

Mobile Über Alles

Future Scenarios

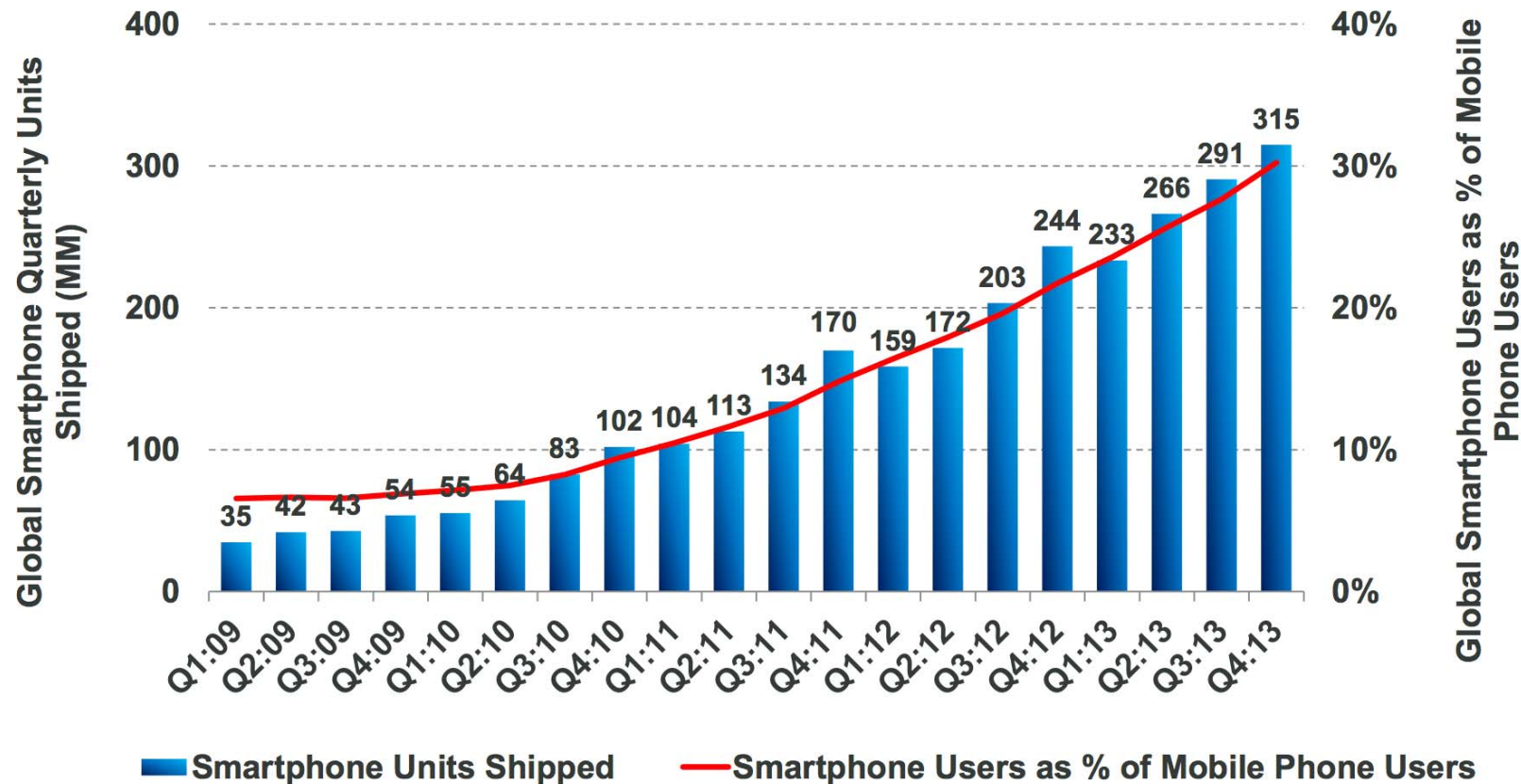
Rick Stevens

Argonne National Laboratory

The University of Chicago

Smartphone Users = Still Lots of Upside... @ 30% of 5.2B Mobile Phone User Base

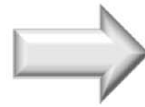
Global Smartphone Quarterly Unit Shipments & Smartphone Users as % of Mobile Phone Users, 2009 – 2013



Sensors = Big / Broad Business, Rapid Growth, Rising Proliferation /N Devices...

Apple

iPhone (2007)
3 Sensors



iPhone 5s (2013)
5 Sensors

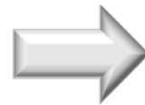


- Accelerometer / proximity / ambient light

- 3-axis gyro / fingerprint / accelerometer / proximity / ambient light

Samsung

Galaxy S (2010)
3 Sensors



Galaxy S5 (2014)
10 Sensors

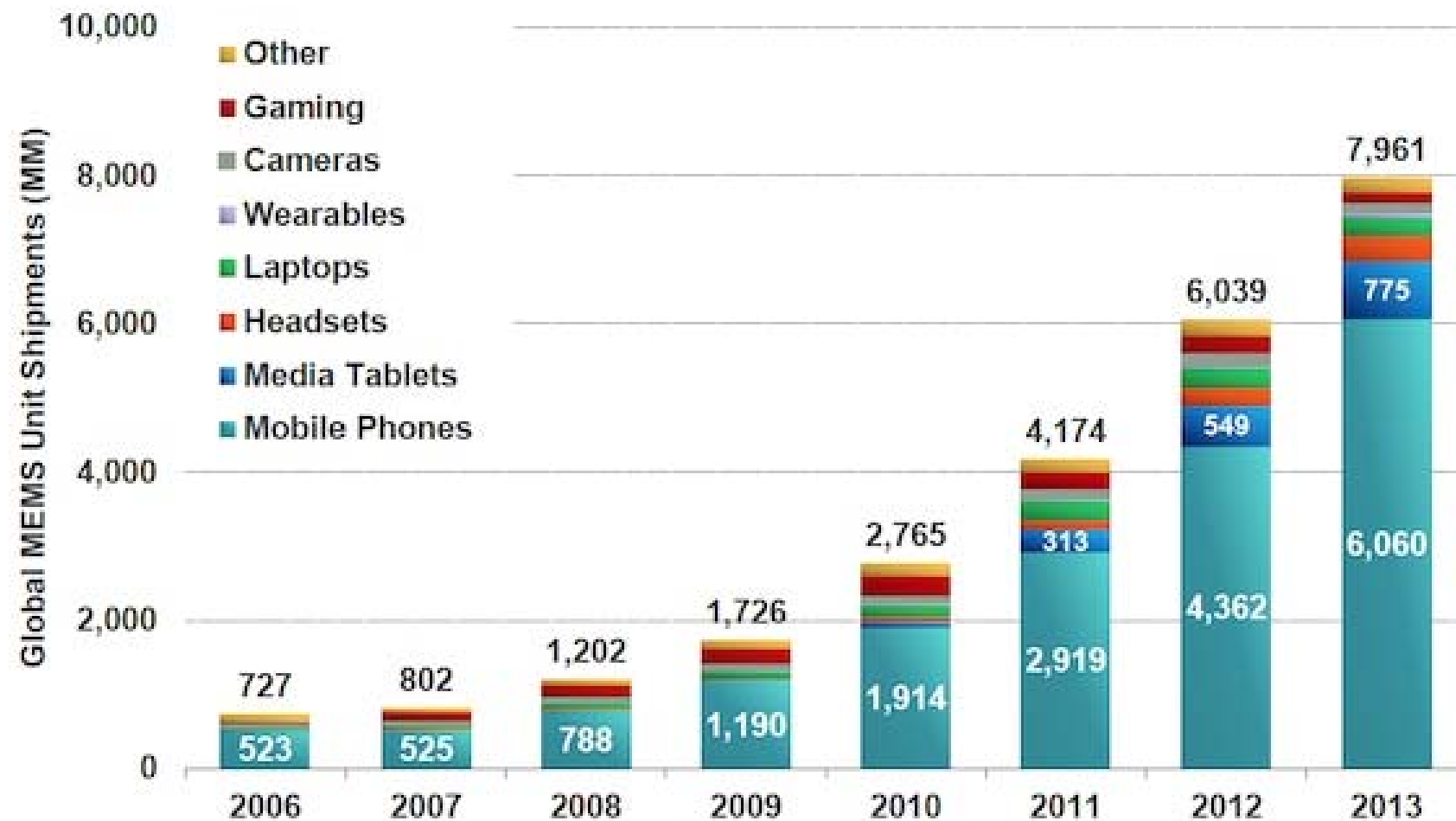


- Accelerometer / proximity / compass

- Gyro / fingerprint / barometer / hall (recognizes whether cover is open/closed) / RGB ambient light / gesture / heart rate / accelerometer / proximity / compass

...Sensors = Big / Broad Business (+32% Y/Y to 8B) Rising Proliferation OF Devices

Global MEMS Unit Shipments by Consumer Electronics Device, 2006 – 2013

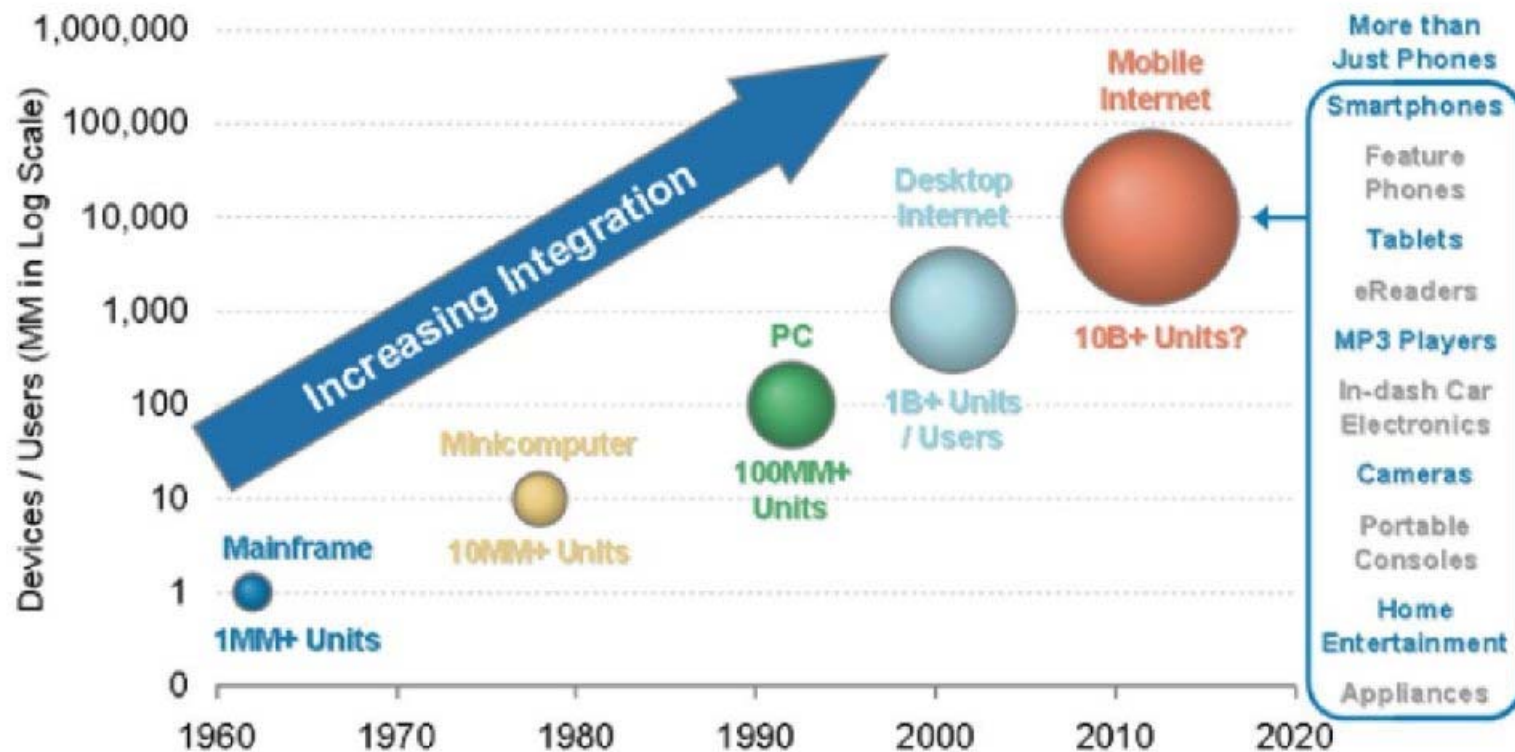


Each New Computing Cycle = 10x > Installed Base than Previous Cycle

Exhibit 29

Each new computing cycle typically generates around 10x the installed base of the previous cycle

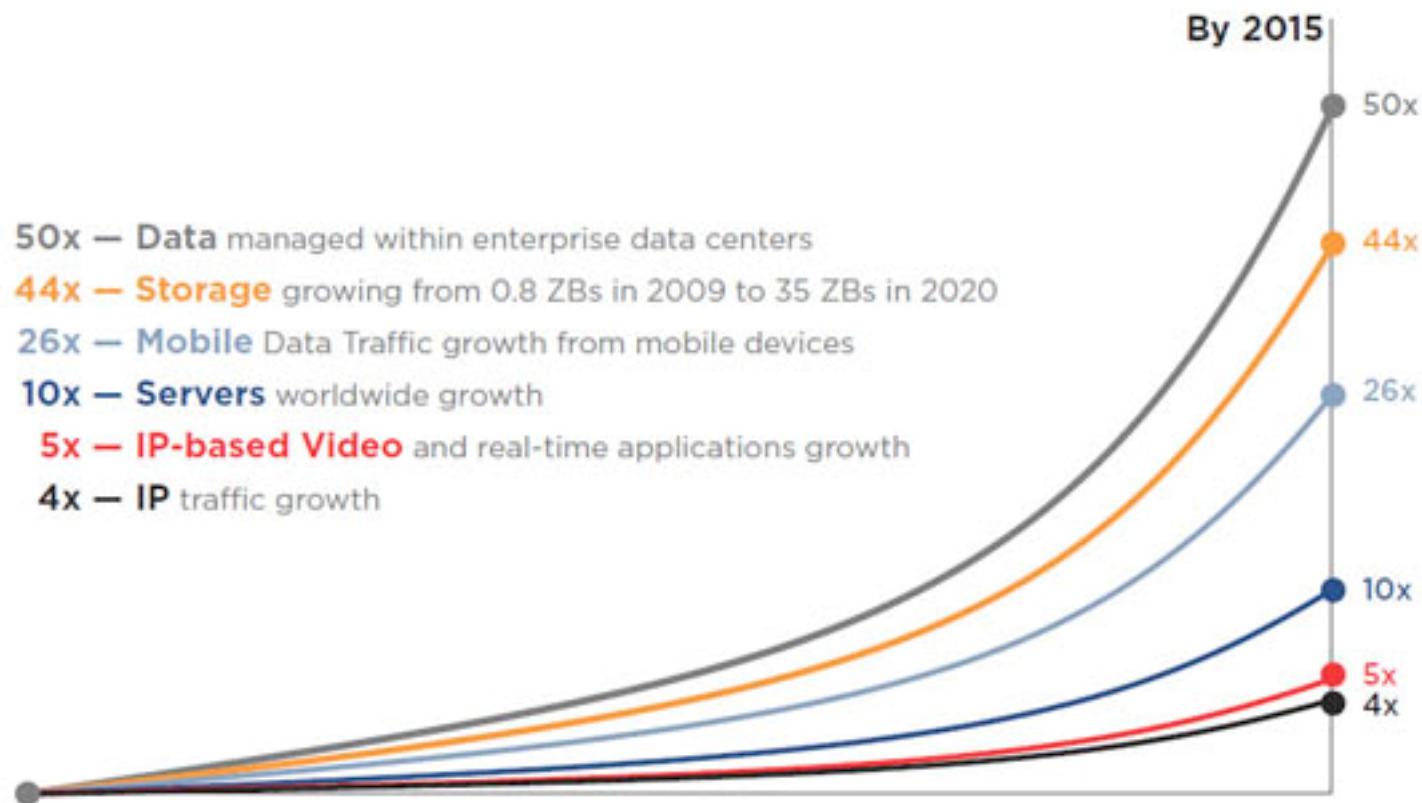
Devices or users in millions; logarithmic scale



Big Picture Trends

- **Mobile Smartphone, Tablet (MSPt+) market is increasingly dominating the computing industry with ~1.5 billion devices**
 - Mary Meeker report suggests this
 - Chip maker actions confirm this
- **Emerging internet of things (IoT) could be larger (perhaps much more so) if trends continue**
 - Long term bets in Arduino by Intel suggests this
 - NEST purchase suggests this
 - I'll call this the “**Catlett Hypothesis**” there is also a wearables conjecture that is related to this
- **The hyper growth is in edge (consumer) devices**
 - By design the market at the consumer level is more visible and played out in public
 - Notice Apple hiring luxury goods execs
 - Qualcomm market cap ~ intel market cap

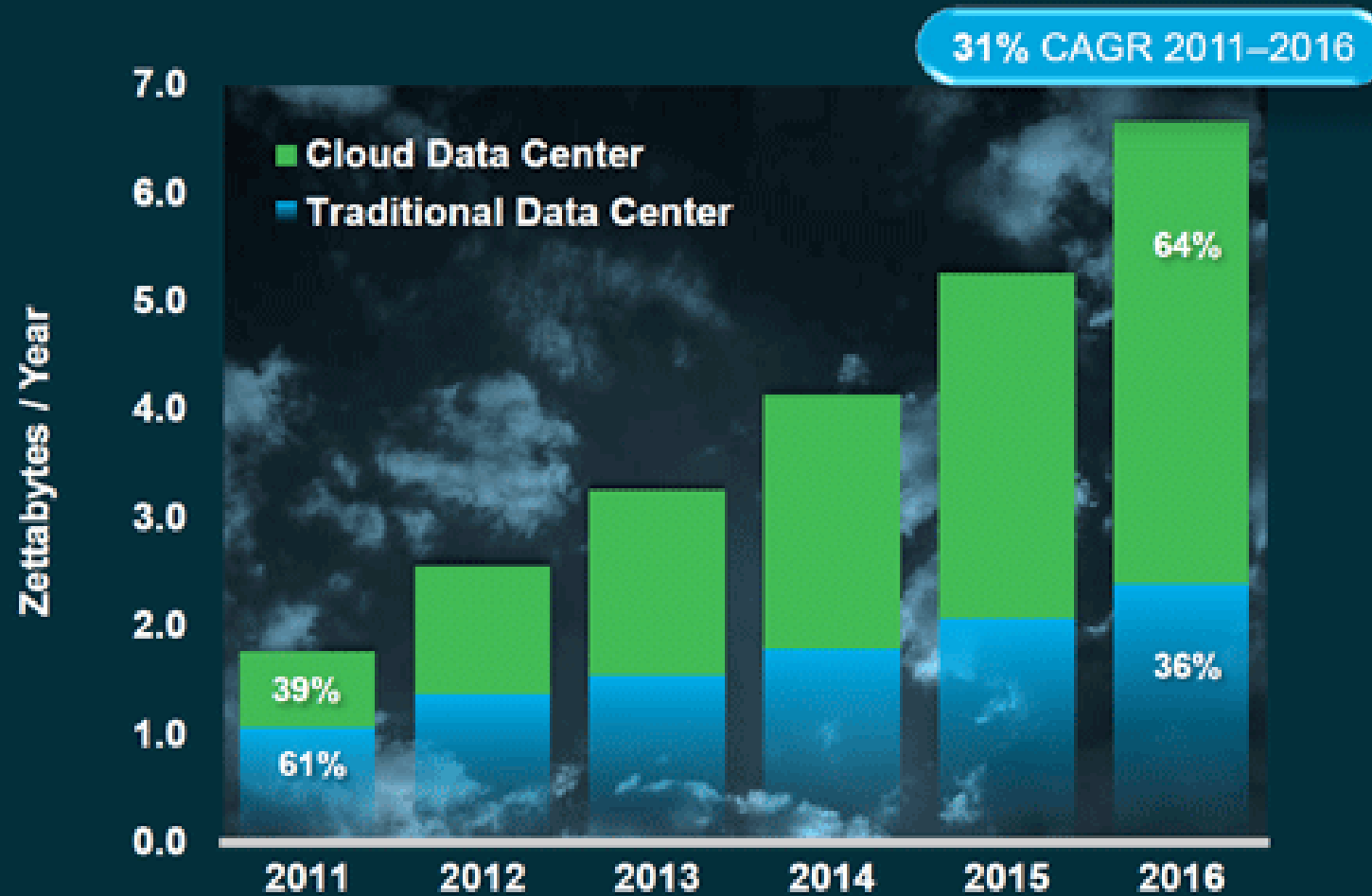
Relative Growth Rates



Source: Cisco VNI, June 2011; Gartner, 2009 & 2011; IDC, 2011

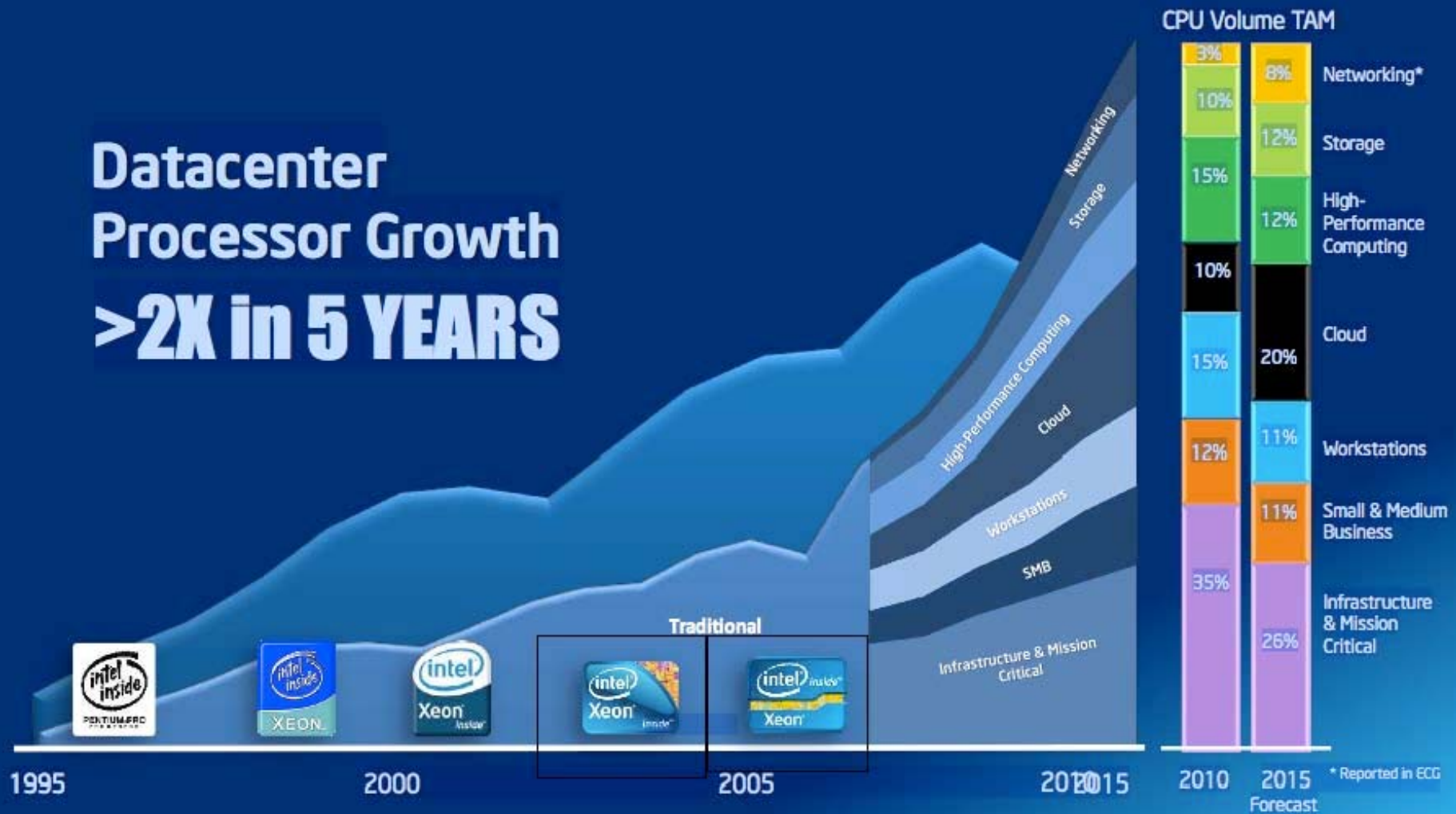
Global Data Center Traffic: Traditional vs. Cloud

Cloud Accounts for Nearly Two-Thirds of Data Center Traffic by 2016



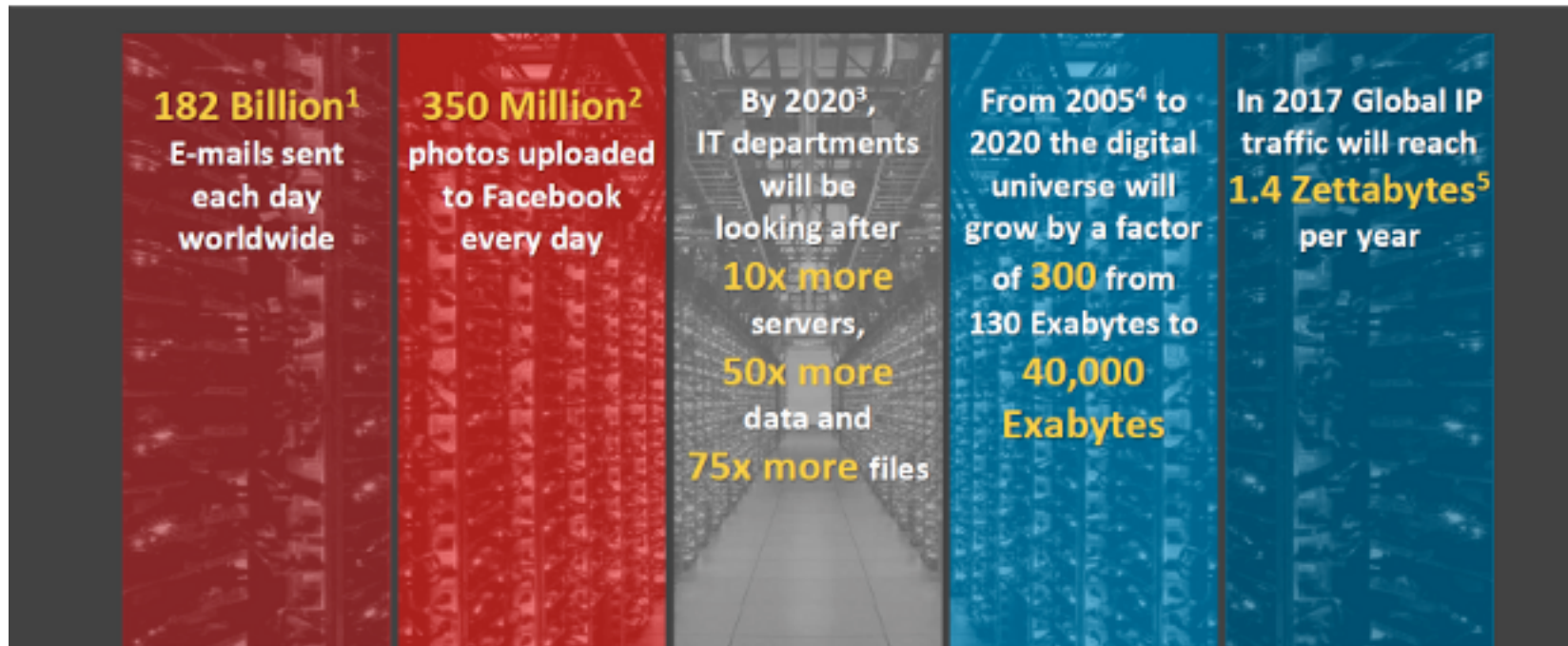
Source Cisco

Datacenter Processor Growth >2X in 5 YEARS



Data Volume Growth

The Cloud is Driving a Data Growth Explosion

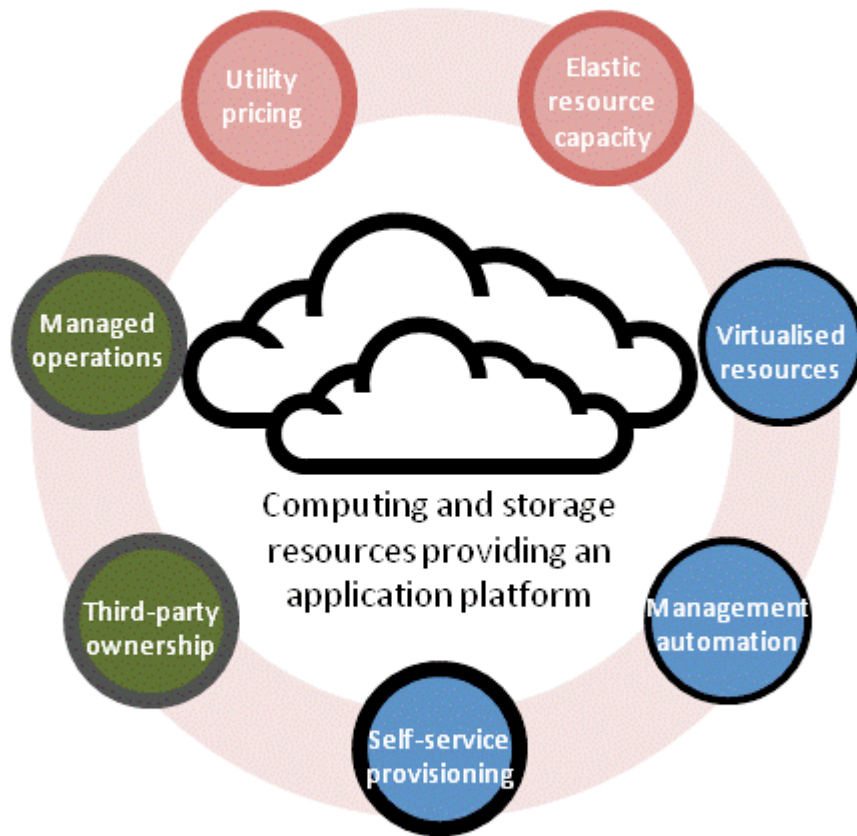


Source IDC

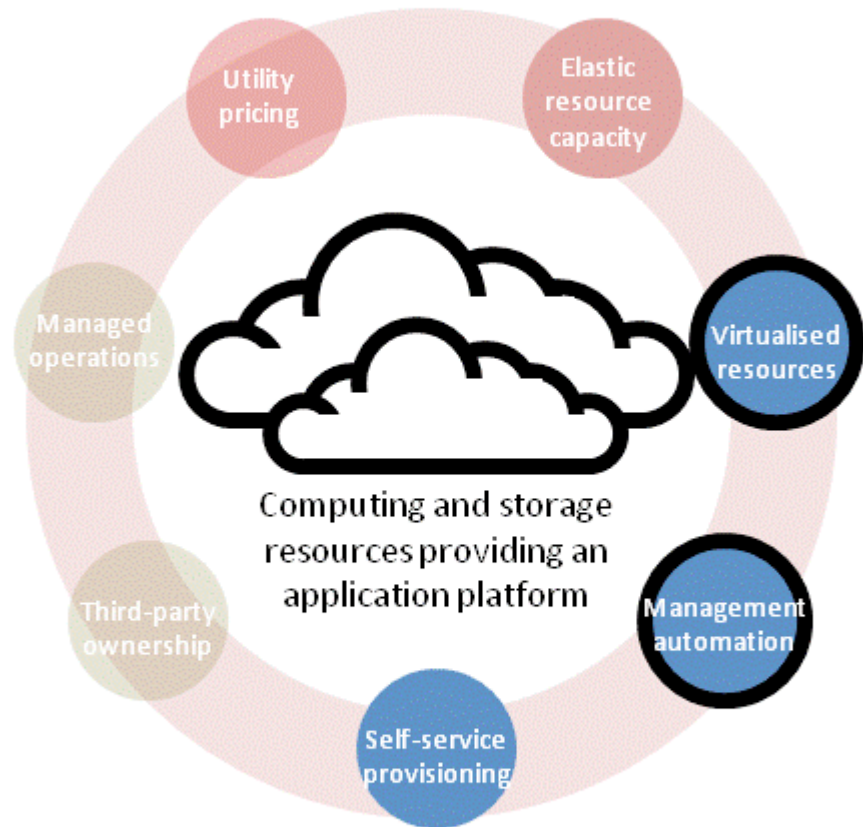
Mobile Drives Cloud

- **The growth rate in backend services/servers is somewhat lower and competition very difficult**
 - The market has driven down margins close to zero
 - IBM is divesting from x86 (ditto from “disks”, “PCs”, “networking”, “fabrication” and other businesses that have failed to drive profit)
- **Cloud deployments (non-customer owned, remotely accessed) appear to be the primary delivery mode**
 - Relatively small number of global cloud providers
 - Many services hosted on larger players infrastructures (Big/Little business models)
- **Cloud increasingly will encroach on “on-premises” or “enterprise owned and operated” infrastructures**
 - New ventures tend to start in cloud
 - Refactoring of applications targets cloud when possible
 - Move towards X as a service hosted on public cloud

Public clouds



Private clouds

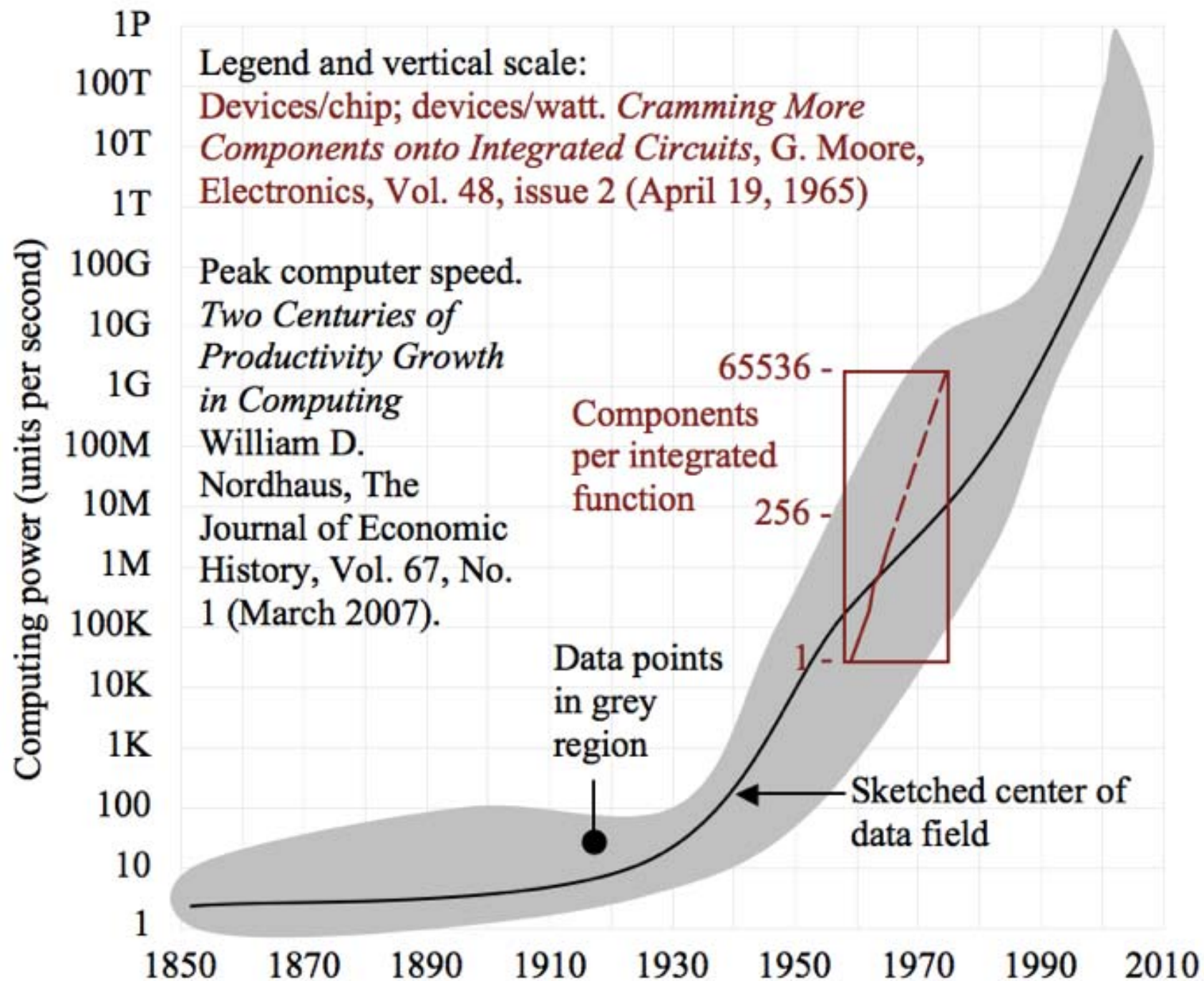


Public/Private Cloud?

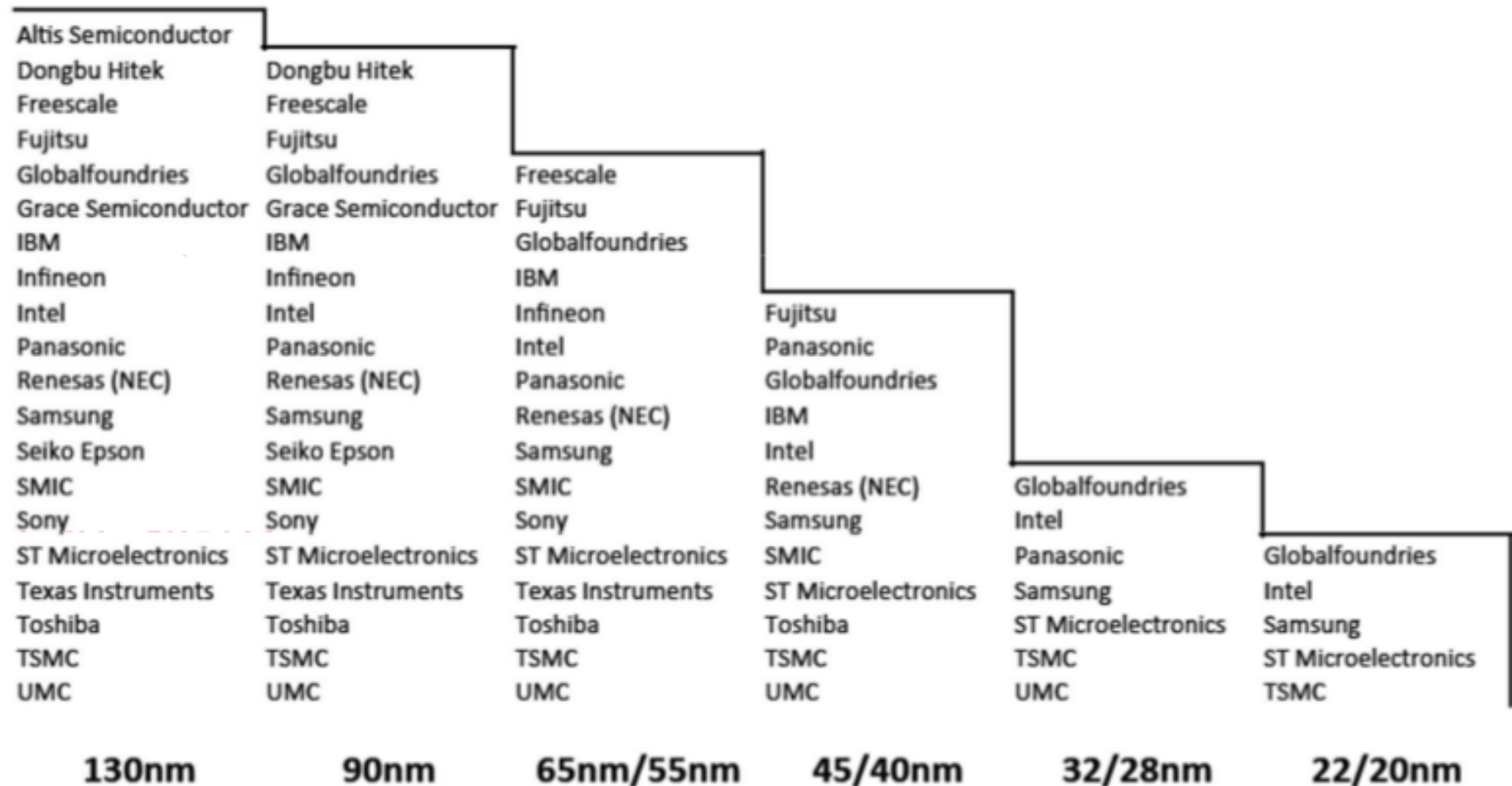
- **Private Cloud/Public Cloud hybrids will emerge for those areas of special (institutional) concerns**
 - “Same Stack Different Datacenter”
 - Legal uncertainty in data privacy/ownership
 - Government surveillance pressures
- **Many of the future technical needs of backend service “clouds” are converging with what has been developed for HPC systems**
 - I call this the “**Satoshi Hypothesis**”
 - Driven by the ratio of internal to external actions and derived events (HPC is an extreme limit of this)
 - I summarize this as tighter coupling, (store, memory, compute, network) (SoC + NoC + MoP)

The End of Moore's Law as we've known it and Implications

- We are now roughly ten years into the rollover on clock, power etc.
- Understanding the timing of transitions, options and impacts is hard
- Nothing is emerging that has a strong consensus to replace CMOS in the next ten years
- Later this week we have a speculation session
- So what's after the CMOS-based transistor? Carbon nanotubes and graphene get the most attention. Over time, the industry could migrate towards stacked-die or monolithic 3D devices. All told, there are nearly 20 viable next-generation transistor candidates on the table, although there is a possibility that CMOS may prevail over the long term.
- Surprisingly, based on the latest performance benchmarks from Intel, carbon nanotubes, graphene or even 3D devices failed to make the cut. Conducted in the lab, Intel's benchmarks are based on throughput, power consumption and other criteria. In simple terms, **the most promising devices on Intel's current list are narrowed down to five technologies—spin-majority gate; spin-wave devices; III-V tunnel field-effect transistors (TFETs); heterojunction TFETs; and graphene nanoribbon (GNR) TFETs.**



The Twilight of Moore's Law: Economics



Market volume wall: only the largest volume products will be manufactured with the most advanced technology

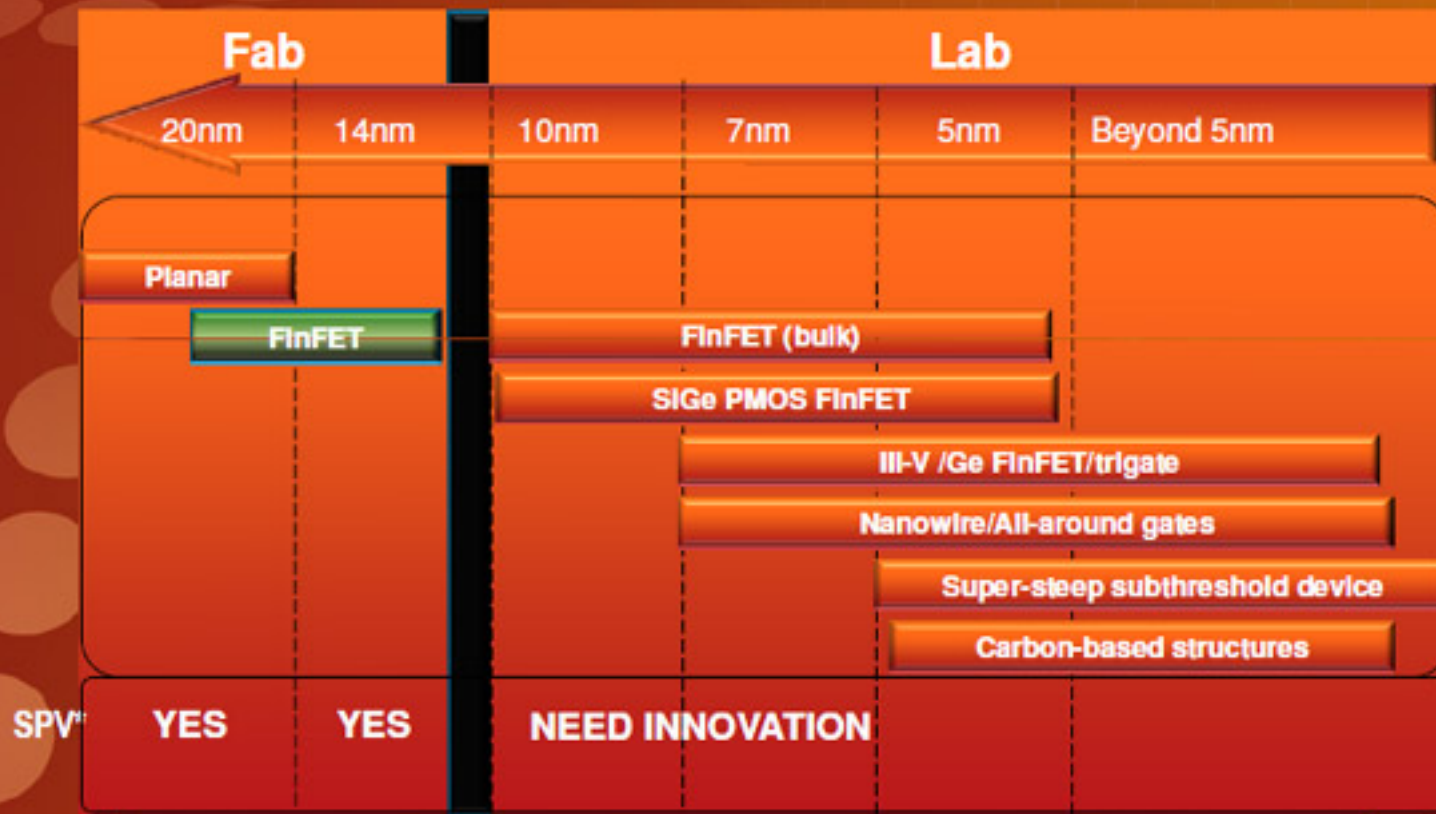


Technology Outlook

High Volume Manufacturing	2008	2010	2012	2014	2016	2018	2020	2022
Technology Node (nm)	45	32	22	16	11	8	6	4
Integration Capacity (BT)	8	16	32	64	128	256	512	1024
Delay Scaling	>0.7			~1?				
Energy Scaling	~0.5			>0.5				
Transistors	Planar			3G, FinFET				
Variability	High			Extreme				
ILD	~3			towards 2				
RC Delay	1	1	1	1	1	1	1	1
Metal Layers	8-9	0.5 to 1 Layer per generation						

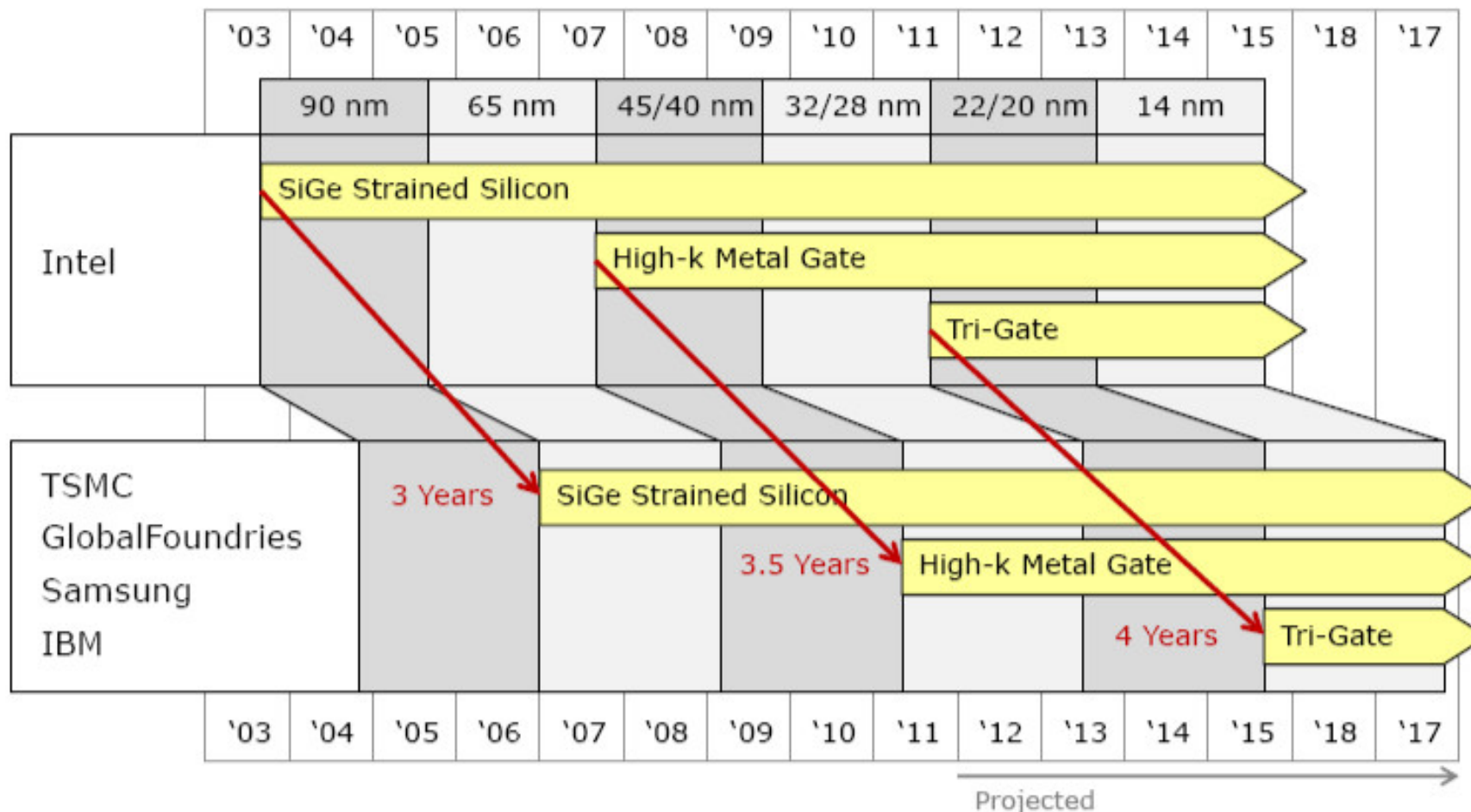
Captured from some Intel Investor Briefing a few years ago

Crossing 'Lab to Fab' Chasm



*SPV = SoC Product Value on current R&D trajectory

Intel® Transistor Leadership



Intel leads the industry in introducing new technology generations and revolutionary transistor technologies

Technology Competition

Comparison of Process Roadmaps (for Volume Production)						
	2011	2012	2013	2014	2015	2016
Intel	22nm tri-gate transistor			14nm	10nm	
GlobalFoundries	28nm			20nm	14nm finFET, 20nm BEOL	10nm 14nm BEOL
Samsung	28nm			20nm	14nm finFET, 20nm BEOL	10nm
TSMC	28nm		20nm	16nm finFET, 20nm BEOL	10nm	
UMC	28nm			14nm finFET, 20nm BEOL	10nm	

Source: Companies, conference reports, IC Insights

Figure 2

Preliminary Worldwide Ranking of the Top 20 Suppliers of Semiconductors in 2012
(Ranking by Revenue in Millions of U.S. Dollars)

2011 Rank	2012 Rank	Company Name	2011 Revenue	2012 Revenue	Percent Change	Percent of Total	Cummulative Percent
1	1	Intel	48,721	47,543	-2.4%	15.7%	15.7%
2	2	Samsung Electronics*	28,563	30,474	6.7%	10.1%	25.7%
6	3	Qualcomm	10,198	12,976	27.2%	4.3%	30.0%
3	4	Texas Instruments	13,967	12,008	-14.0%	4.0%	34.0%
4	5	Toshiba	12,729	10,996	-13.6%	3.6%	37.6%
5	6	Renesas Electronics Corporation	10,648	9,430	-11.4%	3.1%	40.7%
8	7	SK Hynix	9,293	8,462	-8.9%	2.8%	43.5%
7	8	STMicroelectronics	9,735	8,453	-13.2%	2.8%	46.3%
10	9	Broadcom	7,160	7,840	9.5%	2.6%	48.9%
9	10	Micron Technology	7,365	6,955	-5.6%	2.3%	51.2%
13	11	Sony	5,015	6,025	20.1%	2.0%	53.2%
11	12	Advanced Micro Devices (AMD)	6,436	5,300	-17.7%	1.7%	54.9%
12	13	Infineon Technologies	5,312	4,826	-9.1%	1.6%	56.5%
16	14	NXP	3,831	4,096	6.9%	1.4%	57.9%
17	15	nVidia	3,608	3,923	8.7%	1.3%	59.2%
14	16	Freescale Semiconductor	4,408	3,775	-14.4%	1.2%	60.4%
21	17	MediaTek	3,309	3,472	4.9%	1.1%	61.6%
15	18	Elpida Memory	3,887	3,414	-12.2%	1.1%	62.7%
22	19	ROHM Semiconductor	3,267	3,170	-3.0%	1.0%	63.7%
19	20	Marvell Technology Group	3,393	3,113	-8.3%	1.0%	64.8%
Top 20 Companies			200,845	196,251	-2.3%	64.8%	
All Others			109,360	106,768	-2.4%	35.2%	
Total Semiconductor			310,205	303,019	-2.3%	100.0%	

*Significant impact on growth due to Samsung Electronics acquisition of Samsung Electro-Mechanic's 50% share of Samsung LED

Source: IHS iSuppli Research, December 2012

2012 Top 25 Semiconductor Sales Leaders Ranked by Growth Rate (\$M, Including Foundries)

2012 Rank	Company	Headquarters	2011 Tot IC	2011 Tot O-S-D	2011 Tot Semi	2012 Tot IC	2012 Tot O-S-D	2012 Tot Semi	2012/2011 % Change
1	Qualcomm**	U.S.	9,828	0	9,828	13,177	0	13,177	34%
2	GlobalFoundries*	U.S.	3,480	0	3,480	4,560	0	4,560	31%
3	TSMC*	Taiwan	14,600	0	14,600	17,167	0	17,167	18%
4	Sharp	Japan	1,658	1,250	2,908	1,799	1,505	3,304	14%
5	MediaTek**	Taiwan	2,969	0	2,969	3,366	0	3,366	13%
6	Broadcom**	U.S.	7,160	0	7,160	7,793	0	7,793	9%
7	Nvidia**	U.S.	3,939	0	3,939	4,229	0	4,229	7%
8	NXP	Europe	2,855	1,292	4,147	2,931	1,226	4,157	0%
9	UMC*	Taiwan	3,760	0	3,760	3,730	0	3,730	-1%
10	Intel	U.S.	49,697	0	49,697	49,114	0	49,114	-1%
11	Samsung	South Korea	32,703	780	33,483	29,730	2,521	32,251	-4%
12	SK Hynix	South Korea	9,403	0	9,403	9,057	0	9,057	-4%
13	TI	U.S.	12,182	718	12,900	11,442	705	12,147	-6%
14	Fujitsu	Japan	4,035	395	4,430	3,805	357	4,162	-6%
15	Sony	Japan	4,706	1,387	6,093	4,449	1,260	5,709	-6%
16	Micron	U.S.	8,125	446	8,571	7,567	435	8,002	-7%
17	Rohm	Japan	1,952	1,351	3,303	1,792	1,238	3,030	-8%
18	Marvell**	U.S.	3,445	0	3,445	3,157	0	3,157	-8%
19	Infineon	Europe	3,560	2,039	5,599	3,143	1,850	4,993	-11%
20	Toshiba	Japan	10,024	2,721	12,745	9,055	2,162	11,217	-12%
21	Renesas	Japan	8,517	2,136	10,653	7,487	1,827	9,314	-13%
22	ST	Europe	7,117	2,514	9,631	6,227	2,137	8,364	-13%
23	Freescale	U.S.	3,750	641	4,391	3,164	571	3,735	-15%
24	AMD**	U.S.	6,568	0	6,568	5,422	0	5,422	-17%
25	Elpida	Japan	3,891	0	3,891	3,075	0	3,075	-21%

*Foundry

**Fabless

Source: IC Insights' Strategic Reviews Database

Who are the top Foundries

Top 12 2012 IC Foundries

2012 Rank	2011 Rank	Company	Foundry Type	Location	2010 Sales (\$M)	2011 Sales (\$M)	2011/2010 Change (%)	2012 Sales (\$M)	2012/2011 Change (%)
1	1	TSMC	Pure-Play	Taiwan	13,307	14,600	10%	17,167	18%
2	3	GlobalFoundries	Pure-Play	U.S.	3,510	3,480	-1%	4,560	31%
3	4	Samsung	IDM	South Korea	1,205	2,190	82%	4,330	98%
4	2	UMC	Pure-Play	Taiwan	3,965	3,760	-5%	3,730	-1%
5	5	SMIC	Pure-Play	China	1,555	1,320	-15%	1,682	27%
6	6	Hua Hong Grace*	Pure-Play	China	627	850	36%	940	11%
7	7	TowerJazz	Pure-Play	Israel	510	611	20%	644	5%
8	8	Vanguard	Pure-Play	Taiwan	508	519	2%	582	12%
9	9	Dongbu	Pure-Play	South Korea	475	500	5%	540	8%
10	10	IBM	IDM	U.S.	430	420	-2%	435	4%
11	11	MagnaChip	IDM	South Korea	405	350	-14%	400	14%
12	12	WIN**	Pure-Play	Taiwan	221	298	35%	382	28%
—	—	Top 12 Total	—	—	26,718	28,898	8%	35,392	22%
—	—	Top 12 Share	—	—	87%	88%	—	90%	—
—	—	Other Foundry	—	—	4,017	3,972	-1%	3,918	-1%
—	—	Total Foundry	—	—	30,735	32,870	7%	39,310	20%

*Hua Hong NEC and Grace merged in 2012 (includes Shanghai Huali joint venture).

**GaAs foundry

Source: IC Insights, company reports

Samsung, teaming up with [Globalfoundries](#), has landed orders for its 14nm FinFET process from Qualcomm and Apple, reports DigiTimes. Industry sources say related foundry services will begin in early 2015.

(K WPM)	2016				2017			
	16/14nm	20nm	28nm	TOTAL	16/14nm	20nm	28nm	TOTAL
Company A	44	48	76	168	72	54	70	196
Company B	19	46	32	97	34	57	15	106
Company C	8	12	14	34	13	12	10	35
Company D	7	9	11	27	10	12	10	32
Company E	6	15	32	53	14	22	33	69
Others	5	16	44	65	13	19	78	110
TOTAL	89	146	209	444	156	176	216	548

**DEMAND FOR 16/14nm TECHNOLOGY WAFERS
IS CONCENTRATED WITHIN SMALL NUMBER OF COMPANIES
IN 2016 AND POTENTIALLY 2017**

APL0598 339S0176

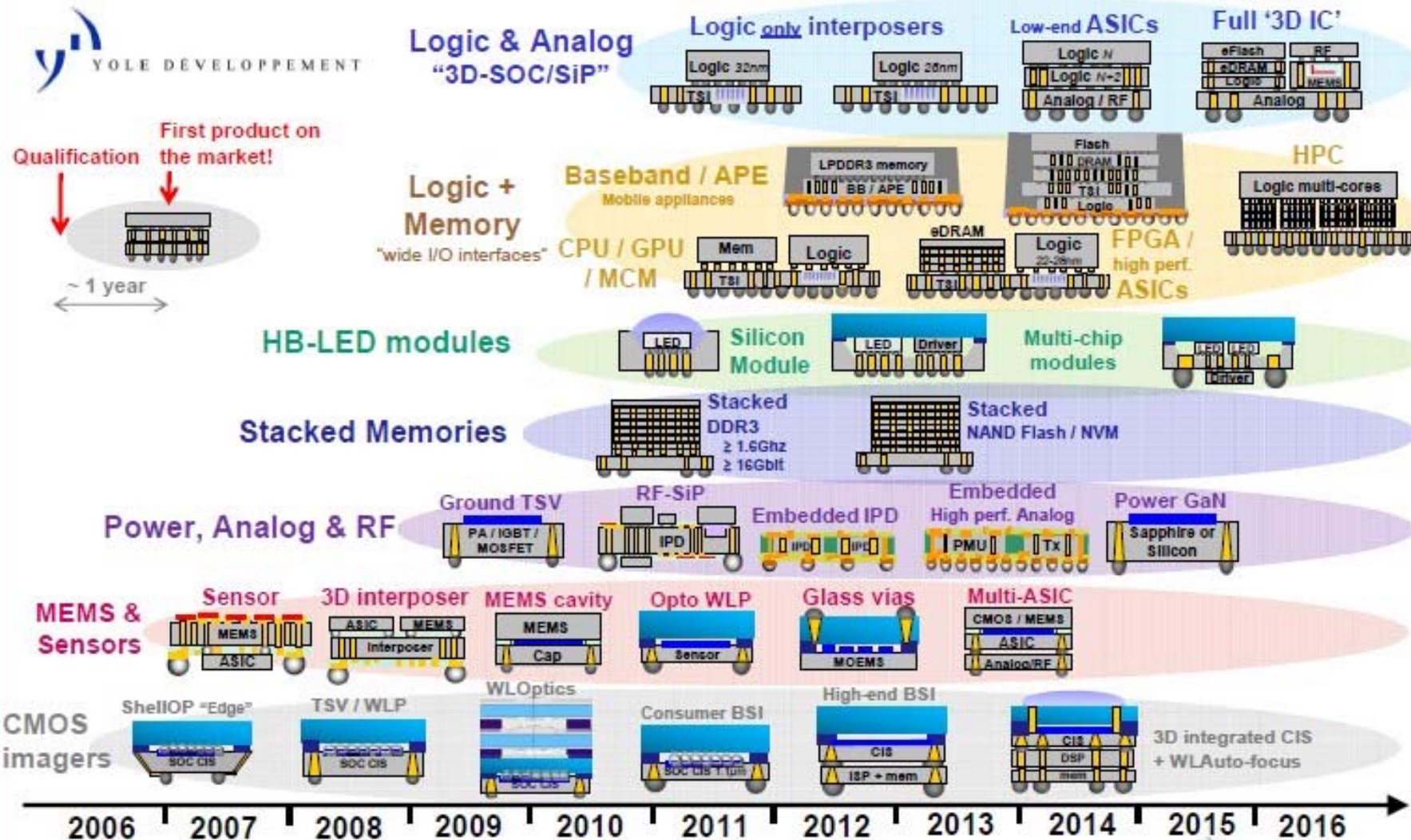


K3PE7E700F - XGC2

GKBC40HI 1213

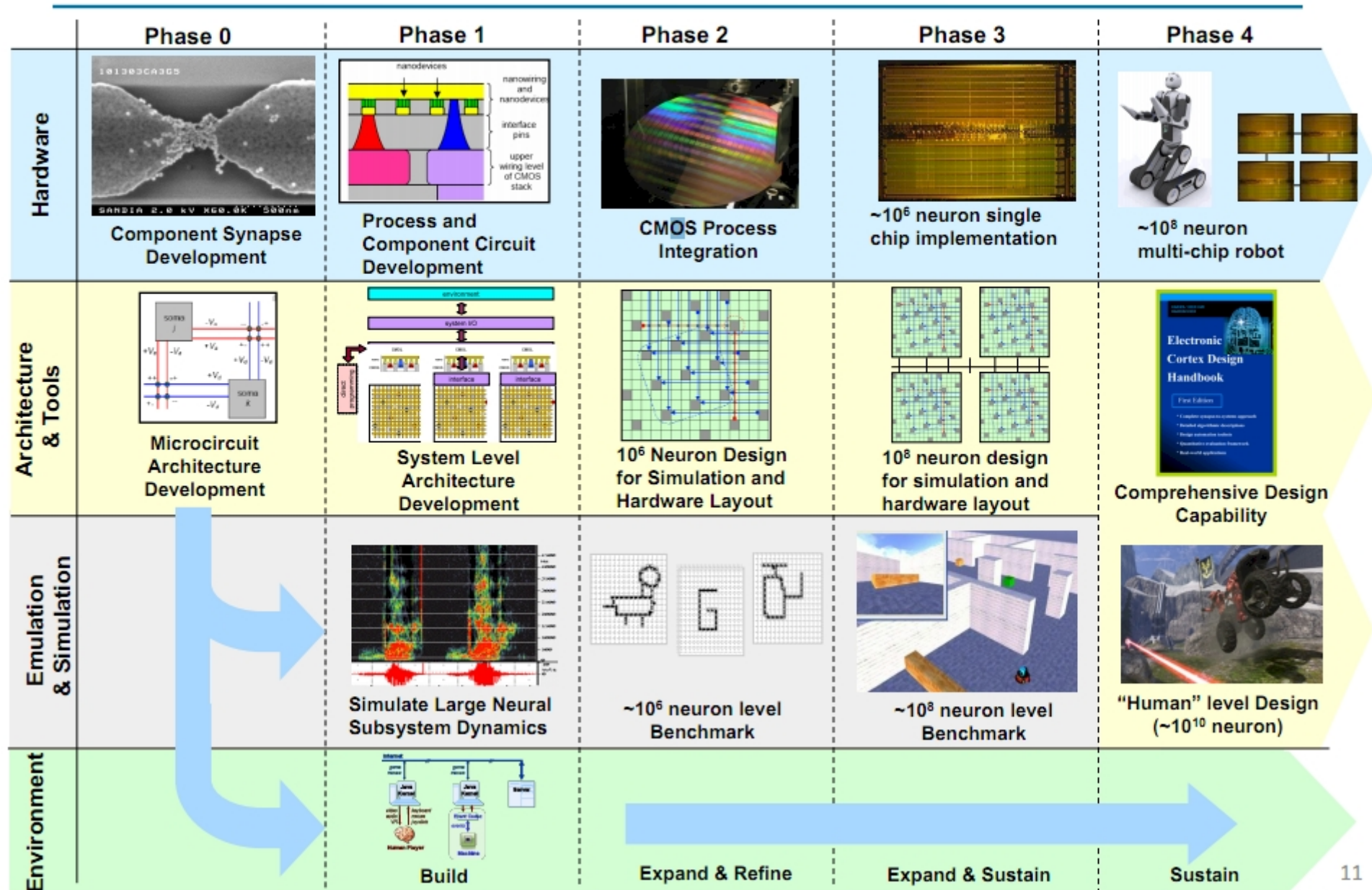
3SC0091 1222

Global Roadmap for 3D Integration with TSV



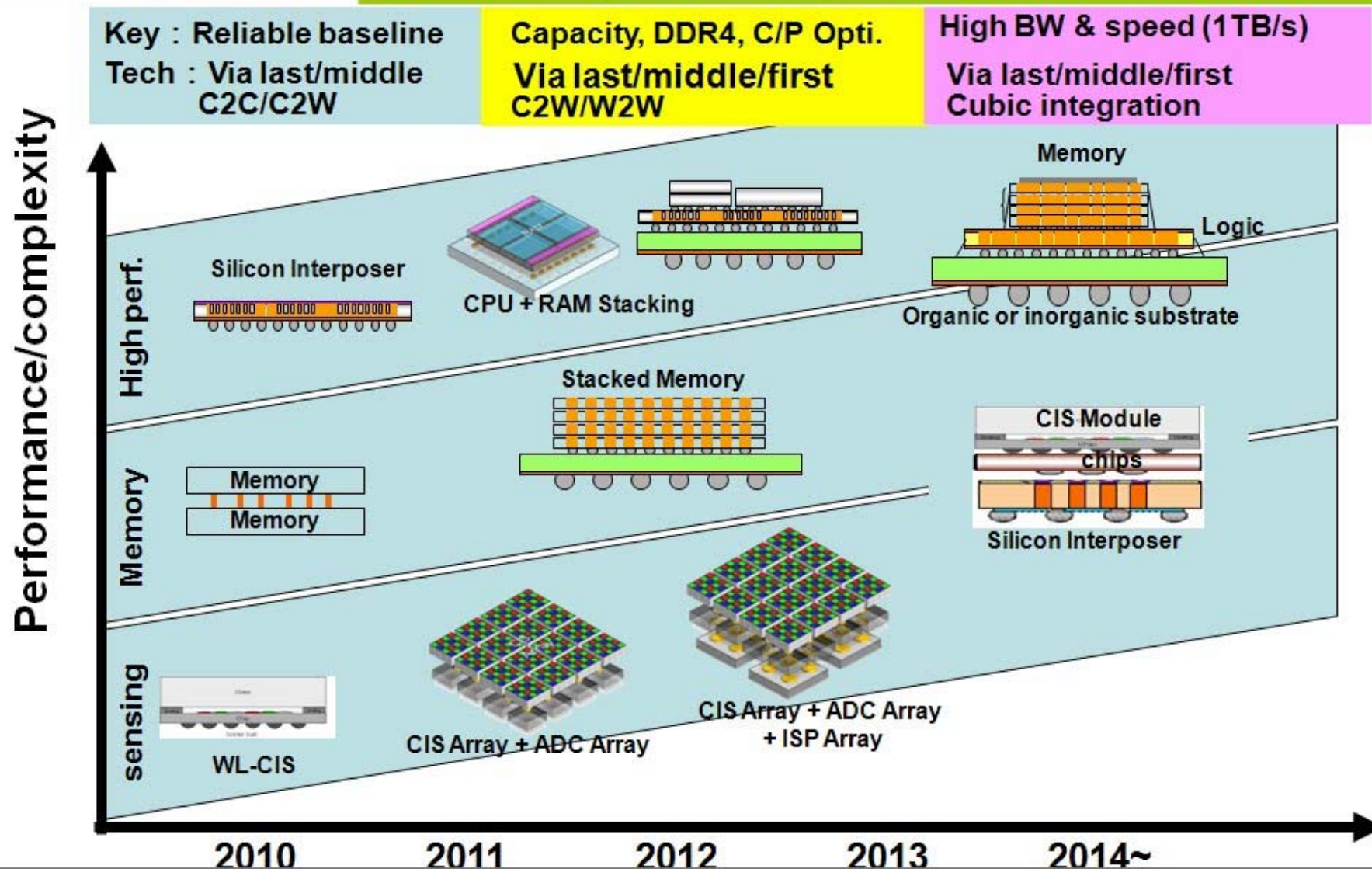


SyNAPSE Program Plan





ITRI's 3DIC Roadmap



What does this mean for HPC?

1. Architecture (and Technology)
2. Infrastructure (and data centers)
3. Software Stack (real-time, big data)
4. Applications (optimization, prediction)
5. Science (health, safety, learning)

Architecture-Technology

- The mobile world is currently a mixture of 32bit and some 64bit and 16bit implementations
 - GPUs, vectors, app specific IP blocks
 - Need for lower power
 - Need for more memory/storage and deeper integration (smaller cheaper packaging)
 - Need for offloading alg/apps to more efficient hardware when possible
- **Mobile and Things are primarily “interfaces” between people and things and the cybereal**
 - Mobile is (thin?) layer that needs to speak human on one side and machine on the backend
 - Things is a layer that needs to speak material objects on one side and machine on the backend
 - Occasionally Mobile and Things will talk to peers but almost always mediated via cyber something
 - Wearables are sorta like Things attached to people.. Bridge between the object side of the people the people side and cyber side

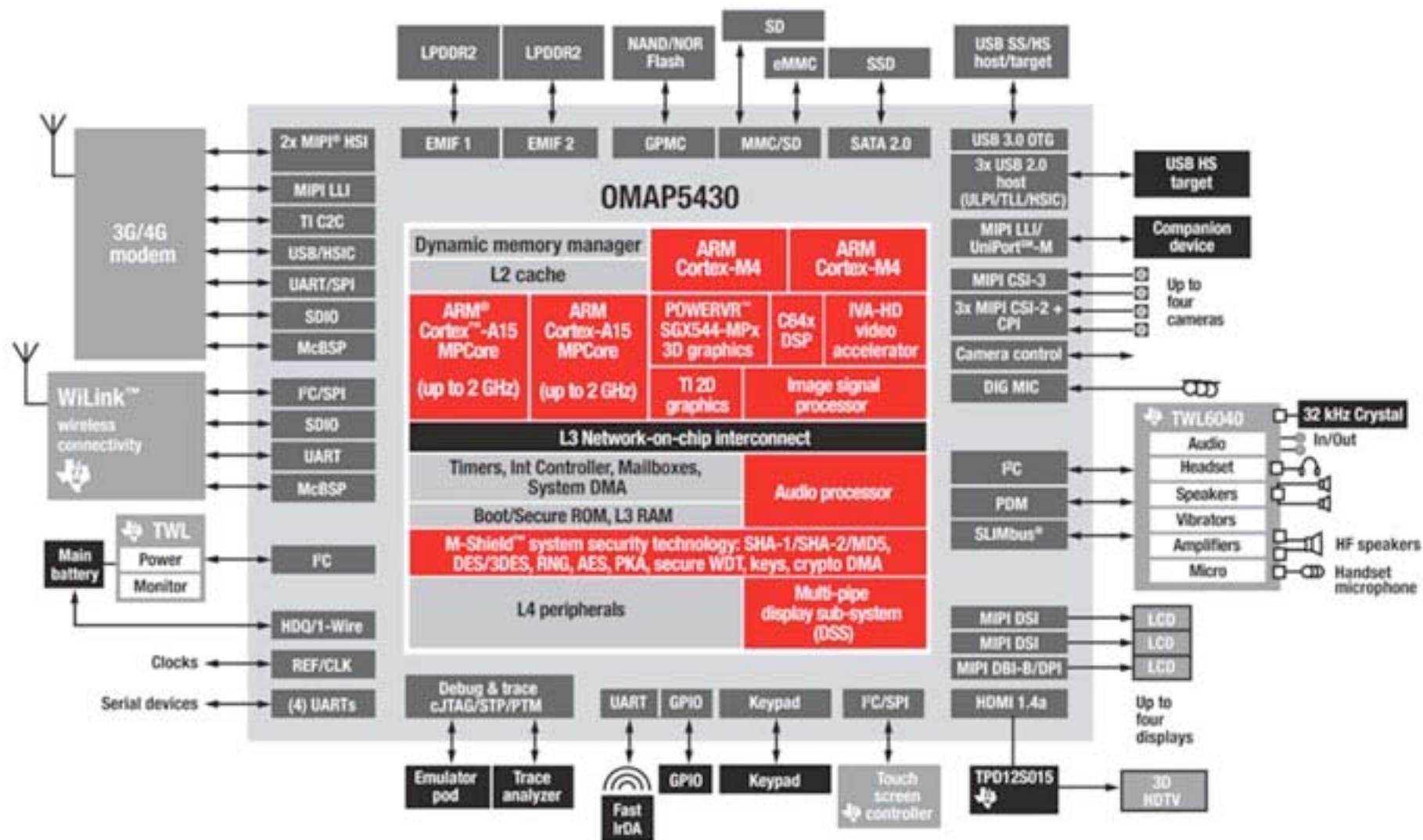
Architecture and Technology

- In the long term it seems to me that Mobile/Wearable become more “biological”, “embedded”, “implicit”
 - E.g. Ender’s game Ansible type things or various sci-fi realm embedded augmentations (vision, audio, comm, thinking. etc.)
 - This doesn’t necessarily imply that they will be less relevant for HPC or future large-scale systems
- However mobile doesn’t today need some things we need (**extreme reliability, scalability, streaming bandwidths, sustained throughput.** etc)
 - Sleep most of the time but wakeup and do many things then go back to sleep
 - Though the bottoms up argument has some potential (ARM + friends pushing up)
 - Top down is also interesting (Intel pushing down)

Architecture-Technology

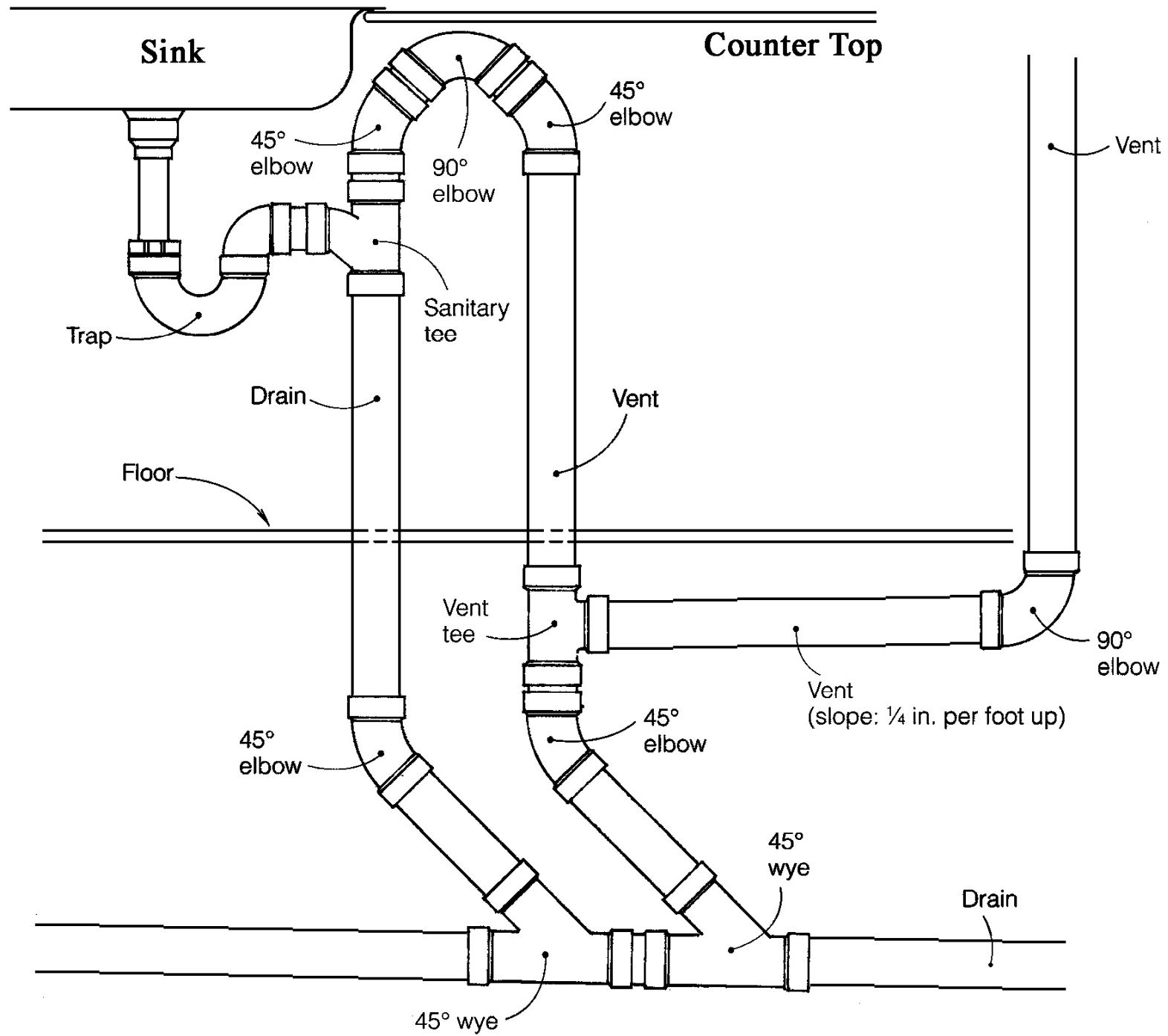
- **The server side of the “mainstream” market does need some of these things and others**
 - More cost efficient platforms (reduced footprint based on services)
 - Driving push away from general purpose x86 towards ARM, GPUs, perhaps even Power! Perhaps retuned architectures from intel
 - Scalability of a sort (infrastructure, and applications)
 - Quickly deploy large-quantities of servers and manage them with few a people as possible
 - Systems need to support clusters of a scale for the given application suite
 - Redundancy is okay as long as its cost effective
 - Continuous deployment
- **There today is little overlap of the specialized IP blocks that are needed in Mobile with the specialized IP blocks that could make HPC systems more efficient**
 - Many of these come from signal processing and media processing which are limited in HPC spaces (though not in cloud see AWS freemuim)

TI OMAP5430 SoC



Infrastructure and data centers

- Server side of the mobile ecosystem
 - Large-scale deployments driven by cost, security, environmental profile (hydro, solar, free cooling)
 - Largely air cooled, commodity (repackaged) building blocks
 - Scale of Machine rooms is 10x-20x the HPC community largest deployments
 - Power and cooling densities typically lower than HPC centers
 - Aggregate bandwidths into the centers can be comparable
 - Lights out remote management, small local teams
 - Common management software across many systems and many sites
 - Integrated interconnect is a disruption in the current hardware model
 - Custom mini-node deployments would be disruptive
 - Virtualization and Containerization
 - Deep software stacks between providers, services and users
 - Same style of infrastructure used for services development



Mobile + Data vs HPC Software Stack

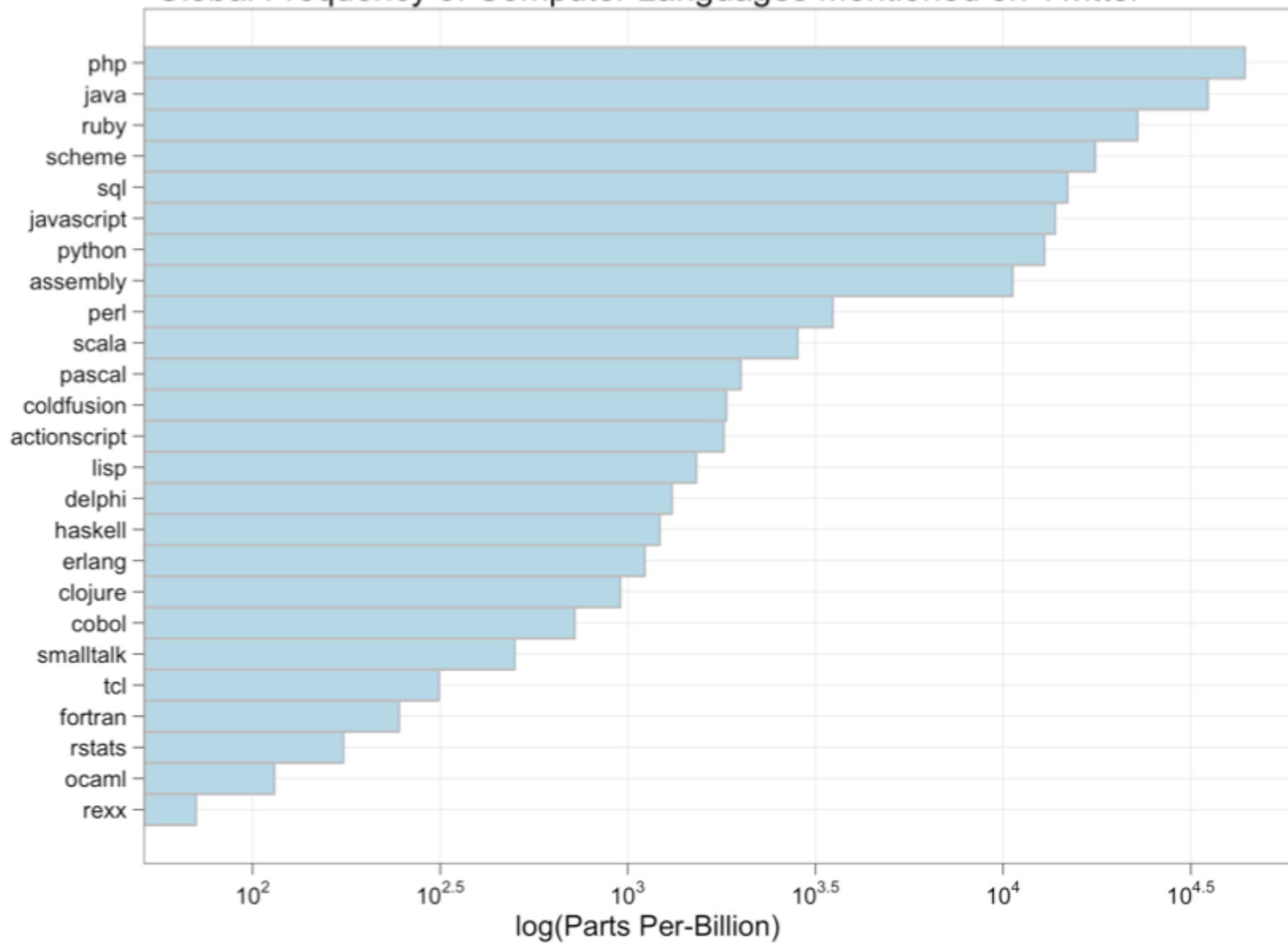
- How much leverage can HPC expect from the mobile world for software?
- In Cloud backend space I see a few things impacting HPC
 - OpenStack and other AWS compatibility layers
 - Docker (reimagined containers)
 - noSQL/KV data management MongoDB, CouchDB, Cassandra, Accumulo, etc.
 - Web software stacks (apache*, etc.)
 - NLP software stacks (NLPT, StanfordNLP, etc.)
- Much plumbing, but little HPC can build on directly to support HPC applications
- Much reinvention (workflow is a good reinvention area)

Programming Language Trends

Position Feb 2013	Position Feb 2012	Delta in Position	Programming Language	Ratings Feb 2013	Delta Feb 2012	Status
1	1	=	Java	18.387%	+1.34%	A
2	2	=	C	17.080%	+0.56%	A
3	5	↑↑	Objective-C	9.803%	+2.74%	A
4	4	=	C++	8.758%	+0.91%	A
5	3	↓↓	C#	6.680%	-1.97%	A
6	6	=	PHP	5.074%	-0.57%	A
7	8	↑	Python	4.949%	+1.80%	A
8	7	↓	(Visual) Basic	4.648%	+0.33%	A
9	9	=	Perl	2.252%	-0.68%	A
10	12	↑↑	Ruby	1.752%	+0.19%	A
11	10	↓	JavaScript	1.423%	-1.04%	A

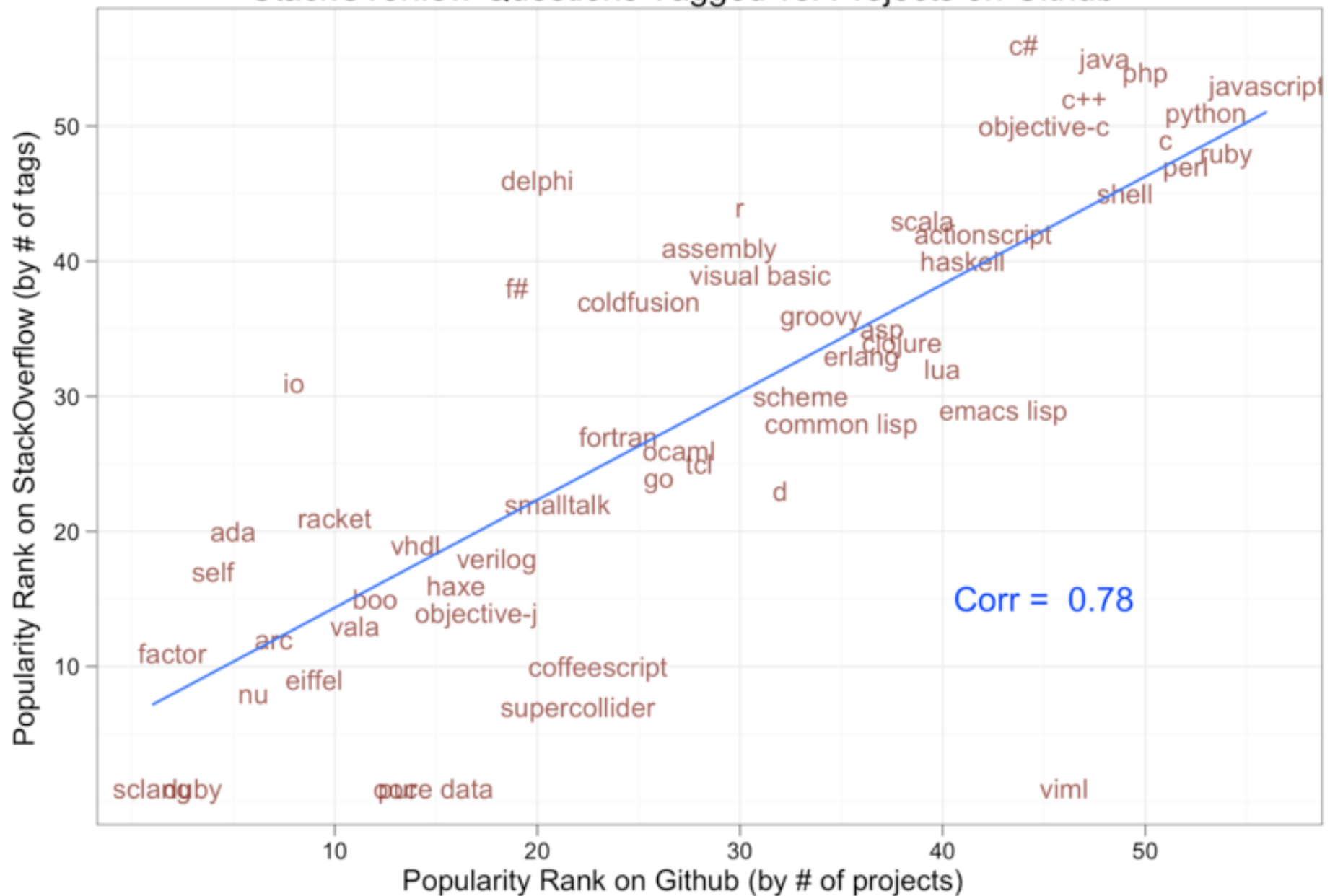
Source TIOBE

Global Frequency of Computer Languages Mentioned on Twitter



Programming Language Popularity

StackOverflow Questions Tagged vs. Projects on Github



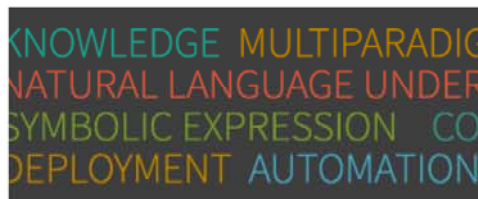


Wolfram Language™

[Home](#)[Principles](#)[Uses](#)[Code Gallery](#)[For Experts](#)[Q&A](#)[Ecosystem](#)[Documentation](#)[Community](#)

Introducing knowledge-based programming...


Designed for the new generation of programmers, the Wolfram Language has a vast depth of built-in algorithms and knowledge, all automatically accessible through its elegant unified symbolic language. Scalable for programs from tiny to huge, with immediate deployment locally and in the cloud, the Wolfram Language builds on clear principles—and 25+ years of development—to create what promises to be the world's most productive programming language.

[Principles and Concepts »](#)[Stephen Wolfram's Introduction »](#)[Where to Use the Wolfram Language »](#)[Scope and Documentation »](#)[For Language Experts »](#)[Products and Ecosystem »](#)

SteganographyWLExample...wolfram language - Googl...

https://programming.wolframcloud.com/app/objects/04bd30ae-e63b-4bd2-9c8f-9a2a536dd691wolfram language

Most VisitedGetting StartedLatest HeadlinesProxyt!RStudio Sign InHome - Parallel P...Rental Mobile Wif...WorldTravelSIM.c...China 3G Mobile ...Travel Abroad wit...Generate Possible...

WOLFRAM PROGRAMMING CLOUD [BETA]Free | Upgrade>>

Search


Cloud FilesNew


RecentHomeAll FilesDeploymentsReportsResources & ControllersShared by YouShared with YouTrash

SteganographyWLExample.nbFileFormatInsertViewDeploy

Hide Secret Messages in Images

Hide messages or other data in an image for covert distribution.



InsertSecretMessage[, "This is a secret message"]

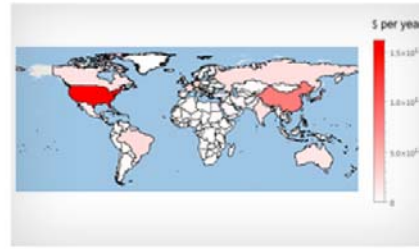
code

```
InsertSecretMessage[carrierImage_Image, msg_] :=  
Block[{carrierBytes, pixelChannels, secretBits, secretBytes},  
  carrierBytes = BitAnd[ImageData[carrierImage, "Byte"], 2 ^ 11 111 110];  
  secretBits = Flatten[IntegerDigits[ToCharacterCode[ToString[msg, InputForm],  
    CharacterEncoding -> "ASCII"]], 2, 8];  
  secretBytes = Fold[Partition, PadRight[Join[IntegerDigits[Length[secretBits], 2, 48], secretBits],  
    Apply[Times, Dimensions[carrierBytes]]], Reverse@Rest[Dimensions[carrierBytes]]];  
  Image[carrierBytes + secretBytes, "Byte"]  
]  
  
ExtractSecretMessage[img_Image] :=  
Block[{secretData, messageLength},
```

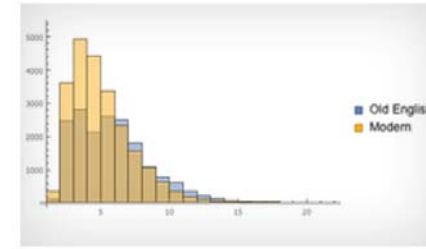
EvaluateDashboardsConsoleCommunityFeedback



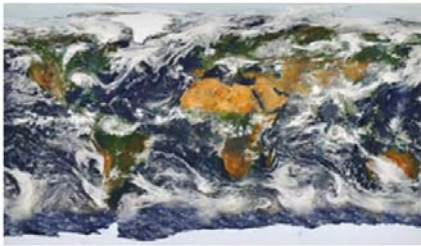
Make a Hipstamatic Filter



Color Countries by GDP



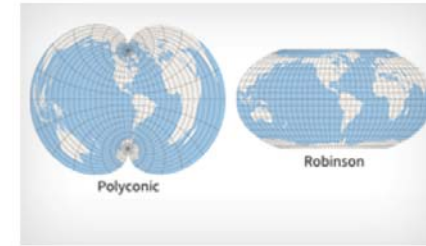
Compare Old and Modern English



Make a Capital Temperature URL



Make Pop Art



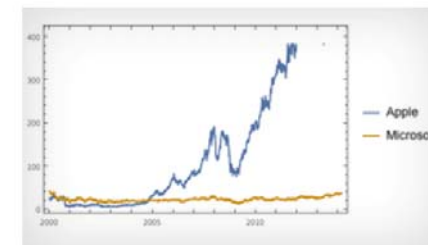
Make a You-Centric World Map

How many days old are you?
`DateDifference["Jan 4, 1987", Today]`
 9927
 How many days until your next birthday?
`DateDifference[Today, "Jan 4, 2015"]`
 300

Find Your Age in Days

Find the population of this country:

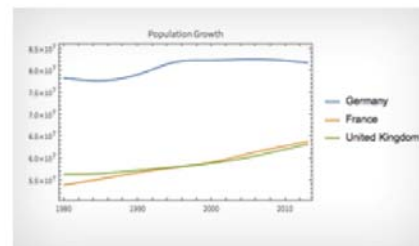
Find the Population of a Country



Plot Histories of Stock Prices



Plan a City Tour



Plot Population Growth



Make an Elevation Map

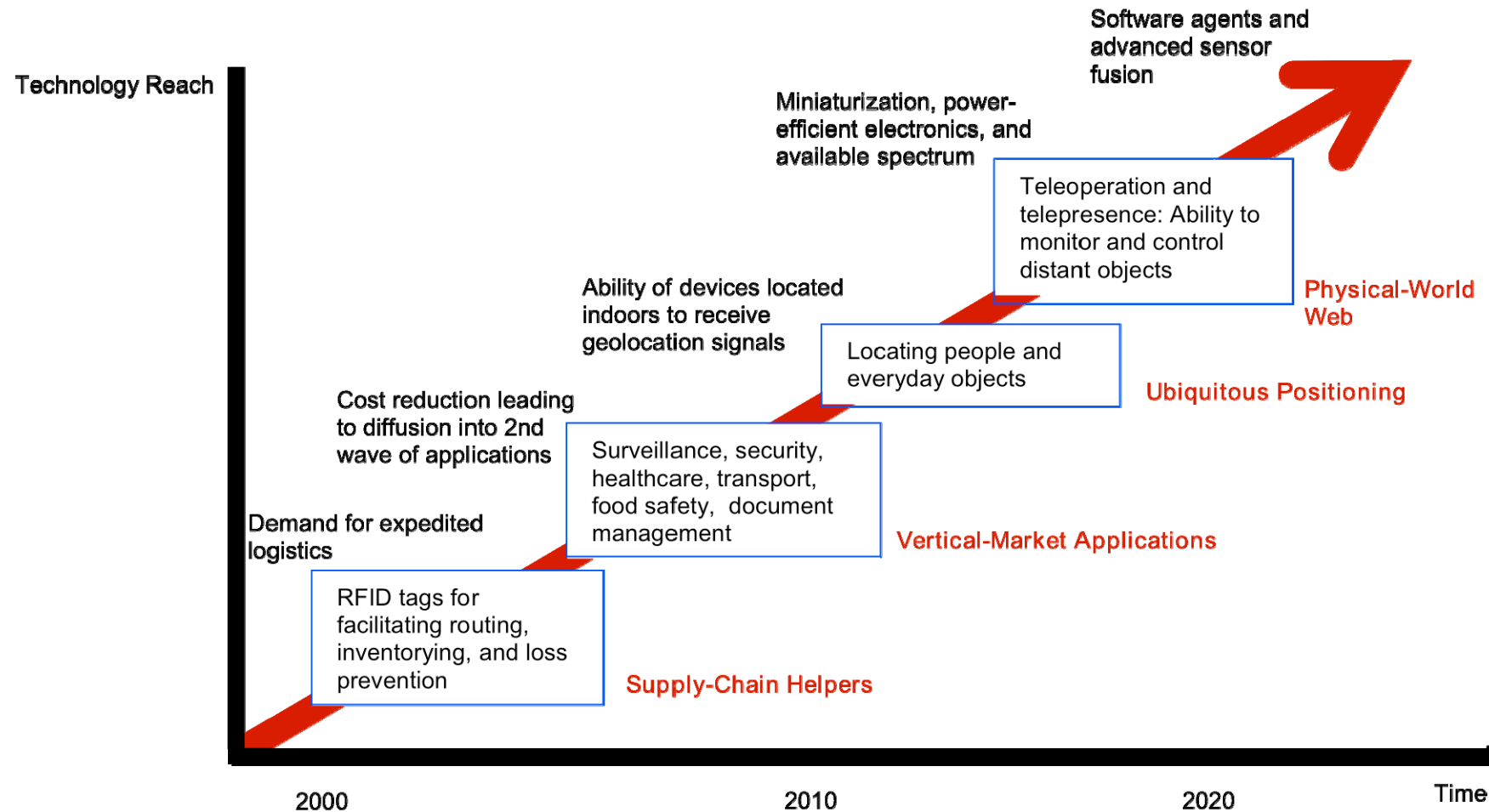
SUMMARY

Internet of Things



- Volume >> that of mobile?
- Downward scaling from mobile platforms
- New software stacks?
- Sensors and New Networks
- Cloud Backends 1-1, 1-N, N-1, N-M
- Data collectors or smart devices?
- Integration with neuromorphic ideas
- Ultra-low power, self powered, disposable?
- Linkage to digital fabrication

TECHNOLOGY ROADMAP: THE INTERNET OF THINGS



Source: SRI Consulting Business Intelligence

Cloud Infrastructures



- Growth paralleling that of mobile, IoT
 - HPC vs Clouds
 - Some HPC/HTC workloads can be moved over to clouds and should be
 - Increasing sophistication of Cloud software stack
 - Virtualization, containerization, runanything"ization"
 - "Wolfram NL programming applied to virtualization" like concepts
 - Internal "core" of clouds becoming more HPC like
 - Integration of fabrics, resource management etc.
 - But much more managed
 - Economics dramatically improving and sophisticated (see AWS TCO app, freemium business models)
-
- Big Data integration will be slower than expected due to imaturing tools, business models and data movement charging models
 - Major players still evolving rapidly in the ecosystem (Amazon, Microsoft, Google, China, etc.)
 - Will have a big impact on server supplier business models

HPC Through 2024

- Two or three main directions
 - SoCores+vectors, Big.little+vectors, CPU.GPU
 - billion way concurrency
- Road to Exascale
 - possible in 2021, definite by 2024 (\$200M-\$300M)
- Memory cost is a dominate \$ problem
- Power is a challenge... raising to 30MW eases things a lot
- Will be a rocky transition of applications
 - MPI+{Threads+ {vectors, gpu}}
 - But more dynamic program/execution models desired
- New programming models that leverage advanced runtime support (e.g. IPM, etc.) could enable new applications
- Knowledge programming is what many domains want (Biology, etc.) and while it needs HPC its much more than that see things like our DOE KBase project

Motivation for New Structures

