Computing Challenges in the SKA Era

Bruce Elmegreen

IBM T.J. Watson Research Center

Yorktown Heights, NY 10598

bge@us.ibm.com

Rutgers University, March 16-18, 2015

Outline: the future of...

- Processors ... Dennard scaling has ended, will Moore's law end too? or is there still "plenty of room at the bottom?" (Feynman 1959)
- Memory ... DRAM, SRAM \rightarrow non-volatile storage
- Storage ... \rightarrow Thermal assist magnetic writing with x100 density
- Data and Data Motion: how big is the SKA on a world scale?

Processors



Gordon Moore 1965: "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year (see graph on next page). Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years."

Exhibit 1: Number of Transistors per Device



Source: Company data, Credit Suisse estimates

http://qz.com/218514/chip-makers-are-betting-that-moores-law-wont-matter-in-the-internet-of-things/



How Lara Croft's changing face illustrates Moore's law

Updated by Timothy B. Lee on February 1, 2015, 11:00 a.m. ET 🛛 tim@vox.com



1996



PERFORMANCE DEVELOPMENT



PROJECTED

Top 500 supercomputers: Sum, #1, #500

http://www.wired.com/2014/06/supercomputer_race/

RANK	SITE	SYSTEM	CORES	RMAX (TFLOP/S)	RPEAK (TFLOP/S)	POWER (KW)
1	National Super ComputerTianhe-2 (MilkyWay-2) - TH-Center in GuangzhouIVB-FEP Cluster, Intel Xeon E5-China2692 12C 2.200GHz, THExpress-2, Intel Xeon Phi 31S1PNUDT		3,120,000 Both	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890

(November 2014)





CPU: Intel E5-2600 Ivy Bridge (10 core version shown), 22nm

Accelerator: Intel Xeon Phi 60 cores, 22 nm







PERFORMANCE DEVELOPMENT



What's next?

PROJECTED

http://www.wired.com/2014/06/supercomputer_race/



AND LECE NUDIA Volta, IBM POWER9 Land Contract Market Stores And Stores And





Lawrence Livermore National Laboratory

SUMMIT

SIERRA

150-300 PFLOPS Peak Performance IBM POWER9 CPU + NVIDIA Volta GPU NVLink High Speed Interconnect 40 TFLOPS per Node, >3,400 Nodes

2017

Major Step Forward on the Path to Exascale

PERFORMANCE DEVELOPMENT



~200 Pf in 2017: Right on track

PROJECTED

http://www.wired.com/2014/06/supercomputer_race/



All per major cycle of processing, assuming 10 cycles (Bojan Nikolic 2014, SDP team)

http://www.wired.com/2014/06/supercomputer_race/

(* Date from John Womersley 2015, STFC)



All per major cycle of processing, assuming 10 cycles (Bojan Nikolic 2014, SDP team)

http://www.wired.com/2014/06/supercomputer_race/

(* Date from John Womersley 2015)



SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

Choose your local minisite

The World's Largest Radio Telescope Takes A Major Step Towards Construction

SKA Organisation HQ – Monday 9 March - At their meeting last week at the SKA Organisation Headquarters near Manchester, UK, the SKA Board of Directors unanimously agreed to move the world's largest radio telescope forward to its final pre-construction phase. The design of the €650M first phase of the SKA (SKA1) is now defined, consisting of two complementary world-class instruments – one in Australia and one in South Africa – both expecting to deliver exciting and transformational science.

The SKA instruments will be located in two countries – South Africa and Australia. In the first phase of the project, South Africa will host about 200 parabolic antennas or dishes – similar to, but much larger than a standard domestic satellite dish – and Australia more than 100,000 'dipole' antennas, which resemble domestic TV aerials.

Hardware Alternatives (Broekema et al. 2015, SDP)

The fiducial design considered by the SDP (2015) is a CPU+GPU combination.

There are other Exascale projects too:

- China: Milky Way 2 upgrade to 0.1 Exaflop in 2015, exascale by 2020
- Japan: \$1.1B to develop 0.2-0.6 Exaflop computer by 2020: "Flagship 2020"
- Europe: 8 separate projects toward the Exascale

For the SKA, it is <u>not just high speed but also low power</u>: -The goal is 5.5 MWatts for a ~300 Pflop computer in SKA1

e.g., Milky Way 2 is 34 Pflops at 24 MWatts – 40x higher per flop!

Worth considering completely different architectures, e.g., FPGA computing, microchips



An example considered by the DOME Project



µServer:

The integration of an entire server node motherboard^{*} into a single microchip except DRAM, Nor-boot flash and power conversion logic.

This does NOT imply low performance!



Ronald P. Luijten - SC14 16-20 Nov 2014

Will Moore's Law Continue?

Appears in the Proceedings of the 38th International Symposium on Computer Architecture (ISCA '11)

Dark Silicon and the End of Multicore Scaling

Hadi Esmaeilzadeh[†] Emily Blem[‡] Renée St. Amant[§] Karthikeyan Sankaralingam[‡] Doug Burger[°] [†]University of Washington [‡]University of Wisconsin-Madison [§]The University of Texas at Austin [°]Microsoft Research hadianeh@cs.washington.edu blem@cs.wisc.edu stamant@cs.utexas.edu karu@cs.wisc.edu dburger@microsoft.com

12 sample simulations what fraction of nodes are not useful for speedup, or must be turned off to limit power density on chip?

Model CPUs + GPUs







Insights for Electronics Manufacturing

HOME	SEMICONDUCTORS	PACKAGING	MEMS	LEDS	DISPLAYS	DESIGN	MAGAZINES	SEM
🖬 Lik	e 145 😏 Tweet 5!	n Share < 178	8+1 1	4				
Мо	ore's Law h	as stop	ped a	at 28	nm			
By Zvi	Or-Bach, President &	CEO of Monolit	hIC (March 2	014)			
While	many have recently pro	edicted the imm	inent der	nise of M	oore's Law, w	ve need to i	recognize that t	his ac-
tually I single	has happened at 28nm device but not at lower	. From this poin cost. And, for r	it on we v nost appl	vill still be ications,	e able to doub the cost will a	ble the amo actually go	unt of transistor up.	s in a

http://electroiq.com/blog/2014/03/moores-law-has-stopped-at-28nm/

Lithography Dominates the Cost Impact of Scaling



http://www.semiconwest.org/sites/semiconwest.org/files/docs/SW2013_Subi%20Kengeri_GLOBAFOUNDRIES.pdf

Technology scale on chip versus λ of light source



Technology Node (nm)



IBM Power 8: 22 nm technology is just 40 atoms wide in a Silicon crystal

PCI Express



3x12GB/s Processor Links

3x38GB/s Processor Links



Pricing: X'over on Transistor Cost



http://www.extremetech.com/computing/123529-nvidia-deeply-unhappy-with-tsmc-claims-22nm-essentially-worthless

Investment Needed For ONE LEADING EDGE FAB

300mm

>\$5B









Feb 10 2015



COMPUTING > AMD, NVIDIA BOTH SKIPPING 20NM GPUS AS TSMC PLANS \$16 BILLION FAB INVESTMENT, REPORT SAYS

AMD, Nvidia both skipping 20nm GPUs as TSMC plans \$16 billion fab investment, report says

By Joel Hruska on February 10, 2015 at 4:26 pm 39 Comments

http://www.extremetech.com/computing/199101-amd-nvidia-both-skipping-20nm-gpus-as-tsmc-plans-massive-16b-fab-investment-report-says



Intel: Moore's Law will continue through 7nm chips

Mark Hachman @markhachman

Feb 22, 2015 12:00 PM 🛛 🖾 🔒

(EP1) Moore's Law Challenges Below 10nm: Technology, Design and Economic Implications

"Intel believes that the current pace of semiconductor technology can continue beyond 10nm technology (expected in 2016 or so), and that 7nm manufacturing (expected in 2018) can be done without moving to expensive, esoteric manufacturing methods like ultraviolet lasers"*





(intel)

http://www.pcworld.com/article/2887275/intel-moores-law-will-continue-through-7nm-chips.html

Following Moore's Law is Expensive: Where does the money come from?

Table 1. Worldwide IT Spending Forecast (Billions of U.S. Dollars)

	2013 Spending	2013 Growth (%)	2014 Spending	2014 Growth (%)	2015 Spending	2015 Growth (%)
Devices	677	1.1	685	1.2	725	5.8
Data Center Systems	140	-0.1	140	0.4	144	2.9
Enterprise Software	300	5.1	321	6.9	344	7.3
IT Services	932	0.0	967	3.8	1,007	4.1
Telecom Services	1,624	-1.2	1,635	0.7	1,668	2.0
Overall IT	3,673	0.0	3,749	2.1	3,888	3.7

Source: Gartner (June 2014)

Tech Companies = Top 1 or 2 Sector by Market Cap in S&P500 for Nearly 2 Decades

20 Years Ago: Dec 1994 - S&P500 = \$3.2T			Peak of NASDAQ: Mar 2000 – S&P500 = \$11.7T			Today: May 2014 – S&P500 = \$17.4T			
Sector	Weight	LargestCompanies	Sector	Weight	Largest Companies	Sector	Weight	Largest Companies	
CONS. STAPLES	14%	COCA-COLA ALTRIA	TECHNOLOGY	35%	MICROSOFT CISCO	TECHNOLOGY	19%	APPLE GOOGLE	
CONS. DISC.	13%	MOTORS LIQUIDATION FORD	FINANCIALS	13%	CITIGROUP AIG	FINANCIALS	16%	WELLS FARGO JPMORGAN CHASE	
INDUSTRIALS	13%	GENERAL ELECTRIC 3M	CONS. DISC.	10%	TIME WARNER HOME DEPOT	HEALTHCARE	13%	JOHNSON & JOHNSON PFIZER	
FINANCIALS	11%	AIG FANNIE MAE	HEALTHCARE	10%	MERCK PFIZER	CONS. DISC.	12%	AMAZON.COM WALT DISNEY	
ECHNOLOGY	11%	IBM MICROSOFT	INDUSTRIALS	8%	GENERAL ELECTRIC TYCO	INDUSTRIALS	11%	GENERAL ELECTRIC UNITED TECHNOLOGIES	
IEALTHCARE	10%	MERCK JOHNSON & JOHNSON	TELECOM	7%	SOUTHWESTERN BELL AT&T	CONS. STAPLES	11%	WAL-MART PROCTOR & GAMBLE	
ENERGY	9%	EXXON MOBIL	CONS. STAPLES	7%	WAL-MART COCA-COLA	ENERGY	10%	EXXON MOBIL CHEVRON	
TELECOM	8%	SOUTHWESTERN BELL GTE	ENERGY	5%	EXXON MOBIL CHEVRON	MATERIALS	3%	DUPONT MONSANTO	
MATERIALS	7%	DUPONT DOW CHEMICAL	MATERIALS	2%	DUPONT ALCOA	UTILITIES	3%	DUKE ENERGY NEXTERA ENERGY	
UTILMES	4%	SOUTHERN COMPANY DUKE ENERGY	UTILITIES	2%	DUKE ENERGY AES	TELECOM	2%	VERIZON AT&T	





Memory

United States Patent Office

3,387,286

Patented June 4, 1968

3,387,286 FIELD-EFFECT TRANSISTOR MEMORY Robert H. Dennard, Croton-on-Hudson, N.Y., assignor to International Business Machines Corporation, Armonk, N.Y., a corporation of New York Filed July 14, 1967, Ser. No. 653,415 21 Claims. (Cl. 340-173)

ABSTRACT OF THE DISCLOSURE

The memory is formed of an array of memory cells controlled for reading and writing by word and bit lines which are connected to the cells. Each cell is formed, in one embodiment, using a single field-effect transistor and a single capacitor. The gate electrode of the transistor is ¹⁵

2

tinent in disclosing various concepts and structures which have been developed in the application of field-effect transistors to different types of memory applications, the primary thrust up to this time in conventional read-write random access memories has been to connect a plurality

of field-effect transisto uration. Memories of t active devices in each quires a relatively large strate. This type of de cells which can be bui necessitates the use o the expense of speed c

10

Word Line eij Gind

onfigber of 11 ret subemory urther nes at

Summary of the invention







Home > Computer Hardware

Apple will consume 25% of all DRAM in the world next year

Samsung 6 Gb DRAM



DRAM production: 1.05 Million chips/month (Electronic News Sept 2014)

http://www.computerworld.com/article/2687940/apple-willconsume-25-of-all-dram-in-the-world-next-year.html 1.05M x 6Gb x 12 months = 7.6E16 bits/yr of DRAM Rice production =4E16 grains/yr <u>One problem</u>: charge in the capacitor continuously leaks

Solution: read it and re-write it every 64 ms

This makes DRAM power hungry

 \rightarrow BAD for Exascale computers



Transistor

 a) Stack capacitors above the transistor



Power Use in Peta- to Exa- Scale Systems



(ref: IBM BlueGene team 2012)

Solutions to Power Problem: "non-volatile memory"

Normal Field Effect Transistor

 Voltage on "Control Gate" pulls up electrons from the body "P" and this allows a current to flow from the "source" to the "bit line" (="drain")



Solutions to Power Problem: "non-volatile memory"

FLASH (Toshiba 1984):

- Voltage on "Control Gate" pulls up electrons from the body "P" and this allows a current to flow from the "source" to the "bit line" (="drain")
- 2. When "Float Gate" is charged, fewer electrons move up and voltage to drive a current is higher.
- Sense state by applying intermediate voltage: if current flows, the "Float Gate" is uncharged, if no current flows, the FG is not charged: bits are 1 or 0 respectively



The Bleak Future of NAND Flash Memory*

(FAST'12 Proceedings of the 10th USENIX conference on File and Storage Technologies, 2012)

Laura M. Grupp[†], John D. Davis[‡], Steven Swanson[†]

[†]Department of Computer Science and Engineering, University of California, San Diego

[‡]Microsoft Research, Mountain View

"future gains in density will come at significant drops in performance and reliability"



Non-volatile Memory in the SKA Era:

Phase Change Memory: chalcogenide glasses (GbSbTe) Crystal phase conducts electricity, amorphous phase does not. Phase is changed by temperature cycling.

Limited production by Numonyx (2008), Samsung (2009), Micron (2012 for Nokia phones, but withdrawn in 2014 to pursue 3D FLASH)



Magneto-resistive Random Access Memory (MRAM)

Under development by IBM, Hynix, Samsung and Toshiba Faster than FLASH, longer life, more reliable



Bit Line

Magnetic Free Layer



Crossbar, Inc. working on resistive RAM (RRAM) to replace FLASH

https://www.youtube.com/watch?feature=player_detailpage&v=EWbikdFFs6A http://www.digitaltrends.com/computing/resistive-ram-how-it-could-change-storage-forever/

Decades away.... IBM Almaden Lab, 0 N Science, 2012

12 Iron atoms ferromagnetically aligned and stable

http://www.extremetech.com/computing/113237-ibm-stores-binary-data-on-12-atoms



Storage



Development targets for HDDs

Reduction of power consumption is a major social issue!

To reduce power consumption and increase information handling capacity...





Perpendicular vs. HAMR Recording

Laser

Heated

spot

Heat above the Curie Temp. reduces coercivity* and allows writing in a smaller area with a given magnetic field

*=resistence to demagnetization

GMR Element N N s s Soft Underlayer Soft Underlayer

Shield

Data Volume

Photos Alone = 1.8B+ Uploaded & Shared Per Day... Growth Remains Robust as New Real-Time Platforms Emerge

Daily Number of Photos Uploaded & Shared on Select Platforms, 2005 – 2014YTD





'Digital Universe' Information Growth = Robust... +50%, 2013

2/3rd's of Digital Universe Content = Consumed / Created by Consumers ...Video Watching, Social Media Usage, Image Sharing...



SKA Archives, to be duplicated at SA and AU (Bolton 2015, SDP)

Long term (e.g., tape): 500 PBy initial with 130 PBy/year additional (transfer rate = 4 GBy/sec) After 5 years: Australia: LOW: 0.575 EBy, SURVEY: 0.575 EBy South Africa: MID: 1.15EBy

Mid term archives (faster IO): LOW: 30 PBy, SURVEY: 70 PBy, MID: 70 PBy

Storage for a full hemisphere of sky: SURVEY: 15 EBy (taking 9 months), MID: 260 EBy (taking 9.5 years)



Data Motion

SKA Data Rates, <u>not stored</u>:

From the telescopes to the correlators: SKA1-LOW: 911 stations @ 10 Gb/s/stat = 9.1 Tb/s SKA1-MID 190 dishes + MeerKAT = 254 dishes at 24 Gb/s/dish = 6.1 Tb/s SKA1-SURVEY 60 dishes ASKAP = 96 dishes @ 2Tb/s = 192 Tb/s (Faulkner 2013; low-bit data)

From the correlators to the Science Data Processor (Dolensky 2015, SDP): LOW: 7.3 TBy/s, MID: 3.3 TBy/s, SURVEY: 4.6 TBy/s

What is the transmission limit for optical fiber? ...

the bandwidth of light, and at 1.5 μ m, that is ~ 200 Tb/s when many wavelengths simultaneously carry the information.

Shannon-Hartley theorem: Max Info Rate = $BW \log_2(1+S/N)$

Ultra-high-density spatial division multiplexing with a few-mode multicore fibre van Uden et al. Nature Photonics (2014)

... demonstrate ... 5.1 Tbit s⁻¹ carrier⁻¹ (net 4 Tbit s⁻¹ carrier⁻¹) on a single wavelength over a single fibre.

... with 50 wavelength carriers ... 255 Tbit s^{-1} over a 1 km fibre link







15 cables to Australia with an estimated 20 Tb/s lit capacity Recall: SKA1 archive transfer rate to other sites: 4 GBy/s (32 Gb/s), no problem...

Summary

- Processors: Moore's law for <u>systems</u> still on track
 - ** \$T/year investment: exponential growth continues to the SKA1 era
 ** also assumed by the SKA SDP 2015 report
- Memory: changing to become more energy efficient (non-volatile)
 - also allows extremely fast large memory spaces ("solid state memory")
 - However, new memory technologies are more expensive now
- SKA1 data volumes are not excessive by world standards
- SKA1 raw data rates are large but the technology should handle it.
- Not discussed: software changes, machine learning, neural networks, ...

Selection of radio pulsar candidates using artificial neural networks

R. P. Eatough,^{1,2★} N. Molkenthin,¹ M. Kramer,^{1,2} A. Noutsos,¹ M. J. Keith,^{1,3} B. W. Stappers¹ and A. G. Lyne¹

¹Jodrell Bank Centre for Astrophysics, Alan Turing Building, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL
 ²Max-Planck-Institut f
ür Radioastronomie, Auf dem H
ügel 69, 53121 Bonn, Germany
 ³Australia Telescope National Facility, CSIRO, PO Box 76, Epping, NSW 1710, Australia

THE ASTROPHYSICAL JOURNAL, 781:117 (12pp), 2014 February 1

doi:10.1088/0004-637X/781/2/11

© 2014. The American Astronomical Society. All rights reserved. Printed in the U.S.A.

SEARCHING FOR PULSARS USING IMAGE PATTERN RECOGNITION

W. W. ZHU¹, A. BERNDSEN¹, E. C. MADSEN¹, M. TAN¹, I. H. STAIRS¹, A. BRAZIER², P. LAZARUS³, R. LYNCH⁴, P. SCHOLZ⁴, K. STOVALL^{5,6}, S. M. RANSOM⁷, S. BANASZAK⁸, C. M. BIWER^{8,9}, S. COHEN⁵, L. P. DARTEZ⁵, J. FLANIGAN⁸, G. LUNSFORD⁵, J. G. MARTINEZ⁵, A. MATA⁵, M. ROHR⁸, A. WALKER⁸, B. ALLEN^{8,10,11}, N. D. R. BHAT^{12,13}, S. BOGDANOV¹⁴, F. CAMILO^{14,15}, S. CHATTERJEE², J. M. CORDES², F. CRAWFORD¹⁶, J. S. DENEVA¹⁷, G. DESVIGNES³, R. D. FERDMAN^{4,18}, P. C. C. FREIRE³, J. W. T. HESSELS^{19,20}, F. A. JENET⁵, D. L. KAPLAN⁸, V. M. KASPI⁴, B. KNISPEL^{10,11}, K. J. LEE³, J. VAN LEEUWEN^{19,20}, A. G. LYNE¹⁸, M. A. MCLAUGHLIN²¹, X. SIEMENS⁸, L. G. SPITLER³, AND A. VENKATARAMAN¹⁵