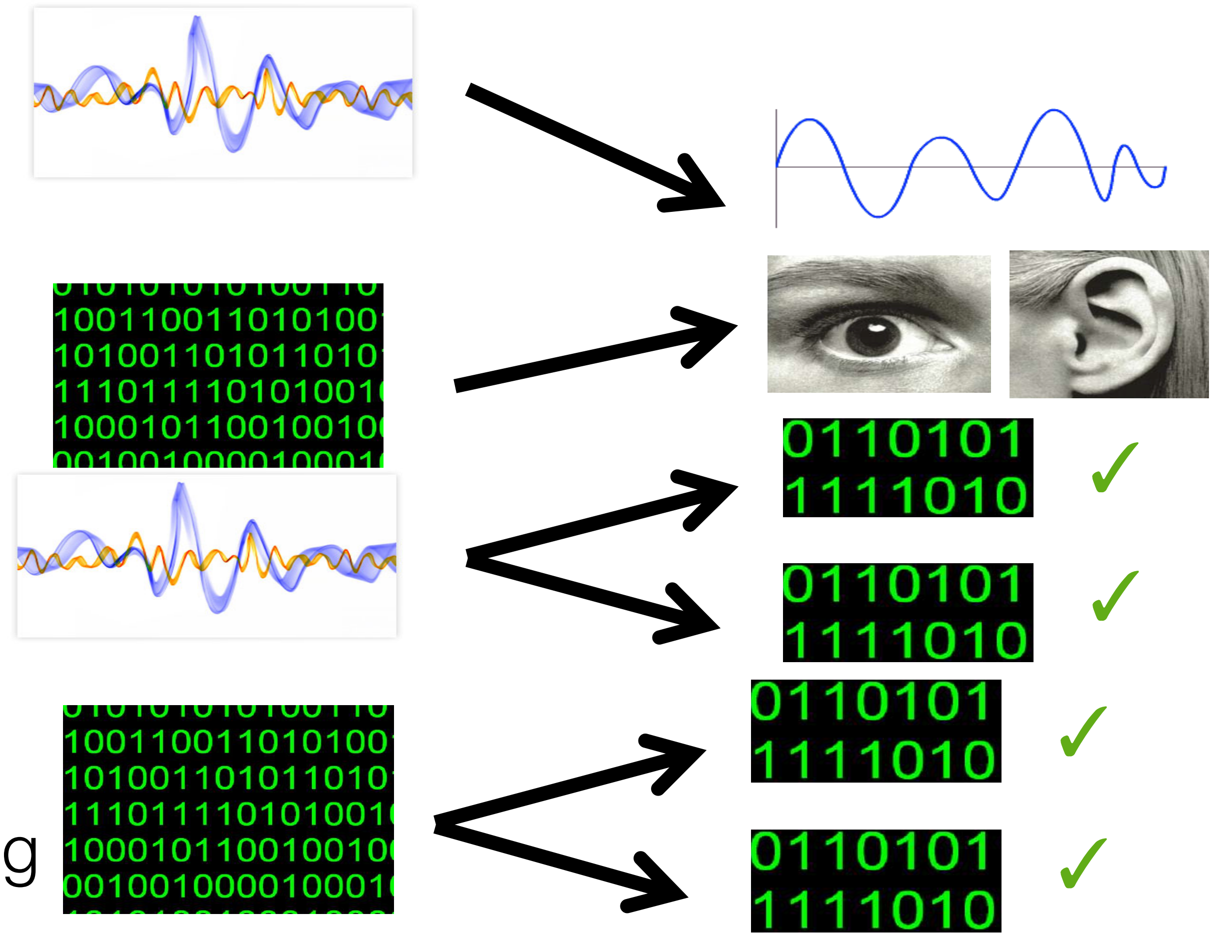
Modern Applications

image, sound
and video processing

image rendering

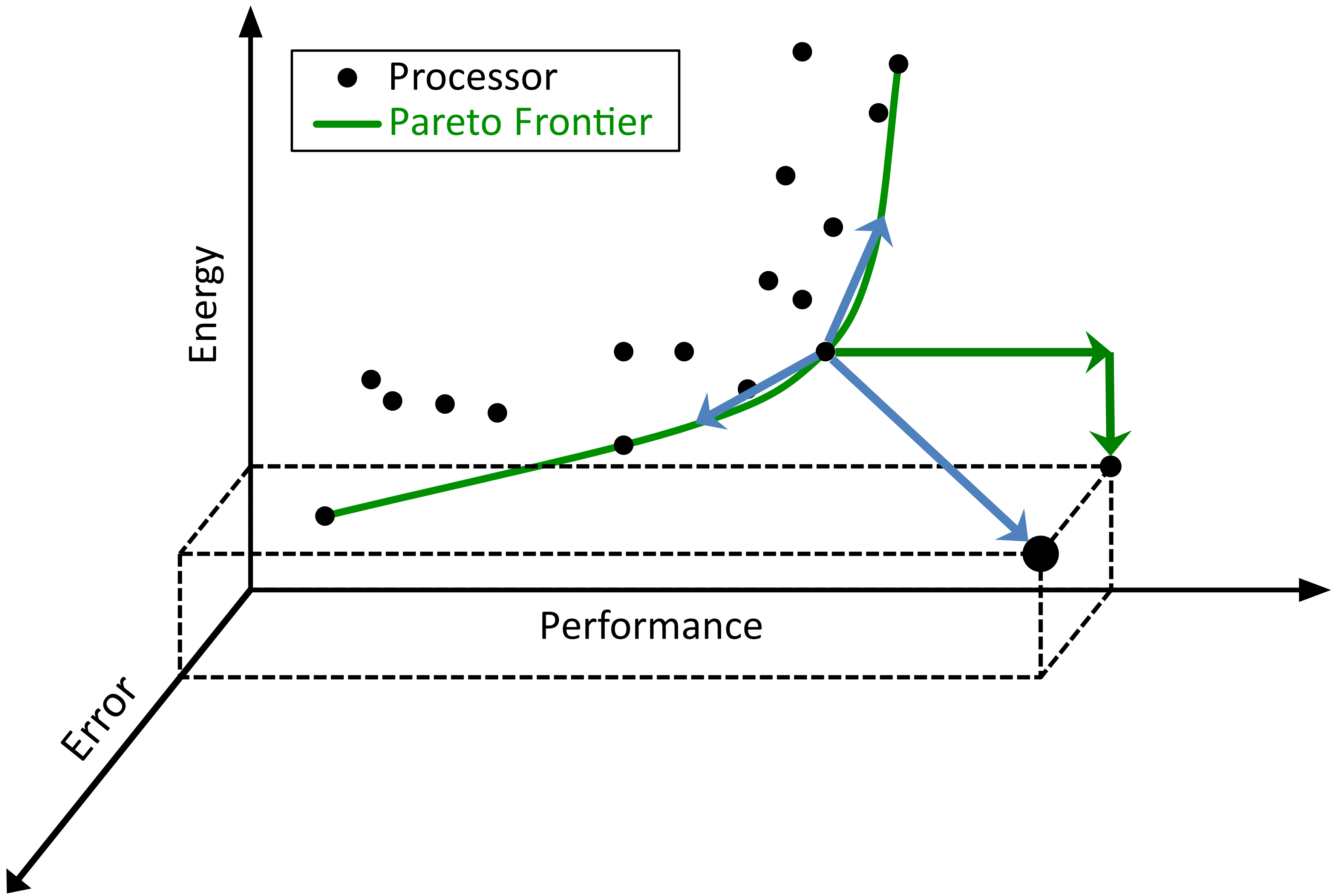
sensor data analysis,
computer vision

simulations, games,
search, machine learning



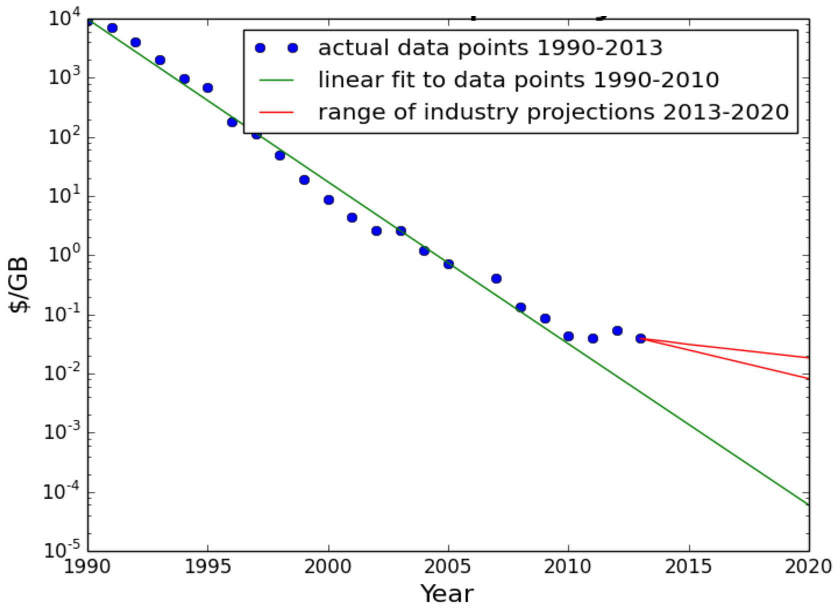
Inexact input data
Approximate/iterative algorithms
Malleable output

Adding a third dimension
Embracing Error



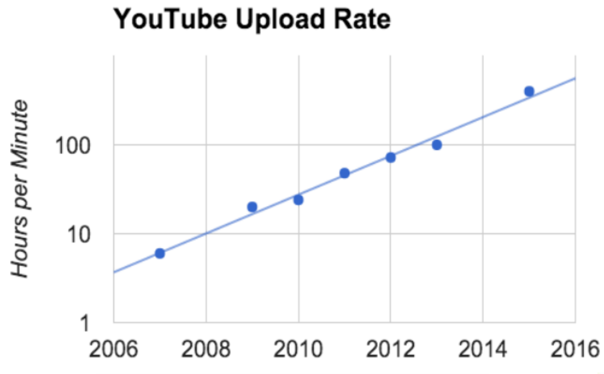
A widening practical gap…

Disk cost-per-byte is not decreasing fast enough

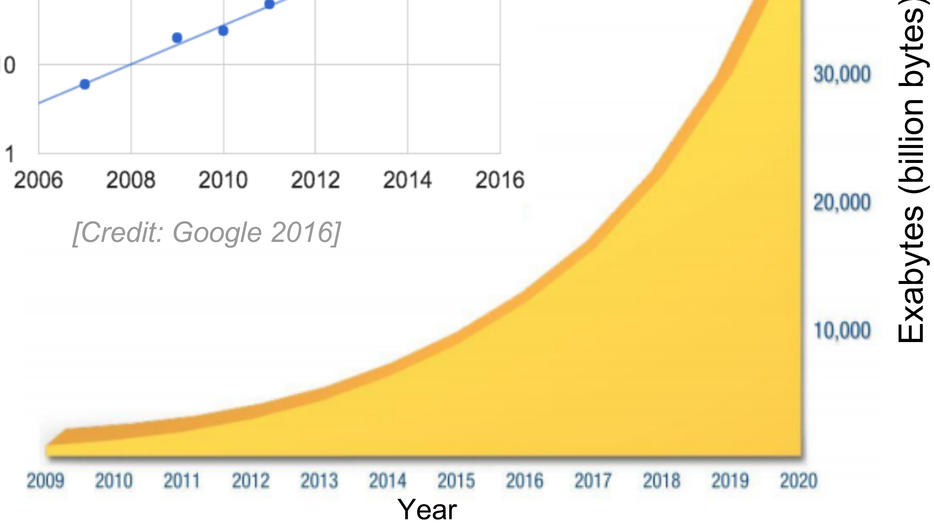


[Credit: David Rosenthal (CMU) and Preeti Gupta (UCSC), 2014]

Information growth

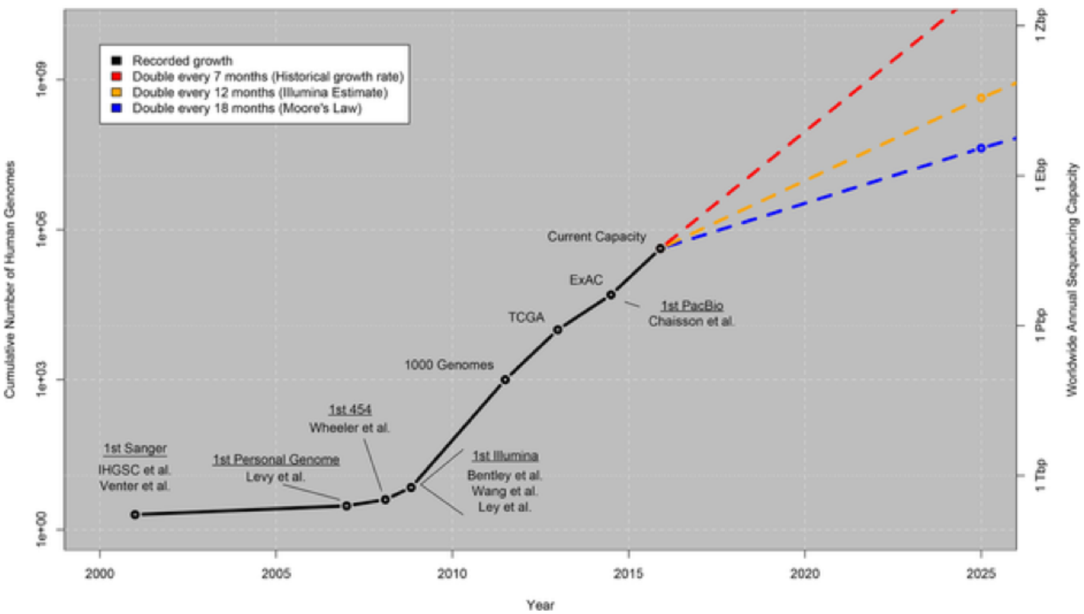


[Credit: Google 2016]

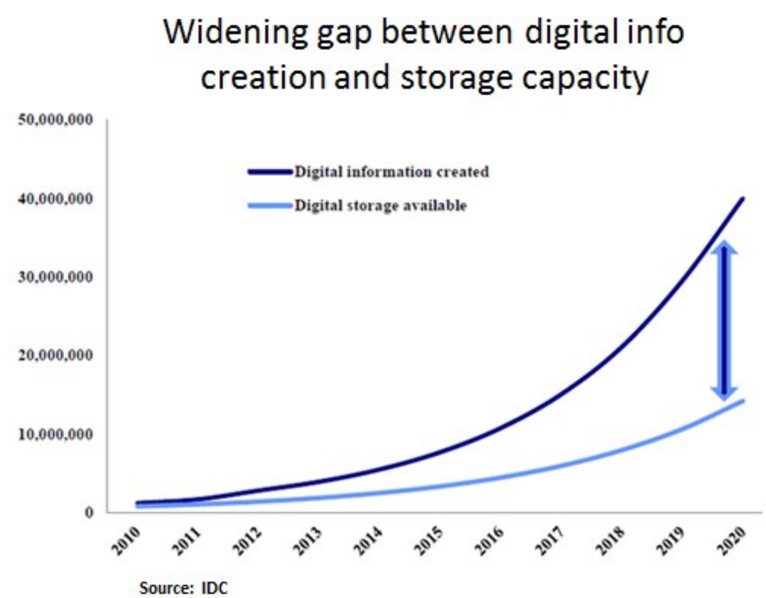


[Credit: EMC 2012]

Is it an inevitable gap?



Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015) Big Data: Astronomical or Genomical?. PLoS Biol 13(7): e1002195. doi:10.1371/journal.pbio.1002195
http://127.0.0.1:8081/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002195

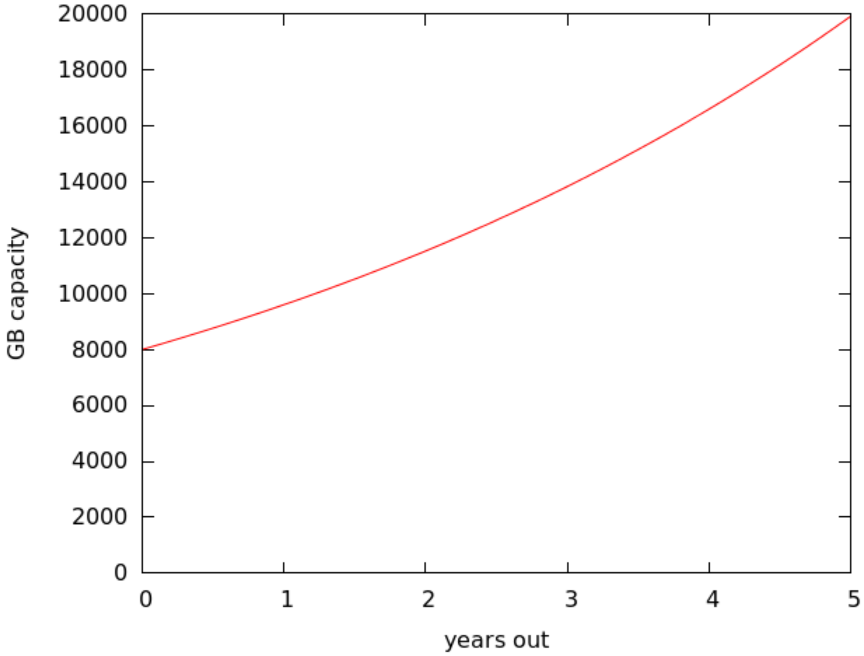


Disks & Data Centers

Larry Greenfield leg@google.com

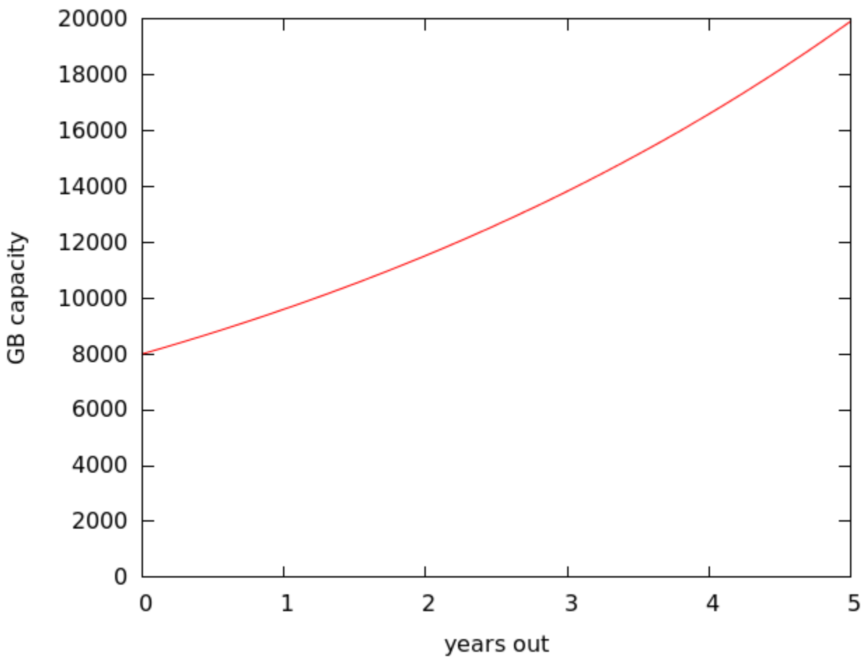
Two big problems...

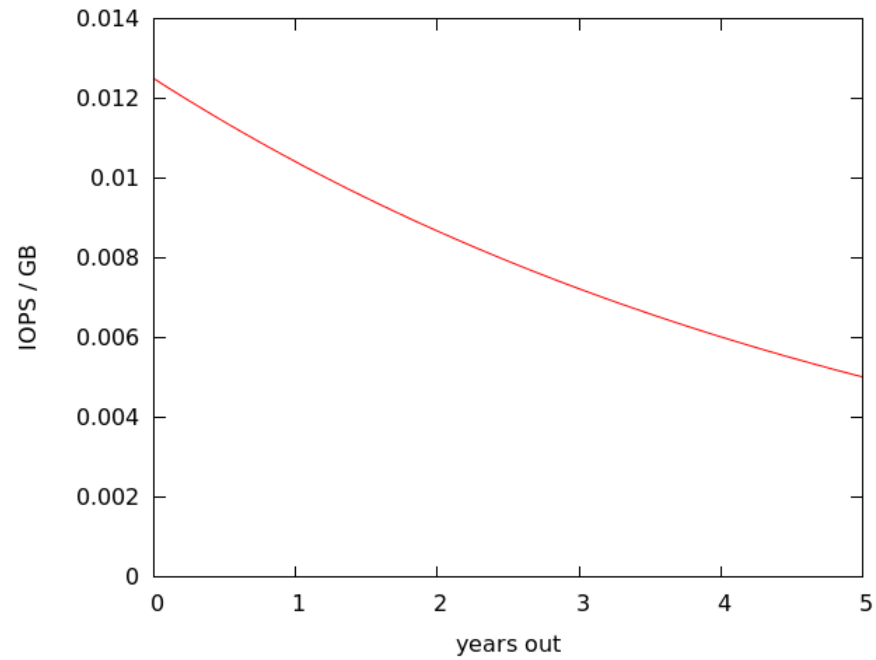
If capacity improvements are slowing to 20% y/y growth, will video storage become uneconomical?



Two big problems...

Or will e-mail storage become uneconomical because we need more IOPS?





Outlook for Memory and Storage

ISAT Future of Storage Workshop 16 May 2016

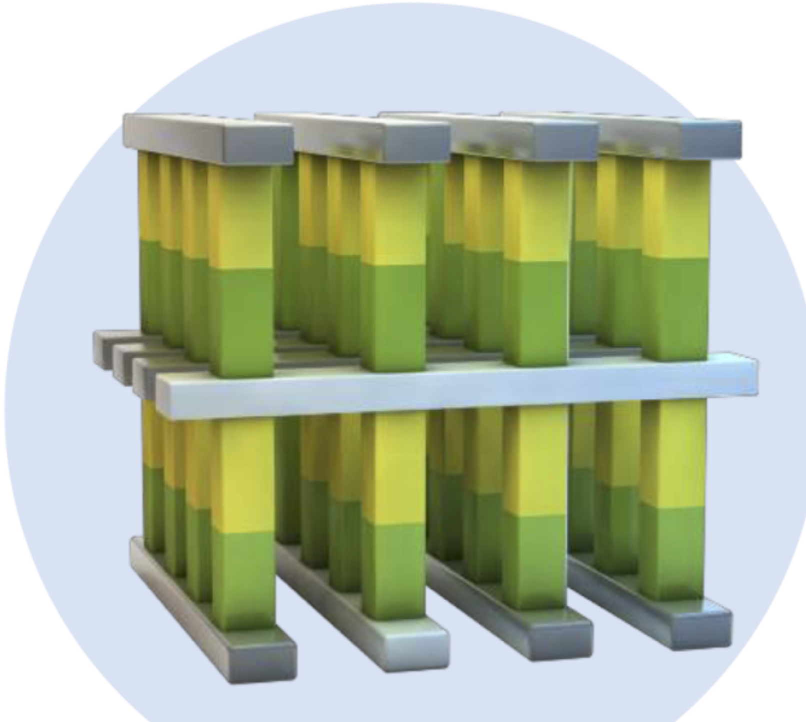
J. Thomas Pawlowski, Chief Technologist, Fellow
jpawlowski@micron.com

Introducing 3D XPoint™ Memory

1000X
FASTER
THAN NAND

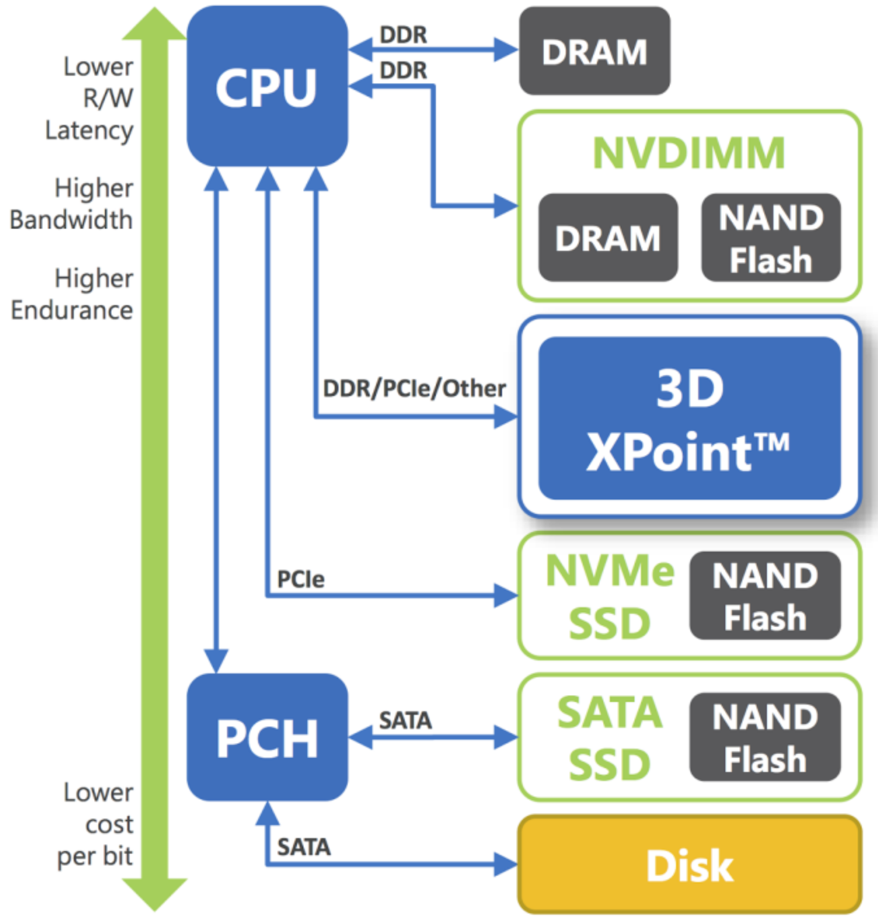
1000X
ENDURANCE
OF NAND

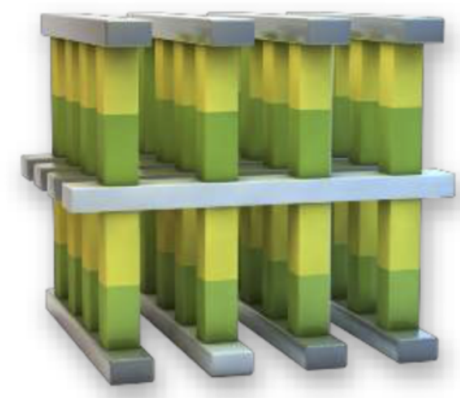
10X
DENSER
THAN CONVENTIONAL MEMORY



3D XPoint

Nonvolatile Memories in Server Architectures





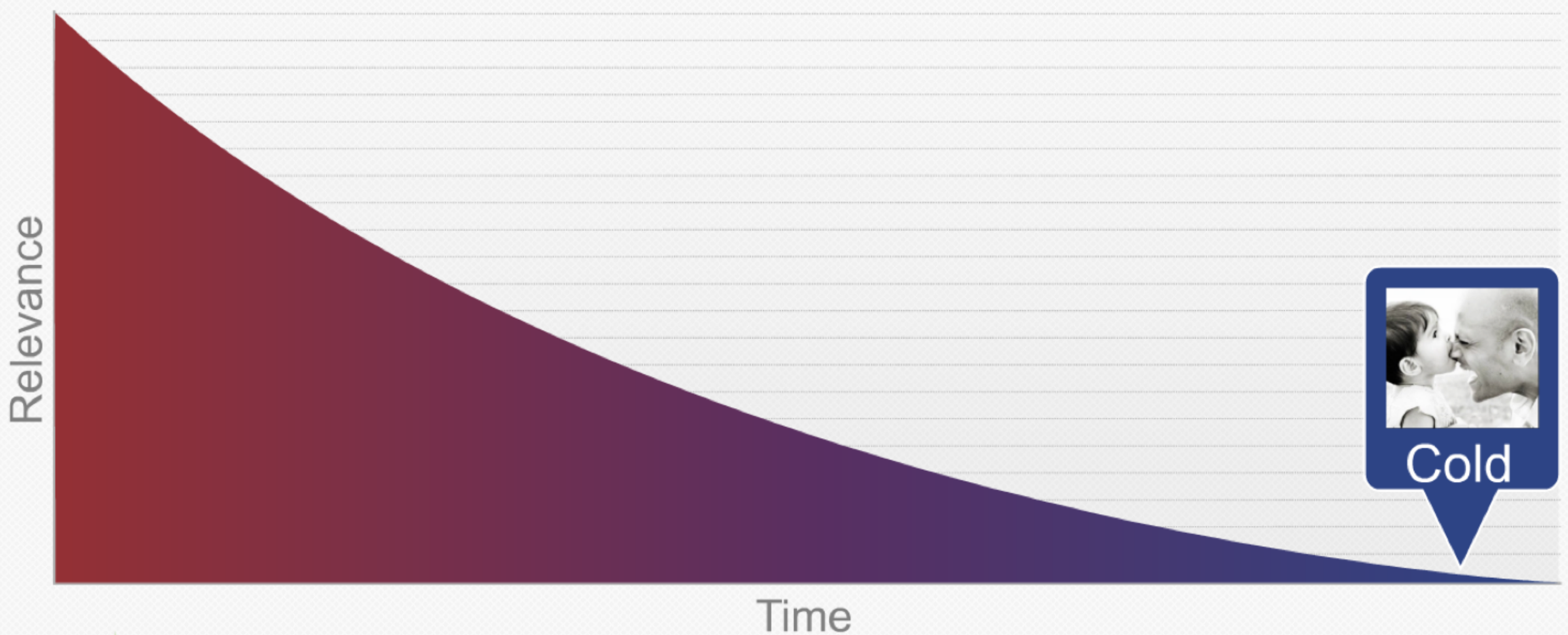
- 3D XPoint™ technology provides the benefit in the middle
- It is considerably faster than NAND Flash
- Performance can be realized on PCIe or DDR buses
- Lower cost per bit than DRAM while being considerably more dense

facebook

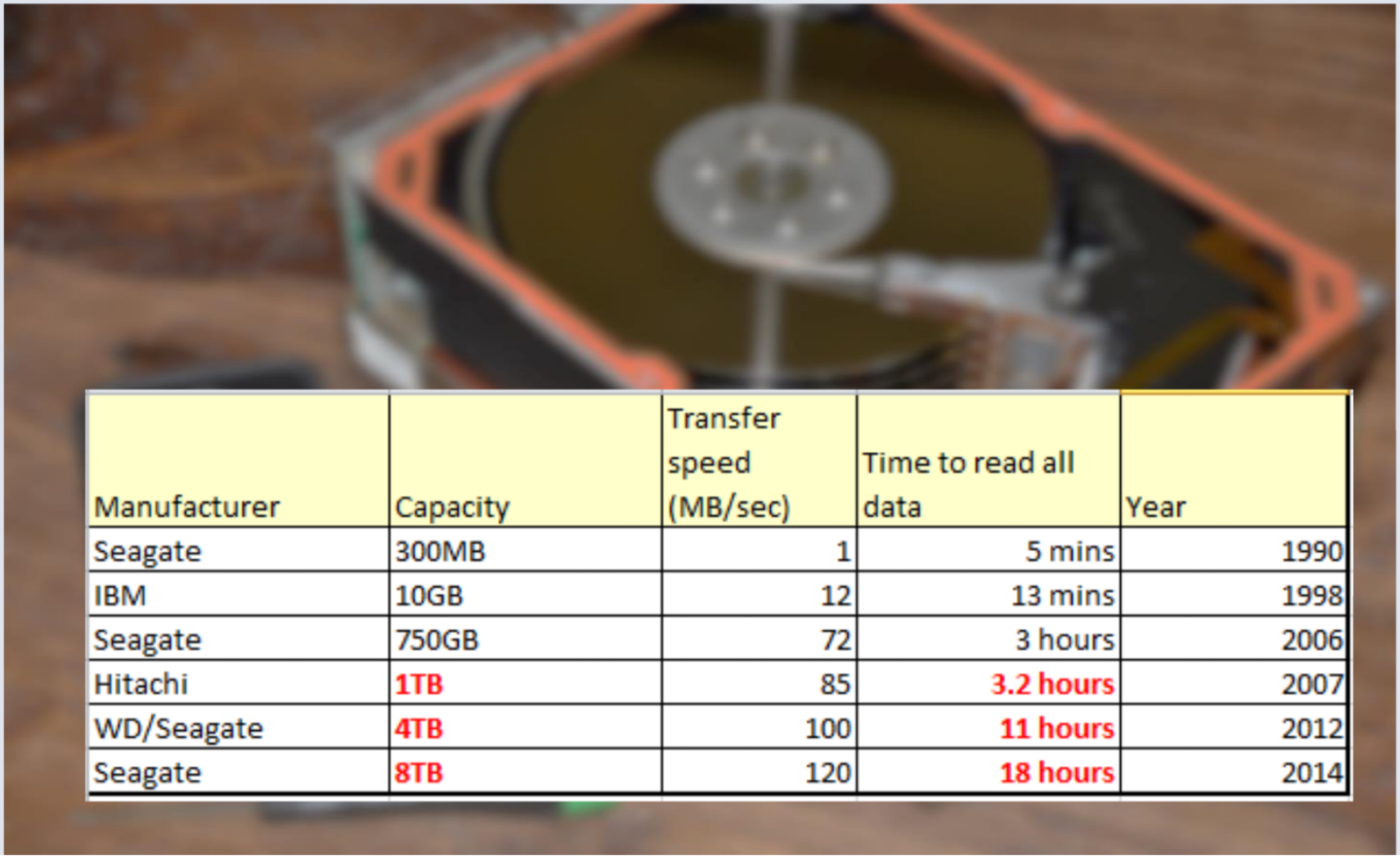
Freezing Exabytes of Data at Facebook's Cold Storage

Kestutis Patiejunas (kestutip@fb.com)

Photo lifecycle



History of Hard Drive data transfer rates



Manufacturer	Capacity	Transfer speed (MB/sec)	Time to read all data	Year
Seagate	300MB	1	5 mins	1990
IBM	10GB	12	13 mins	1998
Seagate	750GB	72	3 hours	2006
Hitachi	**1TB**	85	**3.2 hours**	2007
WD/Seagate	**4TB**	100	**11 hours**	2012
Seagate	**8TB**	120	**18 hours**	2014

Architecture from 36,000 feet

Rack

Server

Goals and non goals

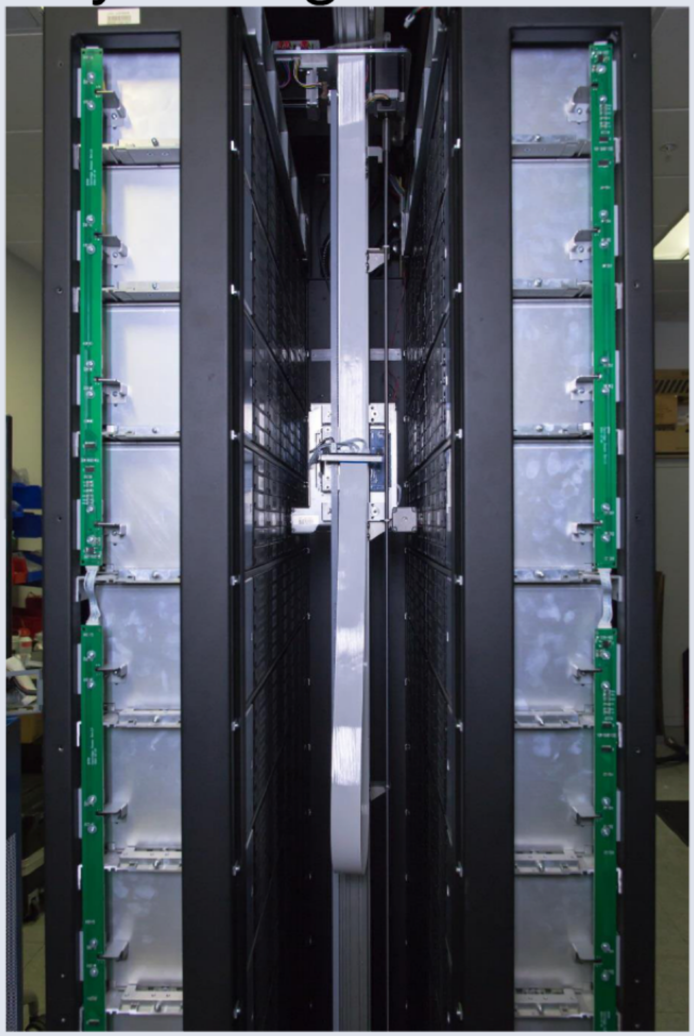
1. Durable
2. High efficiency
3. Reasonable availability
4. Scale
5. Support evolution
6. Gets better as it gets bigger

1. Have low latency for write/read operations
2. Have high availability
3. Be efficient for small objects
4. Be efficient for the objects with short lifetime

Gets better as it gets bigger

Number of racks	Capacity (PB)	Amount of data to read(write) in 1h at 50%	PB in 24 hours
30	52	0.1	1.2
90	156	0.2	3.7
200	346	0.3	8.2
500	865	0.9	20.6
1000	1730	1.7	41.2
2000	3460	3.4	82.4

Facebook Blu-Ray storage rack



Storage at CERN
(and in the LHC computing grid)
lessons learned / remaining challenges / opportunities ahead

Dirk Duellmann, CERN

Status Quo @ CERN: a few numbers

- 186k cores 16k servers
- 105k drives 244 PB raw disk space
- 755 TB RAM
- 104 tape drives 20k cartridges
  - 137 PB used / 47 PB free
- CERN computing centre is split over two sites: Geneva, Switzerland and Budapest, Hungary
- This represents only 20-30% of the total resources required for LHC analysis

Media hierarchy

- We still do tape. Why?
  - $/PB (incl. power)
  - separate physical copy with high “destruction” latency
- We stopped doing “automatic” HSM (Hierarchical Storage Management) for large experiment users
  - file based HSM interface did not allow to specify user priorities
- Disk content is stable and concurrently accessed
  - thousands of streams (jobs) at relatively low rate (cpu bound)
- Tape access for few production-users as explicit archive

YouTube: At a glance

10 years old

100+ hours / video uploaded every minute for the last 3 years ; 400+ 2016

Supporting playback on devices from ancient to modern

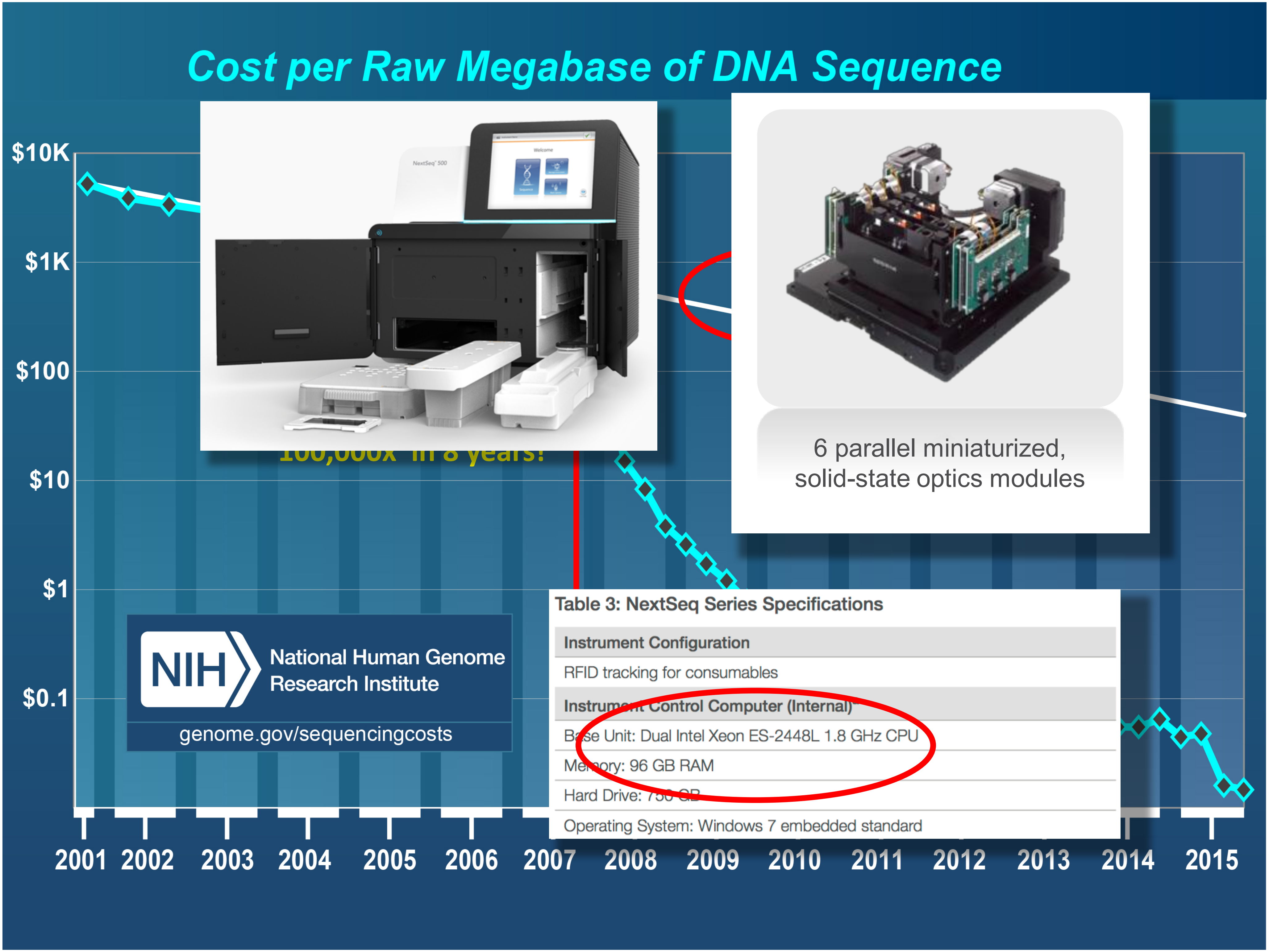
Corpus Effects

Lots of related content ; deduplication / cross-ref opportunities

Many domains: photos

Retrieval granularity the key issue

Silicon meets Biotech:


Cost per Raw Megabase of DNA Sequence
$10K
$1K
$100
$10
$1
$0.1
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
NextSeq 500
Welcome
6 parallel miniaturized, solid-state optics modules
NIH
National Human Genome Research Institute
genome.gov/sequencingcosts
Table 3: NextSeq Series Specifications
Instrument Configuration
RFID tracking for consumables
Instrument Control Computer (Internal)
Base Unit: Dual Intel Xeon ES-2448L 1.8 GHz CPU
Memory: 96 GB RAM
Hard Drive: 750 GB
Operating System: Windows 7 embedded standard

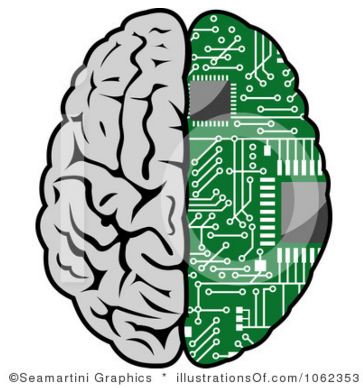
Computer Industry

Helped a **lot**!

Time for payback!

Biotech Industry

Biology
- Density
- Self-assembly
- Efficiency
- Sensitivity
  - Can sense things silicon can’t

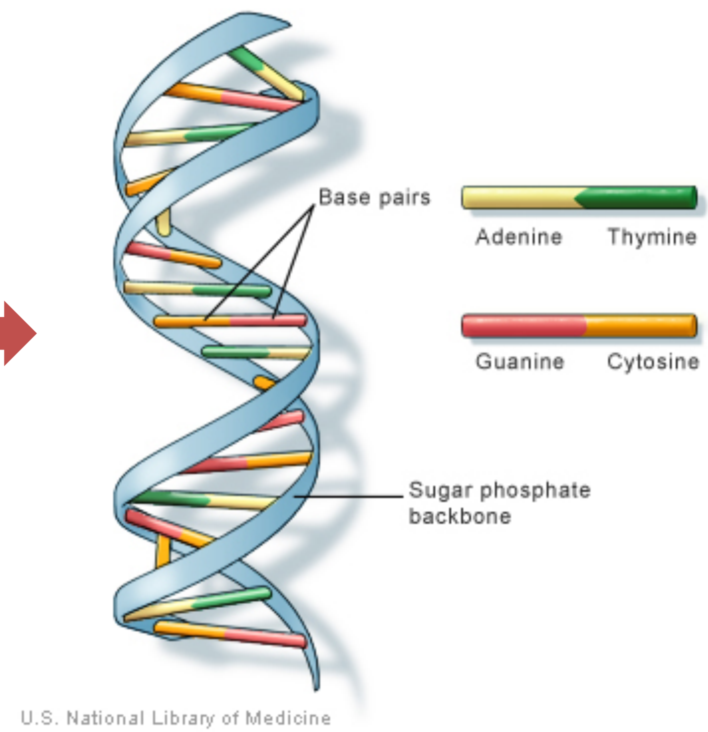


Silicon
- Speed
- Engineerability
- Integration with Infrastructure

Storage

...110101010100...

(images, video, genomic data, ...)

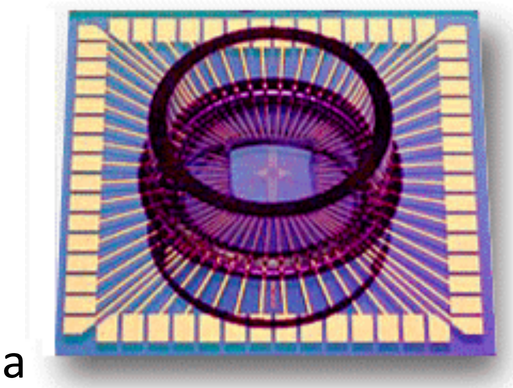


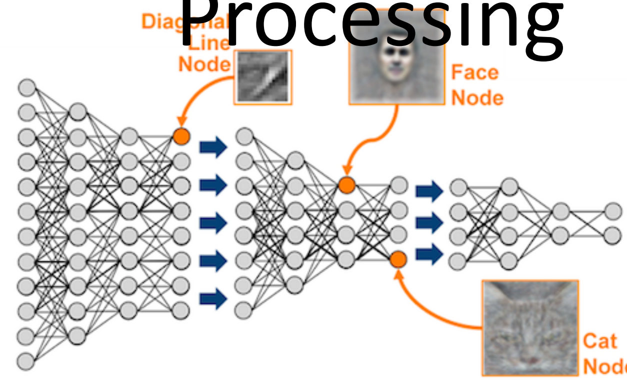
Lifetime (years)	3-5	5	10-30	100	5	5-10	**1000+**

And readers never become obsolete!

Session 7A on Wed @ 11:40am:
A DNA Archival Storage System

Processing





Proteins are ama

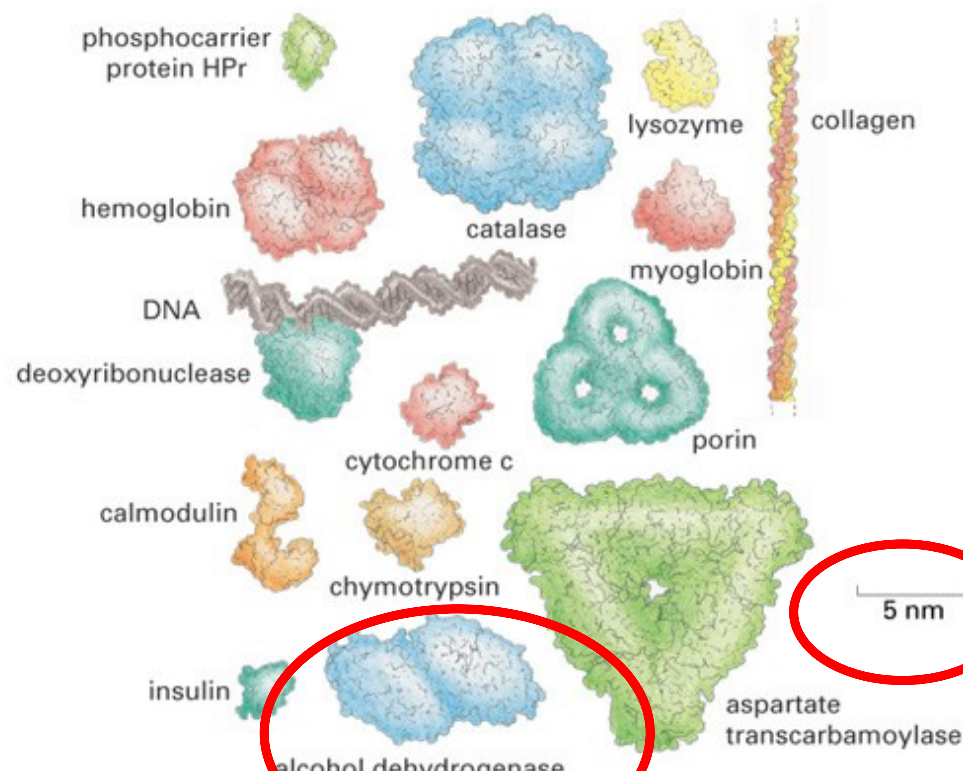
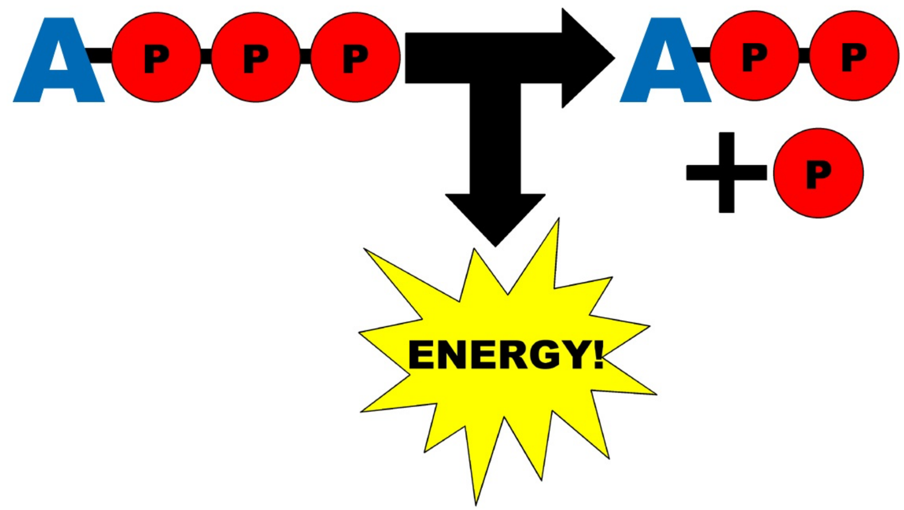


Figure 4-9 Essential Cell Biology, 2/e. (© 2004 Garland Science)



1 ATP molecule = ~ 10^{-21} J

A “proteinistor” would be 10,000x more energy efficient than a CMOS transistor

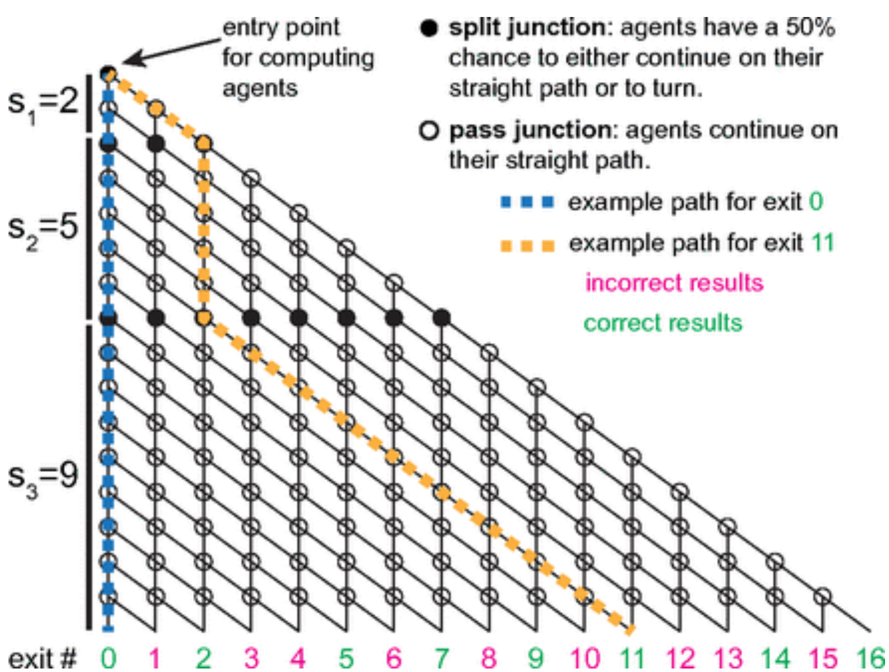
ATP powered “supercomputer”?

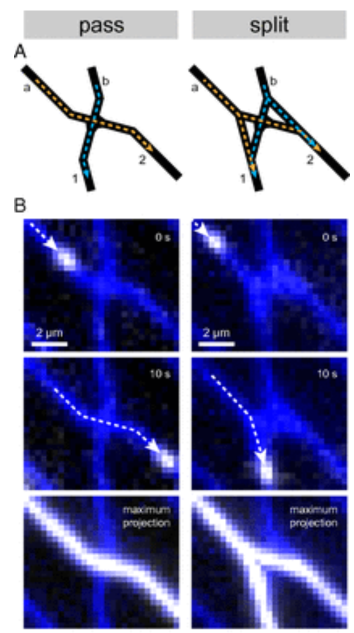
Parallel computation with molecular-motor-propelled agents in nanofabricated networks

Dan V. Nicolau Jr.[a,b,1], Mercy Lard[c,1], Till Korten[d,e,1], Falco C. M. J. M. van Delft[f,2], Malin Persson[g], Elina Bengtsson[g], Alf Månsson[g], Stefan Diez[d,e], Heiner Linke[c,3], and Dan V. Nicolau[h,i,3]

[a]Department of Integrative Biology, University of California, Berkeley, CA 94720-3140; [b]Molecular Sense, Ltd., Wallasey CH44 1AJ, United Kingdom; [c]NanoLund and Solid State Physics, Lund University, S-22100 Lund, Sweden; [d]Center for Molecular Bioengineering (B CUBE) and Center for Advancing Electronics Dresden (cfaed), Technische Universität Dresden, 01069 Dresden, Germany; [e]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany; [f]Philips Research (MiPlaza) and Philips Innovation Services, 5656 AE, Eindhoven, The Netherlands; [g]Department of Chemistry and Biomedical Sciences, Linnaeus University, SE-39182 Kalmar, Sweden; [h]Department of Electrical Engineering & Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom; and [i]Department of Bioengineering, McGill University, Montreal, QC, Canada H3A 0C3

Edited by Hillel Kugler, Microsoft Research, Cambridge, United Kingdom, and accepted by the Editorial Board January 18, 2016 (received for review June 5, 2015)





	a2+b1	a1+b2	n
Actin	97.9%	2.1%	3669
MT	99.7%	0.3%	1801

a1	a2	b1	b2	n
50%	50%	42%	58%	1492
49%	51%	51%	49%	705

Or an ATP battery?

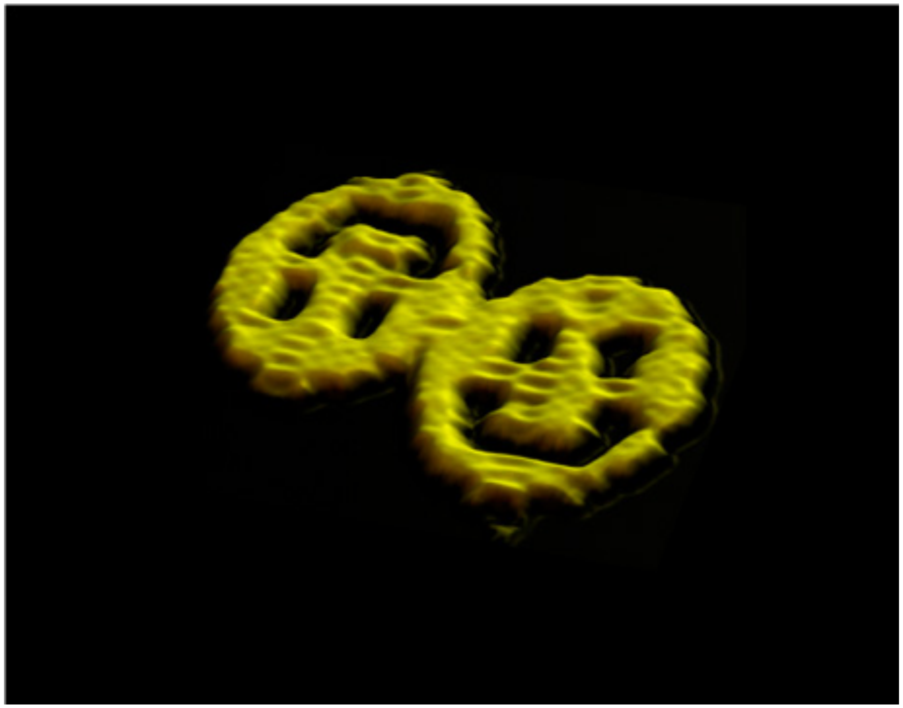
NATURE COMMUNICATIONS | ARTICLE

A high-energy-density sugar biobattery based on a synthetic enzymatic pathway

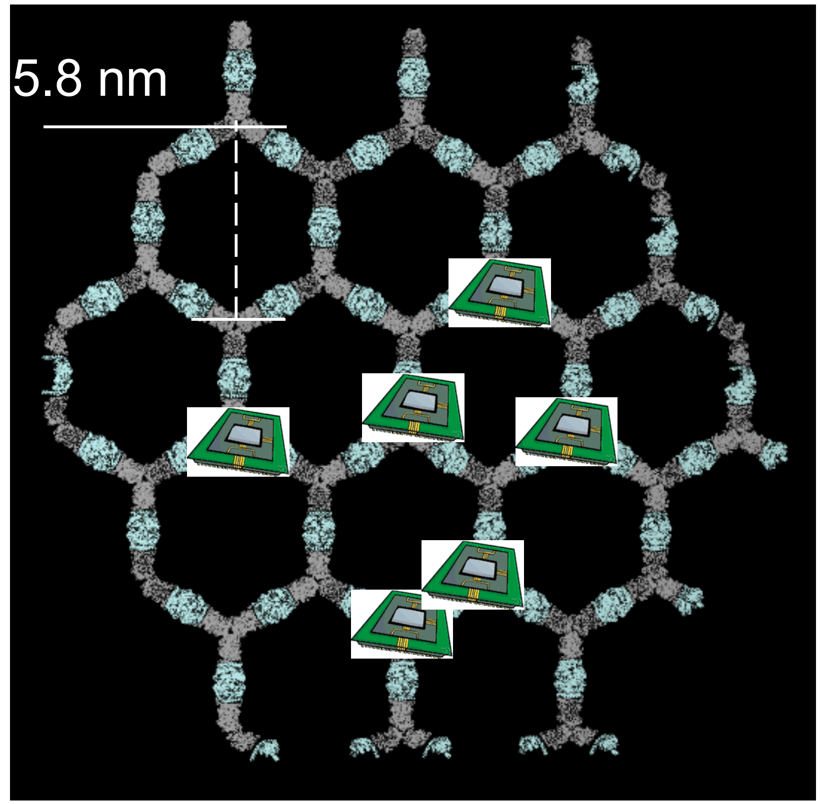
Zhiguang Zhu, Tsz Kin Tam, Fangfang Sun, Chun You & Y. -H. Percival Zhang

10x energy density of lithium

Molecular-Level Self-Assembly *and Reconfiguration*

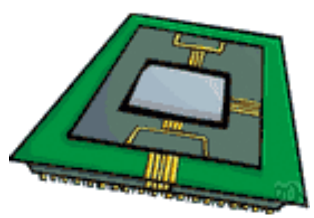


DNA origami



Proteins

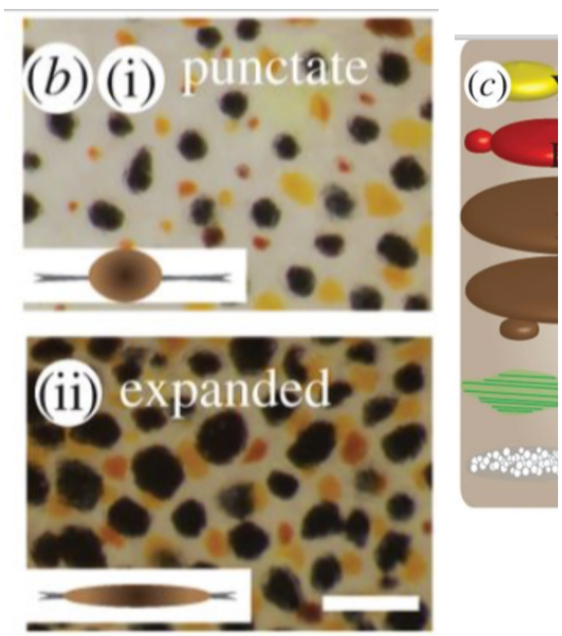
[Caltech, Duke]



Single-Molecule Sensing

Gu et al. Single molecule sensing by nanopores and nanopore devices. 2009, Analyst.

What about output?

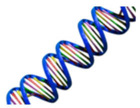


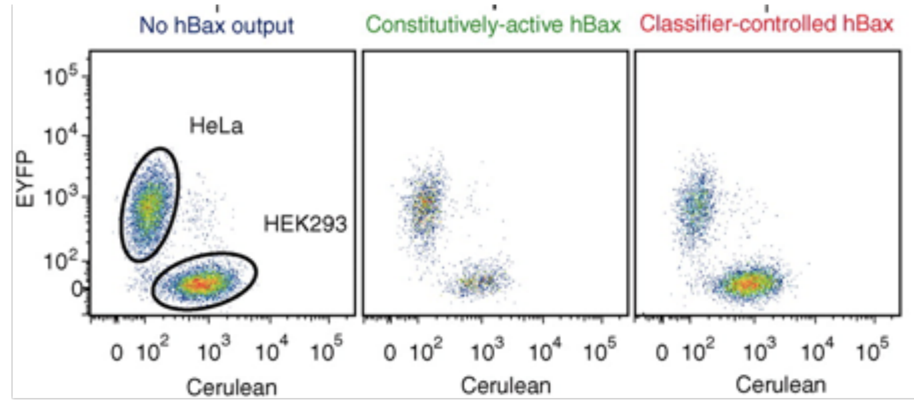
ıigment
ell
MC
eus
rk

CuttlePhone

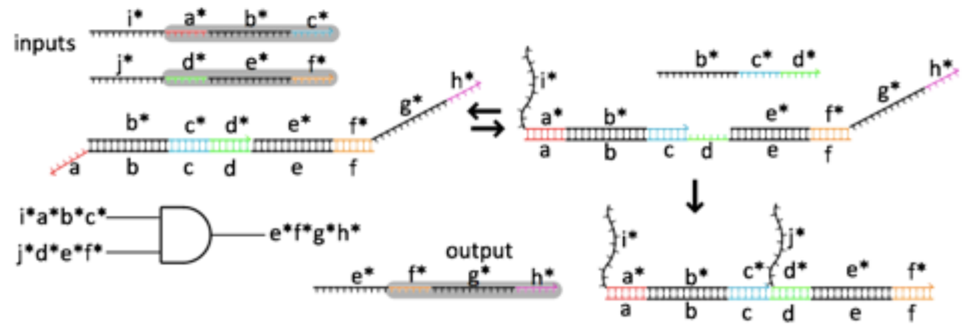
Silicon+Biotech

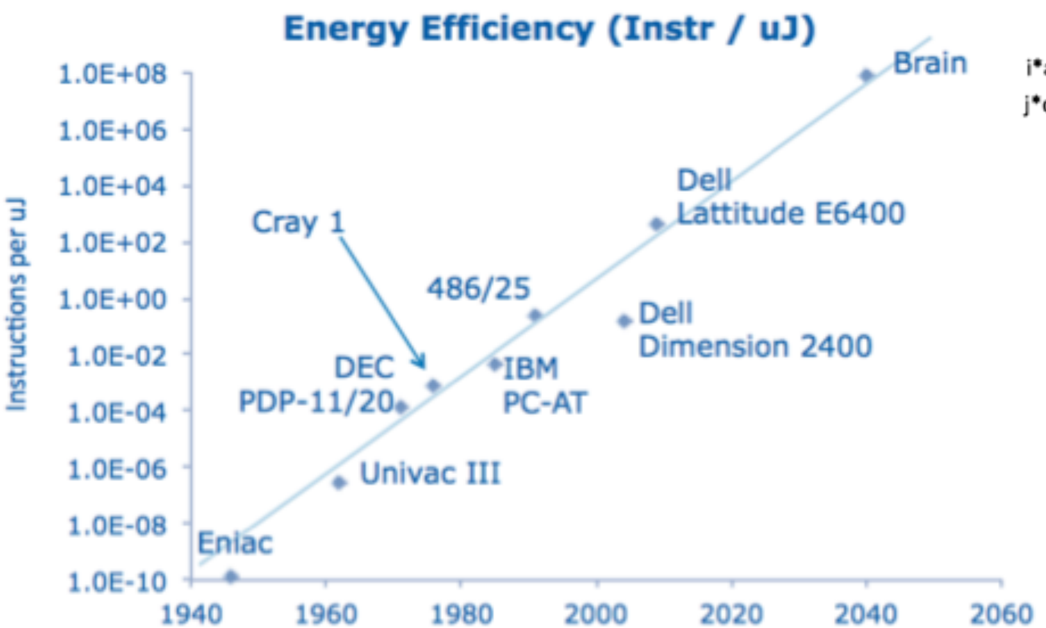
gattaca

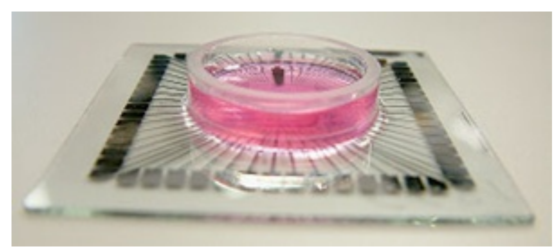
01011011101 —DNA synthesis→  —DNA sequencing→ 01011011101



Hyper-Dense — 1 ZB/cm^3 (~1E8 denser than Flash)
Hyper-Durable — We find readable ~100k-year-old DNA
Eternally relevant — As long as there is DNA-based intelligent life, there will be reasons to read/write DNA

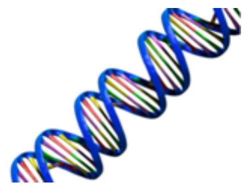






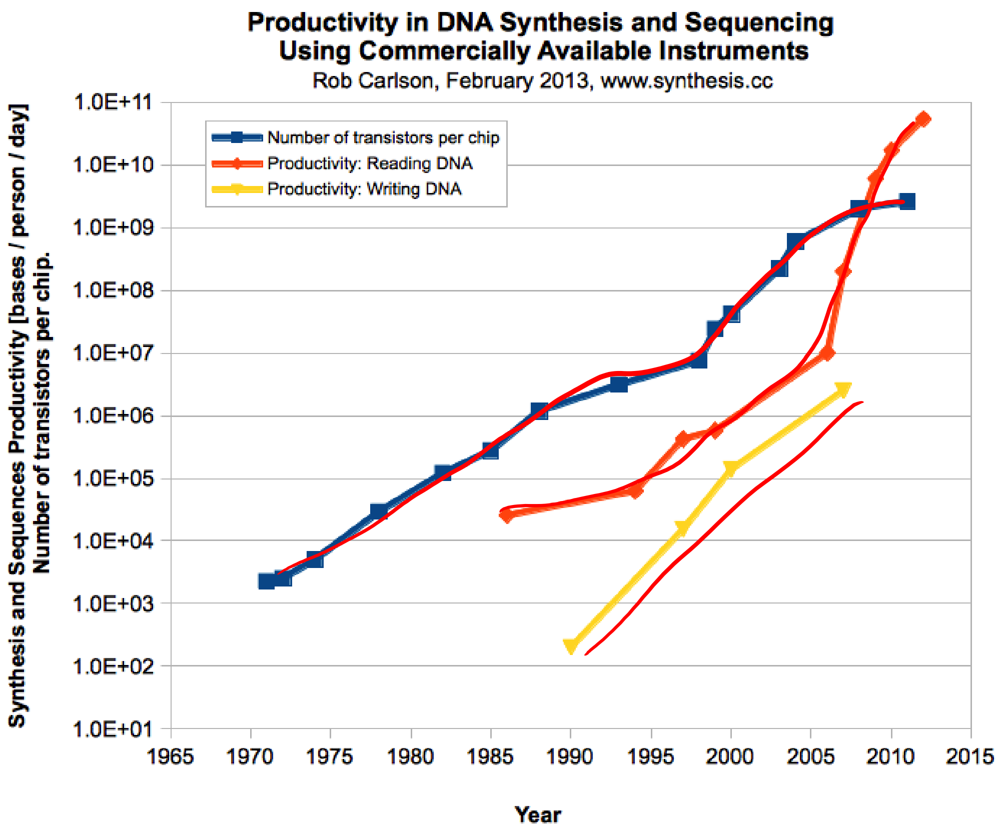
DNA as data storage media

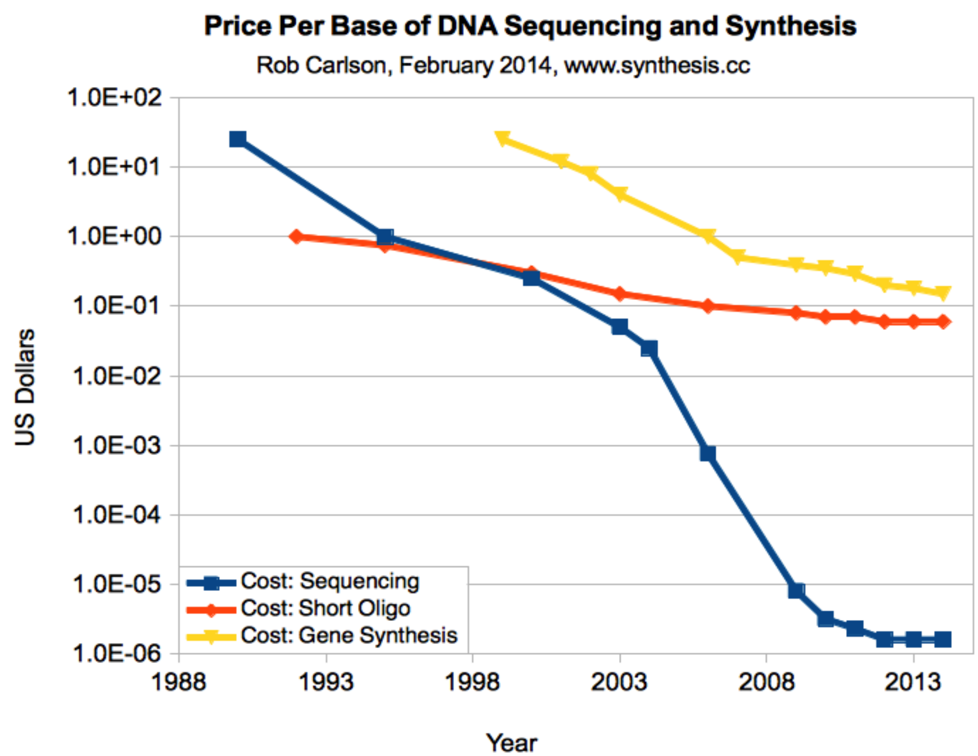
gattaca

01011011101 $\xrightarrow{\text{DNA synthesis}}$  $\xrightarrow{\text{DNA sequencing}}$ 01011011101

Hyper-Dense — 1 exabyte/cm^3 (~1E8 denser than Flash)
Hyper-Durable — We find readable 1M-year-old DNA
Eternally relevant — As long as there is DNA-based intelligent life, there will be reasons to read/write DNA

Cost and speed!?

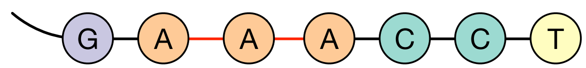




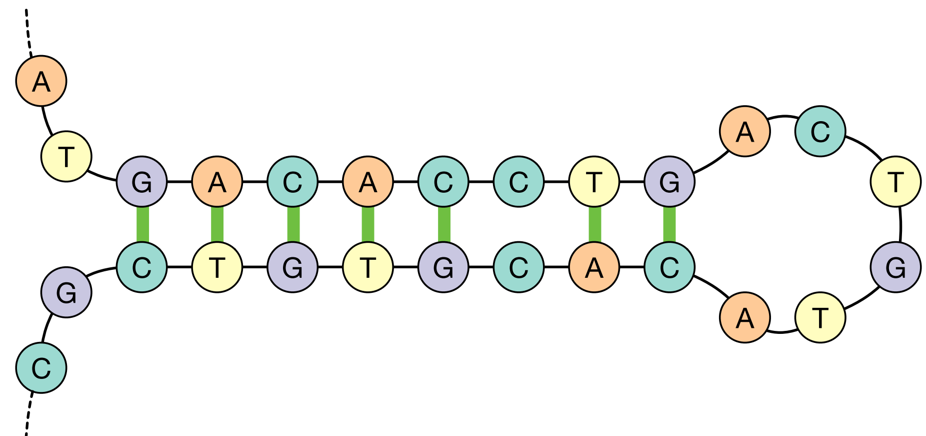
From bits to letters

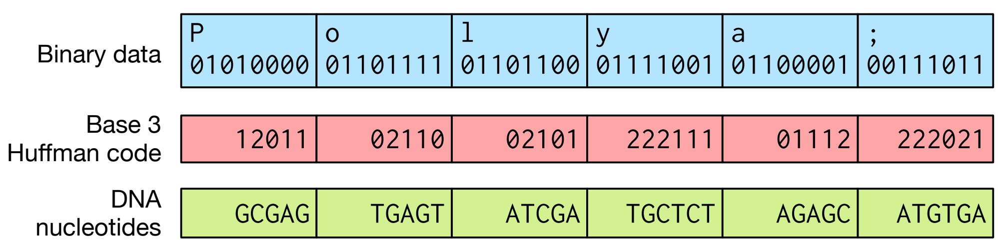
(00, 01, 10, 11) → A, C, G, T

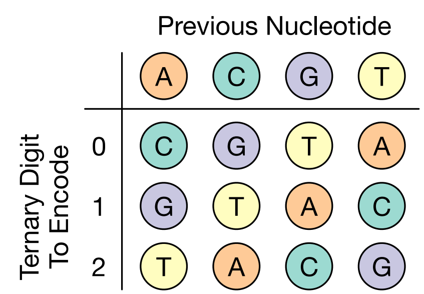
Repeated letters are bad…



Secondary structures are bad…



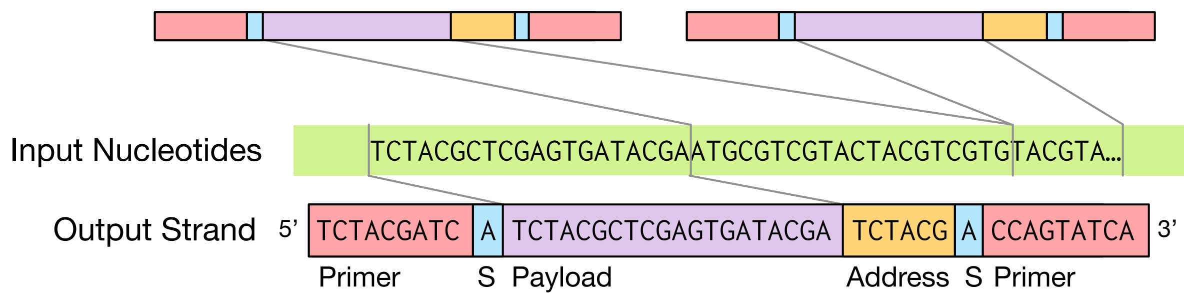




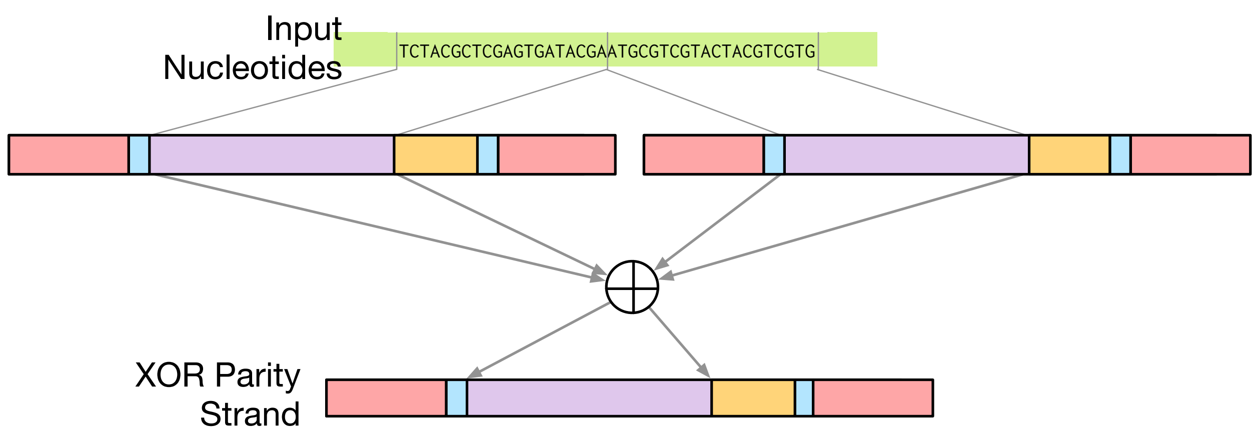
No repeated letters, low chance of secondary structures

Preventing a big mess…

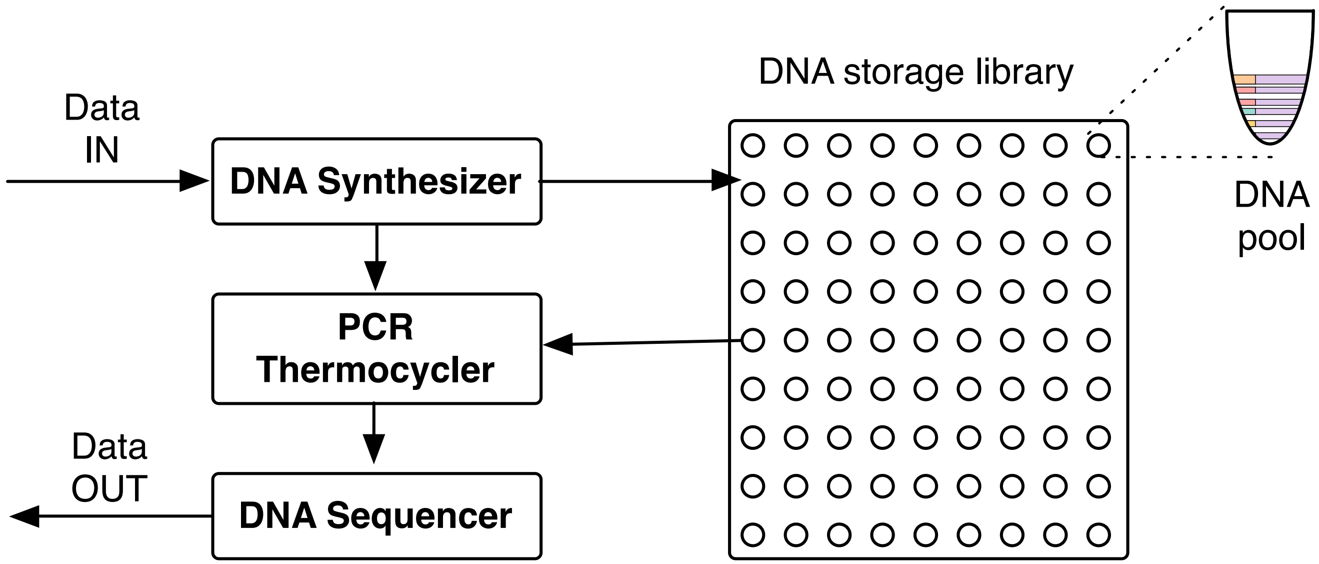
DNA is “uni-dimensional”, so need to embed address as data…

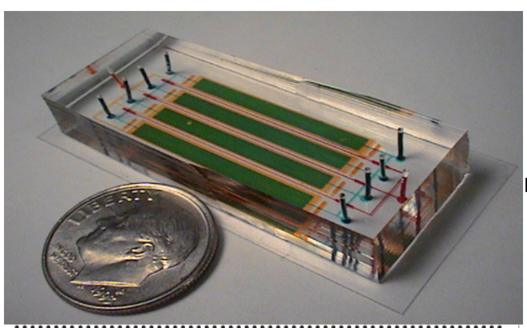


Synthesis is not perfect…



An integrated system





Accurate multiplex gene synthesis from programmable DNA microchips

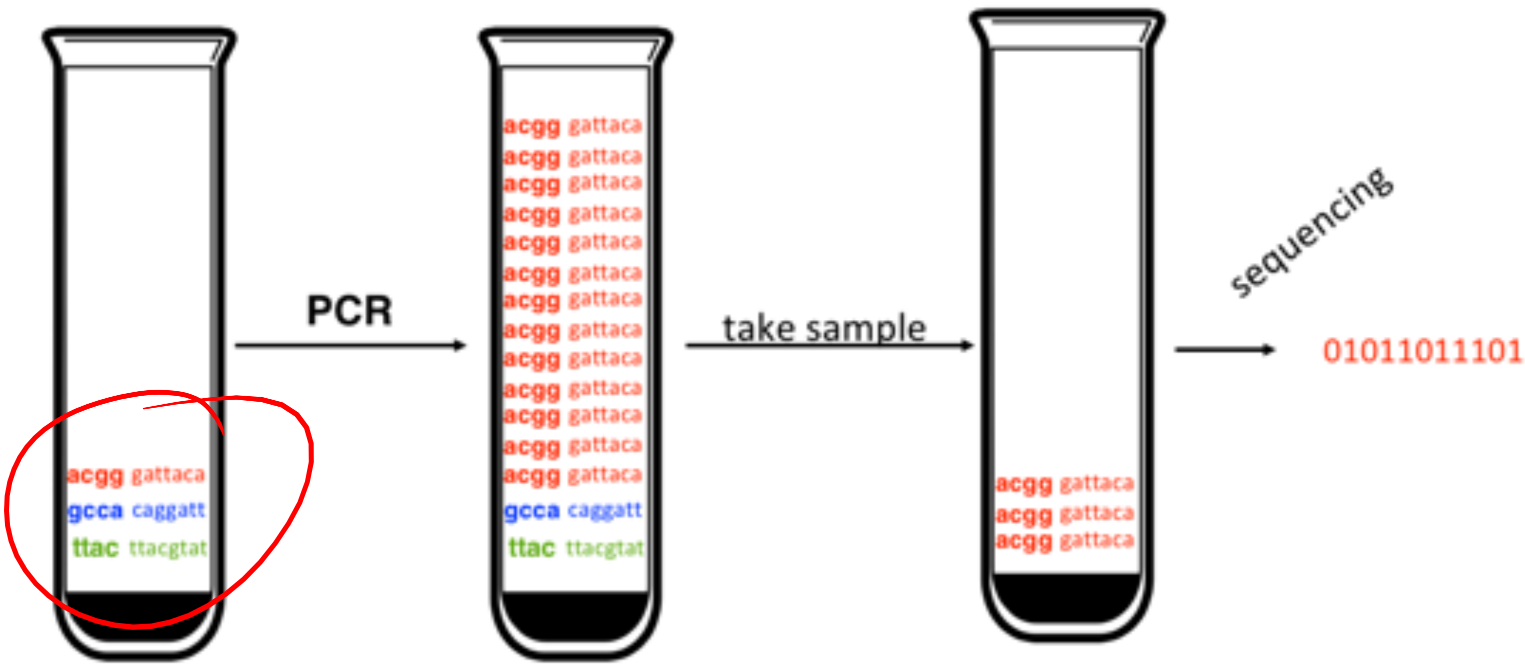
Jingdong Tian[1], Hui Gong[1], Nijing Sheng[2], Xiaochuan Zhou[3], Erdogan Gulari[4], Xiaolian Gao[2] & George Church[1]

+

=

Random access

address	data	PCR primer	data
1110	01011011101	**acgg**	gattaca
1010	11010010000	**gcca**	caggatt
0110	11111000011	**ttac**	ttacgtat



What I hope you got a sense of…

- Power as a first class design constraint
- How "modern" microprocessors work
- ILP/TLP/DLP
- Cache coherence and memory consistency models
  - At least the difference between them ☺
- Silicon scaling future not so good
  - Trade-offs between generality and performance inevitable

Today’s Dominant Target Systems

- Mobile (smartphone/tablet)
  - >1 billion sold/year
  - Market dominated by ARM-ISA-compatible general-purpose processor in system-on-a-chip (SoC)
  - Plus sea of custom accelerators (radio, image, video, graphics, audio, motion, location, security, etc.)
- Warehouse-Scale Computers (WSCs)
  - 1,000,000’s cores per warehouse
  - Market dominated by x86-compatible server chips
  - Dedicated apps, plus cloud hosting of virtual machines
  - Starting to see some GPU usage, but mostly general-purpose CPU code
- Embedded computing
  - Wired/wireless network infrastructure, printers
  - Consumer TV/Music/Games/Automotive/Camera/MP3

Thank you!