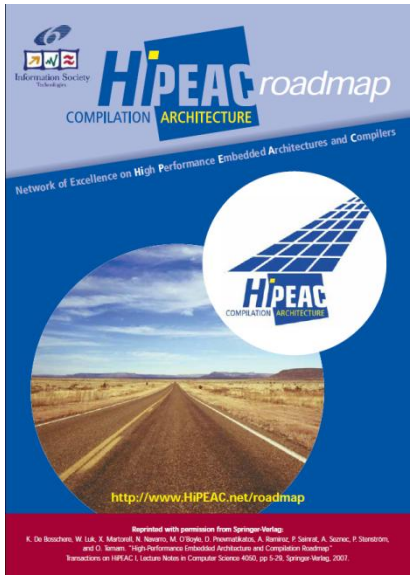
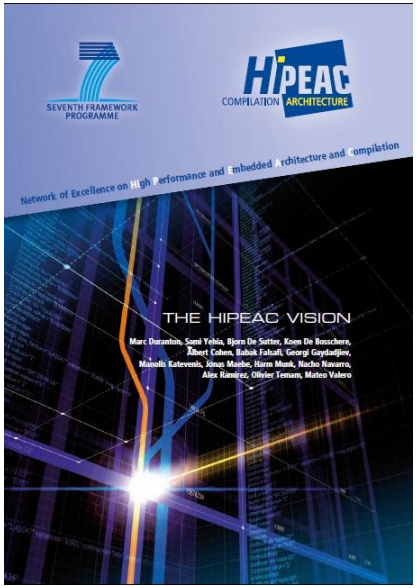
Koen De Bosschere, Ghent University

ARCS 2012 Keynote – February 29, 2012

HiPEAC Roadmaps





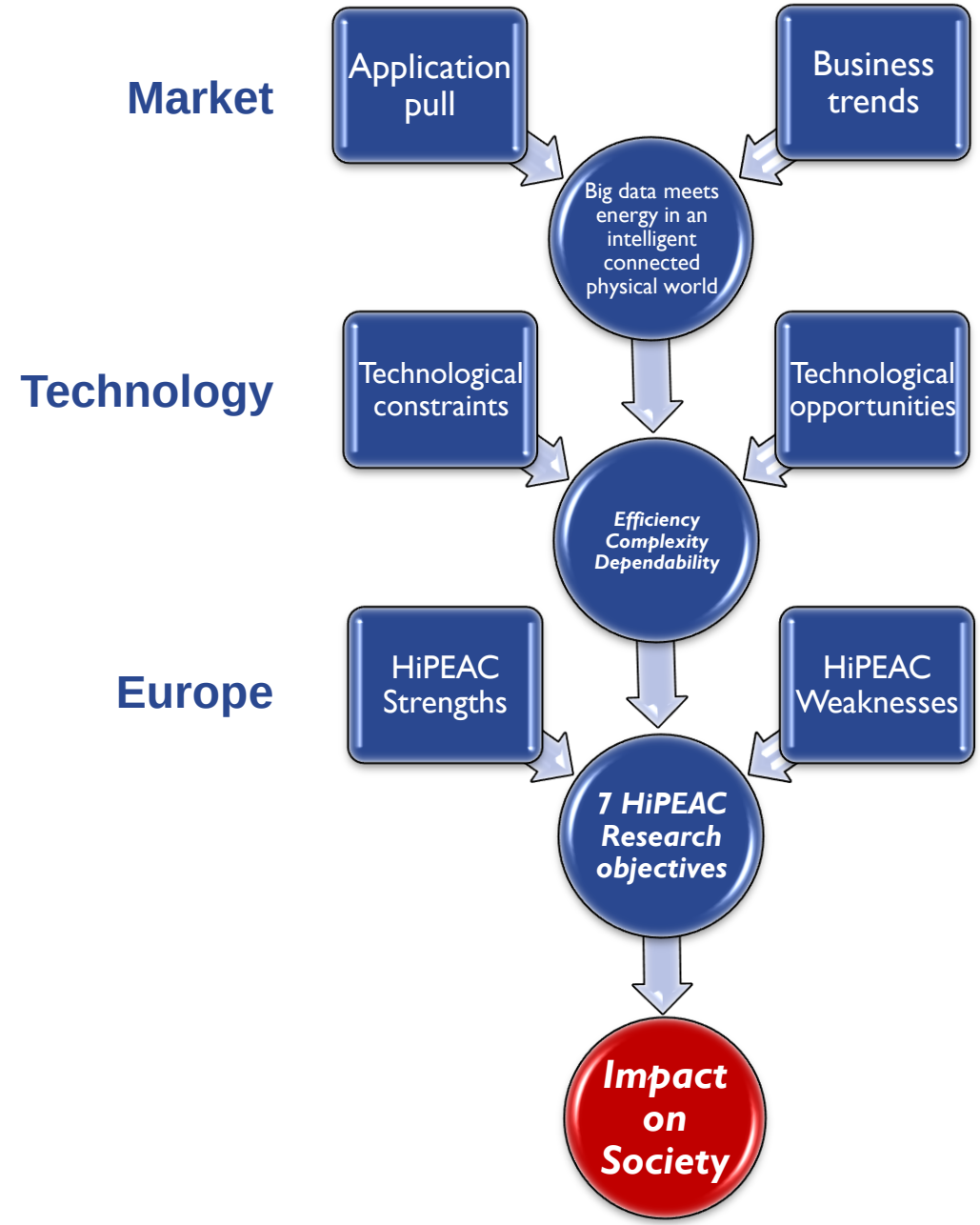
2008 2009 2011

[Marc Duranton, David Black-Shaffer, Sami Yehia, Koen De Bosschere, "Computing Systems: Research Challenges Ahead. The HiPEAC Vision 2011-2012"]

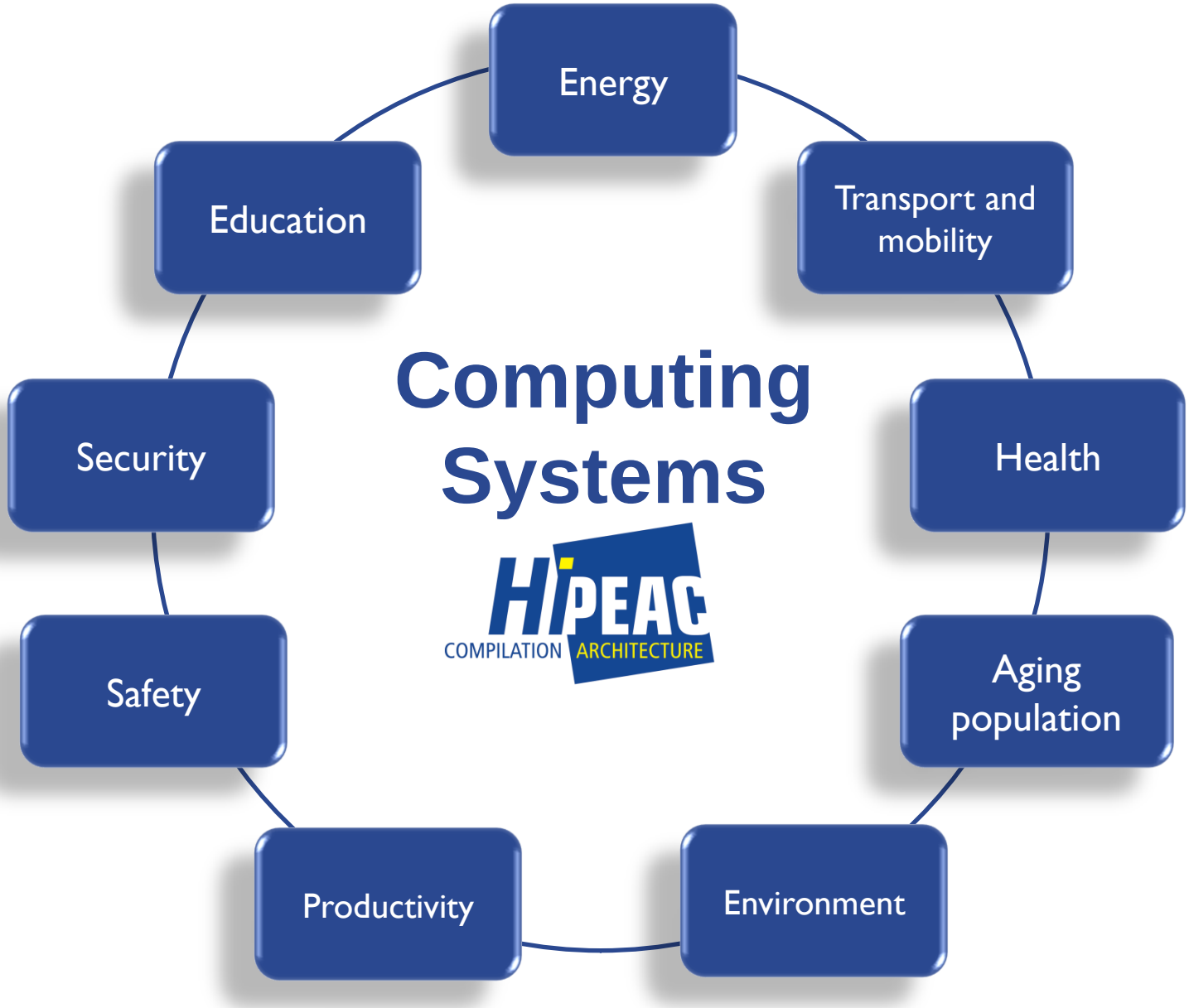
http://www.hipeac.net/roadmap

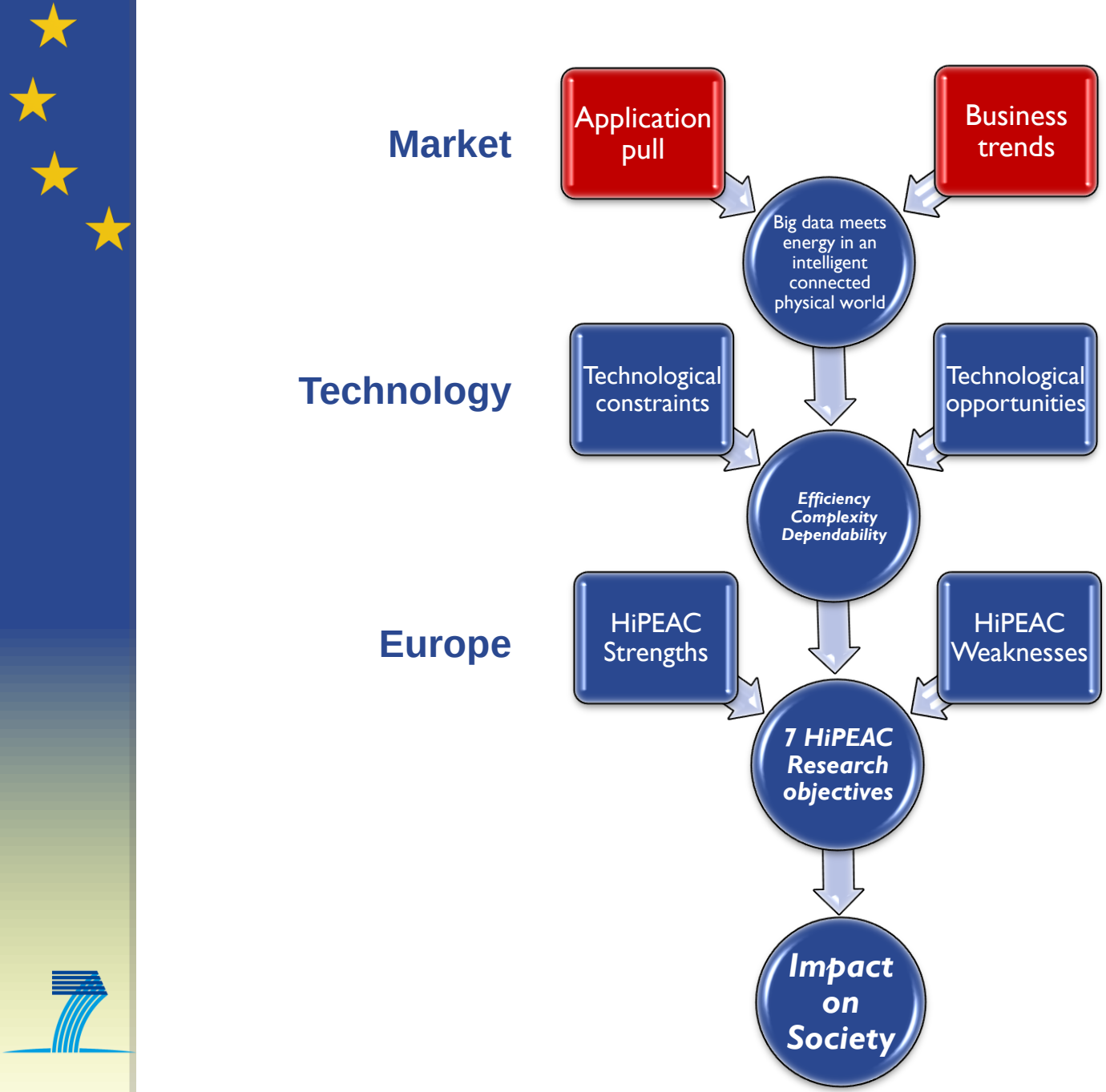
Contributors

- Angelos Bilas (FORTH), Pierre Boulet (INRIA), Albert Cohen (INRIA), Philippe Coussy (CNRS), Raphaël David (CEA), Bjorn De Sutter (Ghent University), Pedro Diniz (inesc-id), Babak Falsafi (EPFL), Paolo Faraboschi (HP), Christian Gamrat (CEA), Georgi Gaydadjiev (TU Delft), Timothy M Jones (University of Cambridge), Manolis Katevenis (FORTH), Stefanos Kaxiras (Uppsala University), Jonas Maebe (Ghent University), Nacho Navarro (UPC), Dimitris Nikolopoulos (FORTH), Alex Ramirez (BSC), Per Stenström (Chalmers), Dirk Stroobandt (Ghent University), Olivier Temam (INRIA), Sid Touati (INRIA), Mateo Valero (BSC), Ayal Zaks (IBM).
- the teachers of the ACACES summer school 2010 and 2011,
- the HiPEAC members,
- Panos Tsarchopoulos, project officer of the HiPEAC network of Excellence.

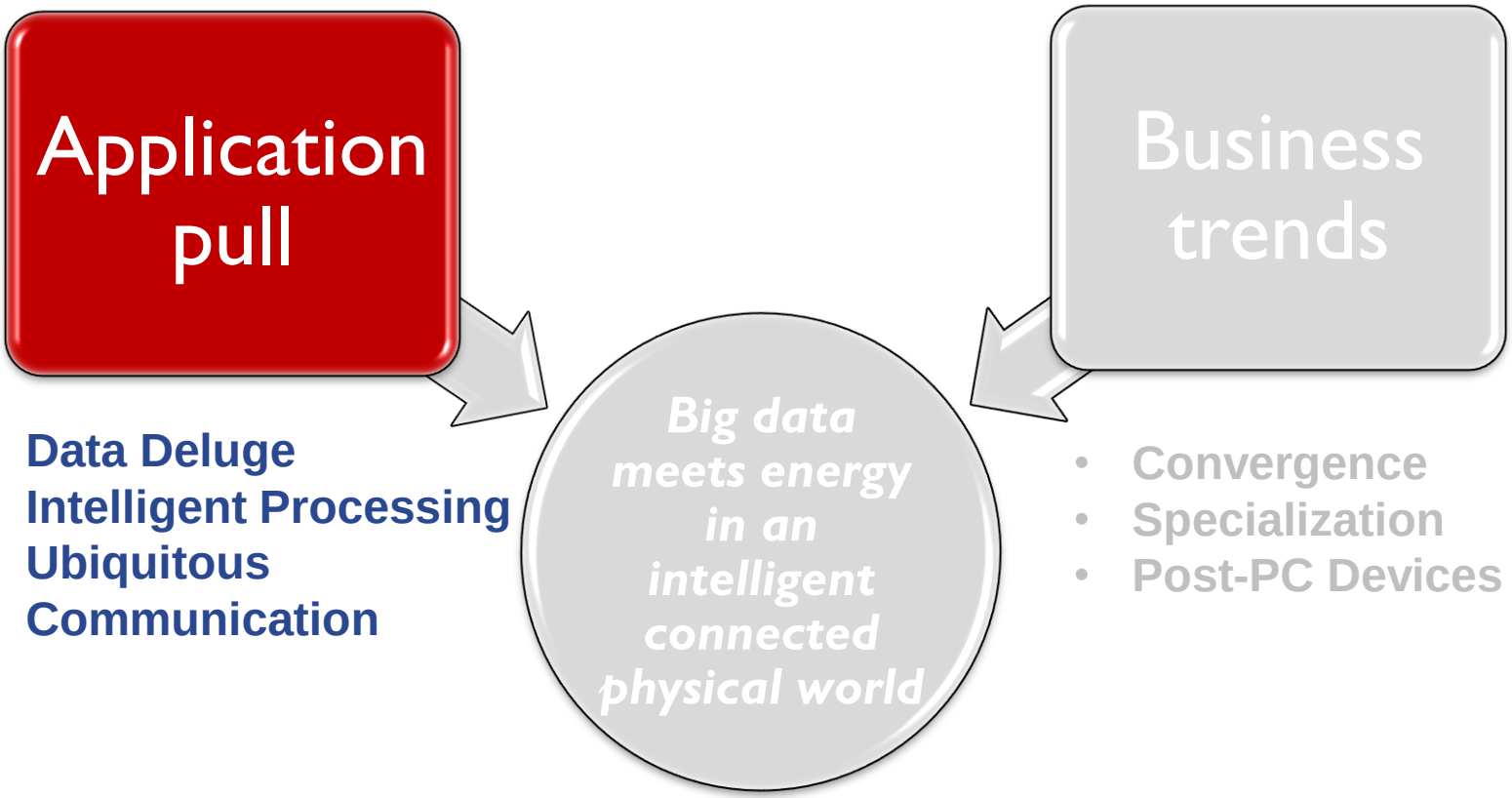

Market
Technology
Europe
Application pull
Business trends
Big data meets energy in an intelligent connected physical world
Technological constraints
Technological opportunities
Efficiency
Complexity
Dependability
HiPEAC Strengths
HiPEAC Weaknesses
7 HiPEAC Research objectives
Impact on Society

Impact on society




Market
Technology
Europe
Application pull
Business trends
Big data meets energy in an intelligent connected physical world
Technological constraints
Technological opportunities
Efficiency
Complexity
Dependability
HiPEAC Strengths
HiPEAC Weaknesses
7 HiPEAC Research objectives
Impact on Society

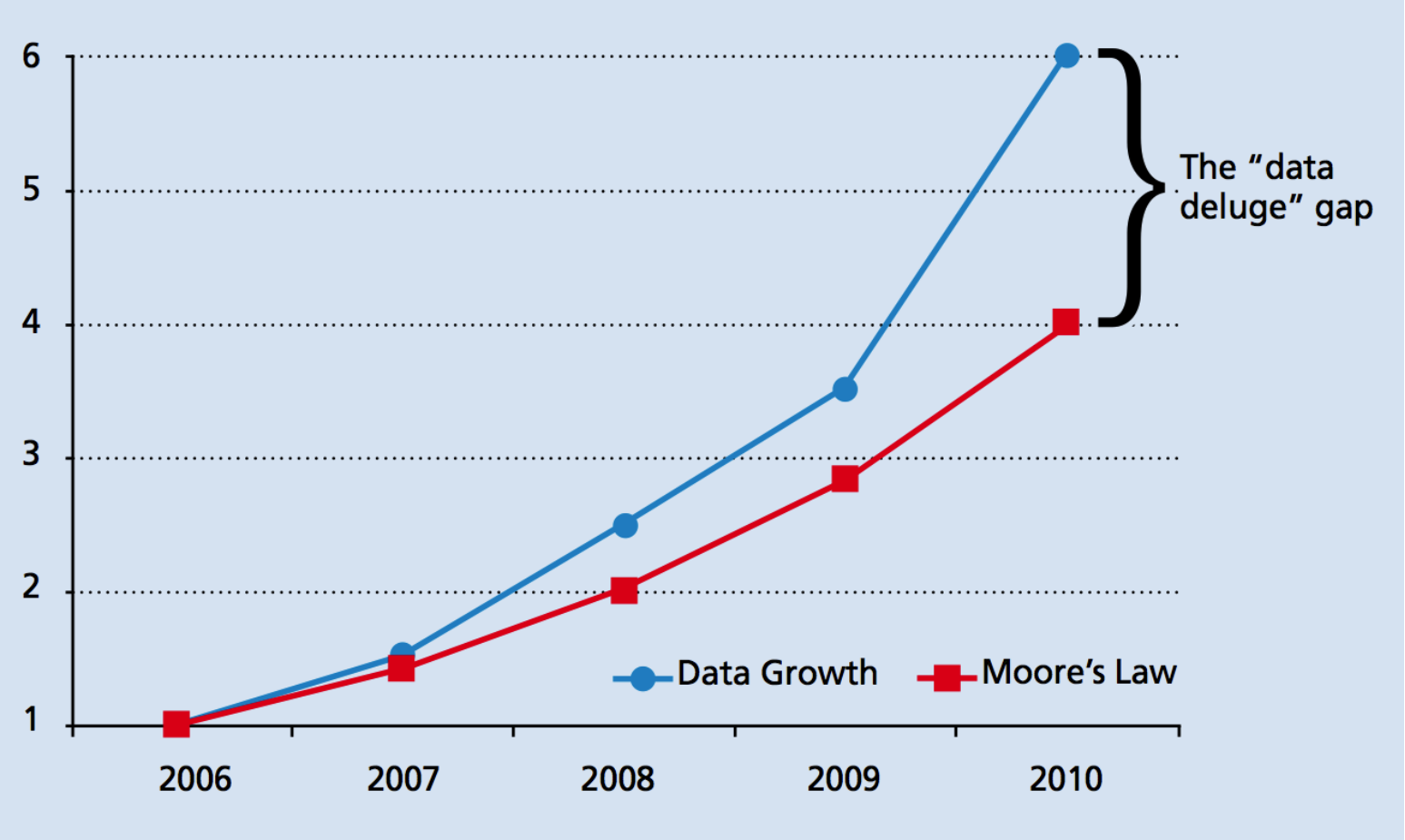
Market forces



- **Data Deluge**
- **Intelligent Processing**
- **Ubiquitous Communication**

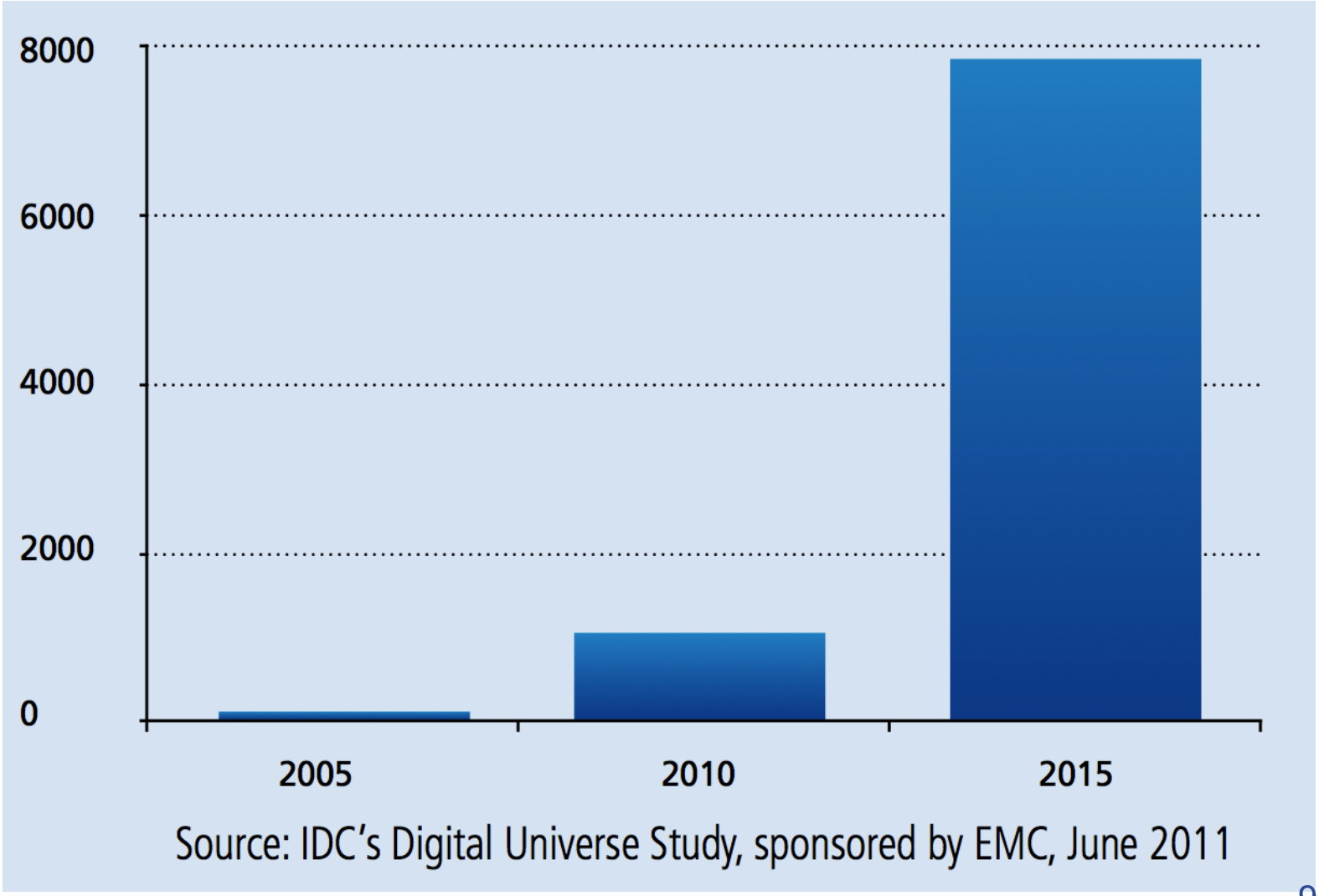
- **Convergence**
- **Specialization**
- **Post-PC Devices**

Data Deluge



[Paolo Faraboschi, HP]

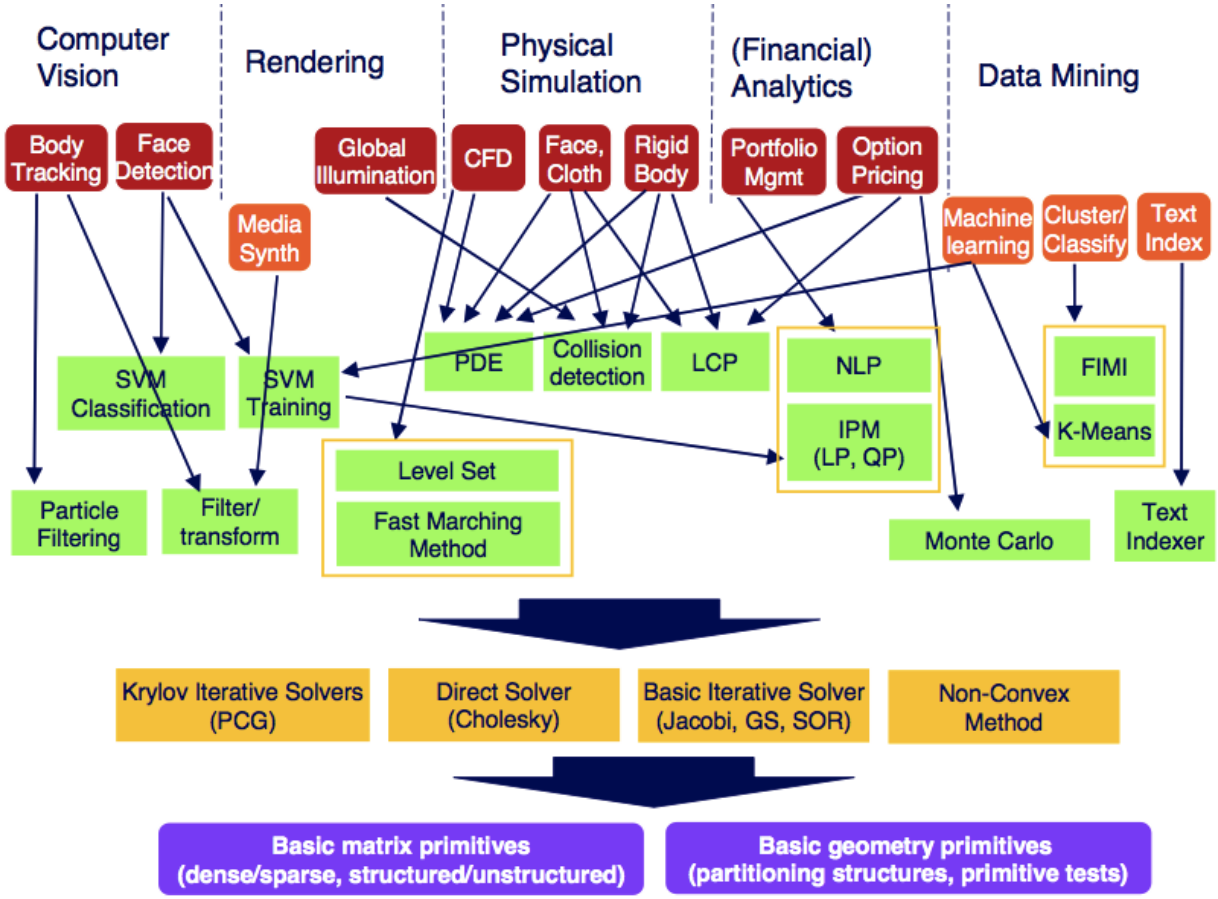
Growth of data storage in Exabytes



Intelligent processing of "natural" data

More and more applications are not only "number crunching"

Recognition, Mining, Synthesis



Intel's RMS and how it maps down to functions that are more primitive. Of the five categories at the top of the figure, Computer Vision is classified as Recognition, Data Mining is Mining, and Rendering, Physical Simulation, and Financial Analytics are Synthesis. [Chen 2006]

Krste Asanovic et al, "The Landscape of Parallel Computing Research: A View from Berkeley"

SMARTER THAN YOU THINK

Google Cars Drive Themselves, in Traffic

Ramin Rahimian for The New York Times

Dmitri Dolgov, a Google engineer, in a self-driving car parked in Silicon Valley after a road test.

Ubiquitous computing in a connected world

Infrastructure
Core (cloud)

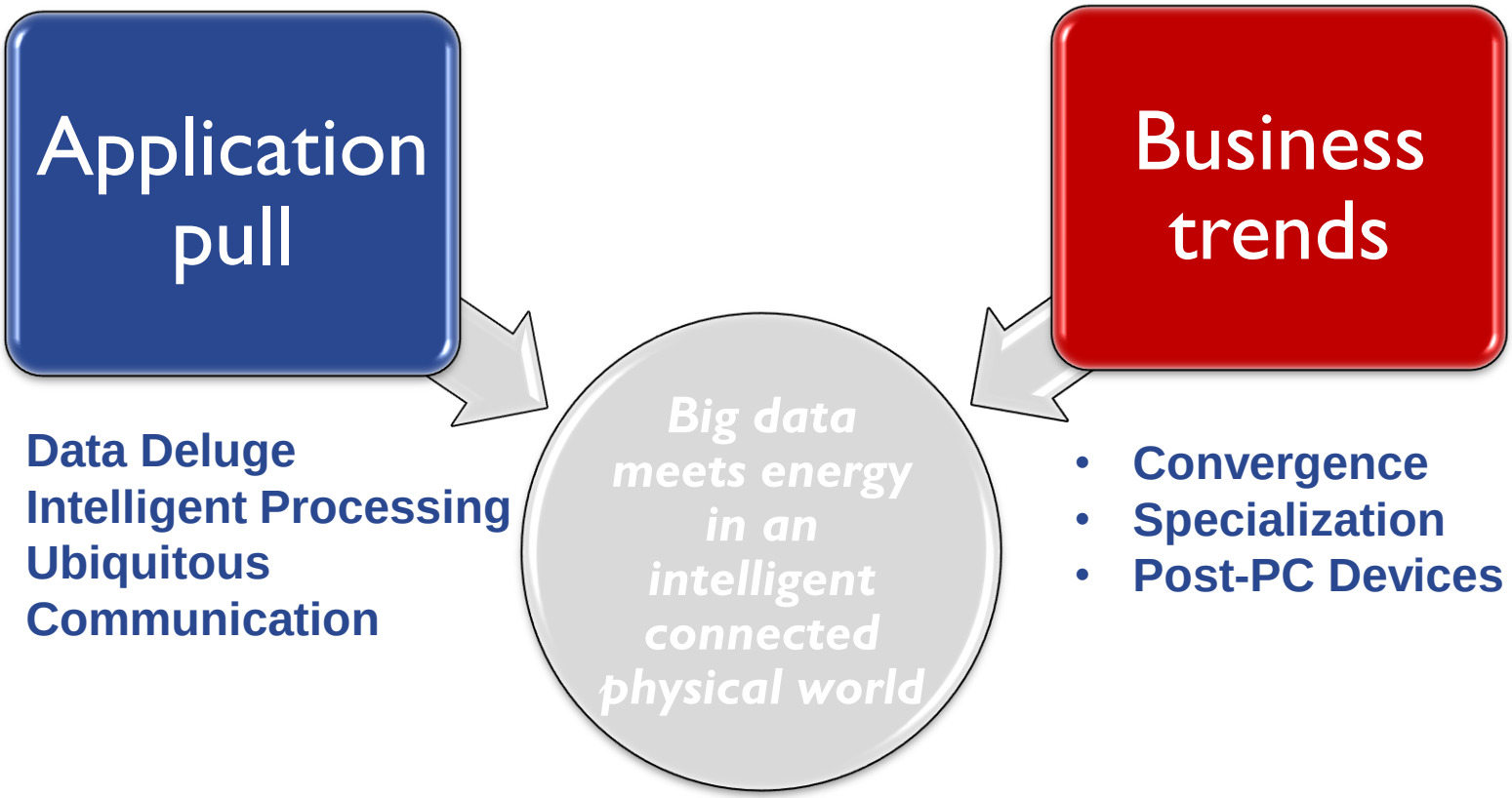
Sensory swarm, actuators and real world data

Smart house cities, …

Mobile access

Courtesy Jan M. Rabaey, UC Berkeley, updated for this HiPEAC vision

Market forces



- Data Deluge
- Intelligent Processing
- Ubiquitous Communication

- Convergence
- Specialization
- Post-PC Devices

Convergence

Broadcast

IP, Internet

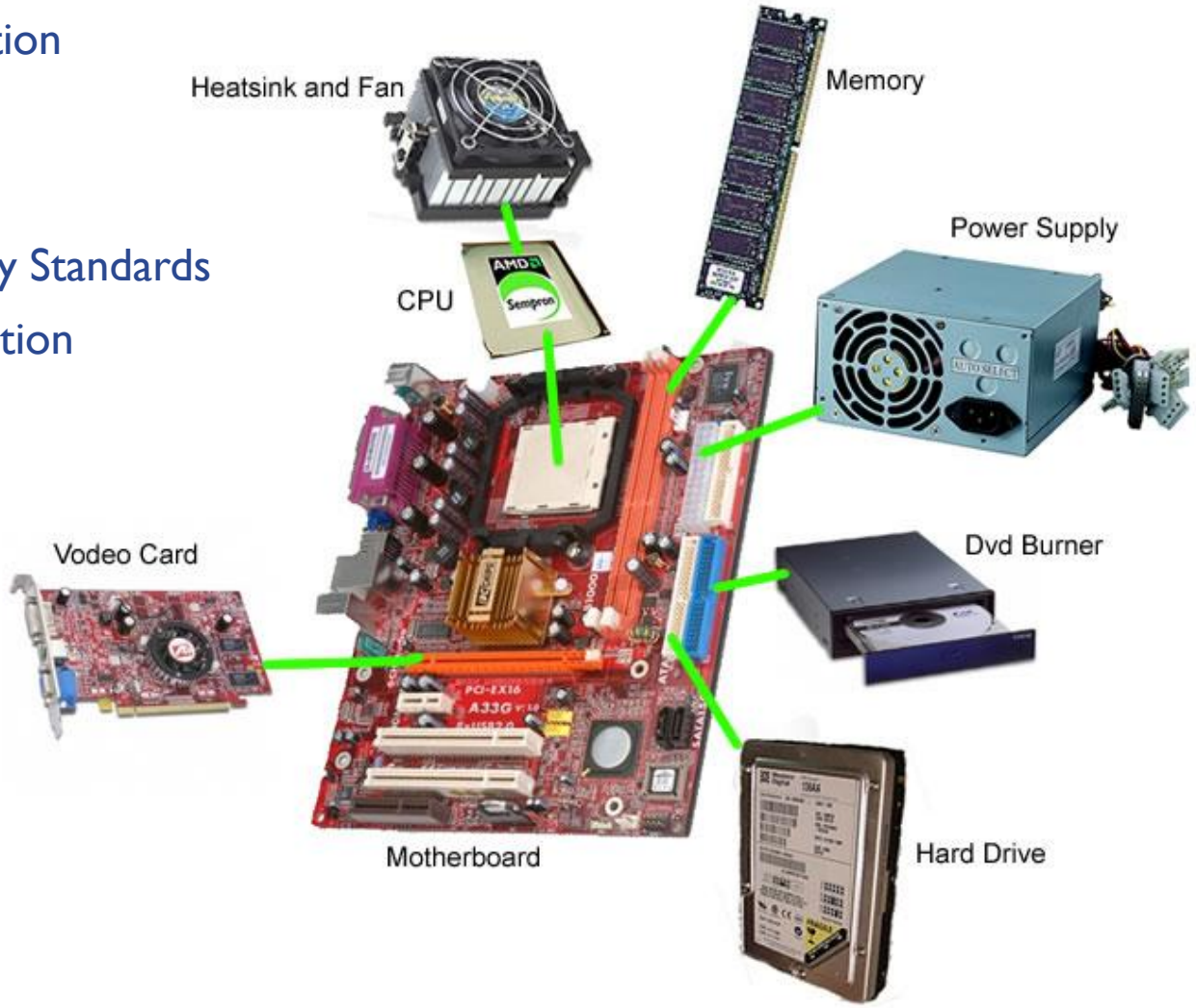
- Business models
- Standard(s)
- Interoperability
- …

Telecom

MacBook image © Jared C. Benedict
Phone, TV images © LG Electronics

Industry specialization

- De-verticalization
- Sharing
- Ecosystem
- Interoperability Standards
- Local optimization

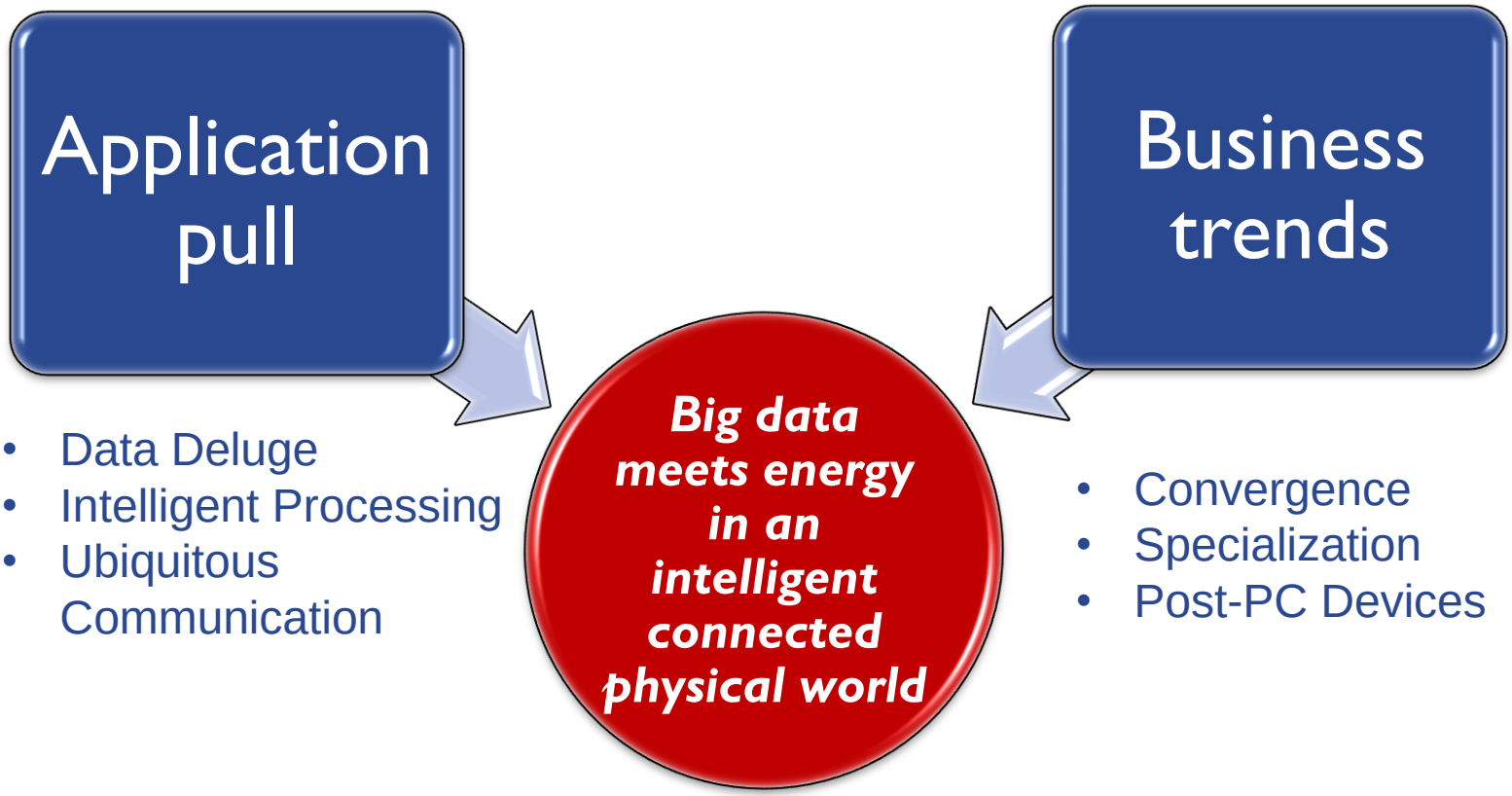


Post-PC devices

Ubiquitous access

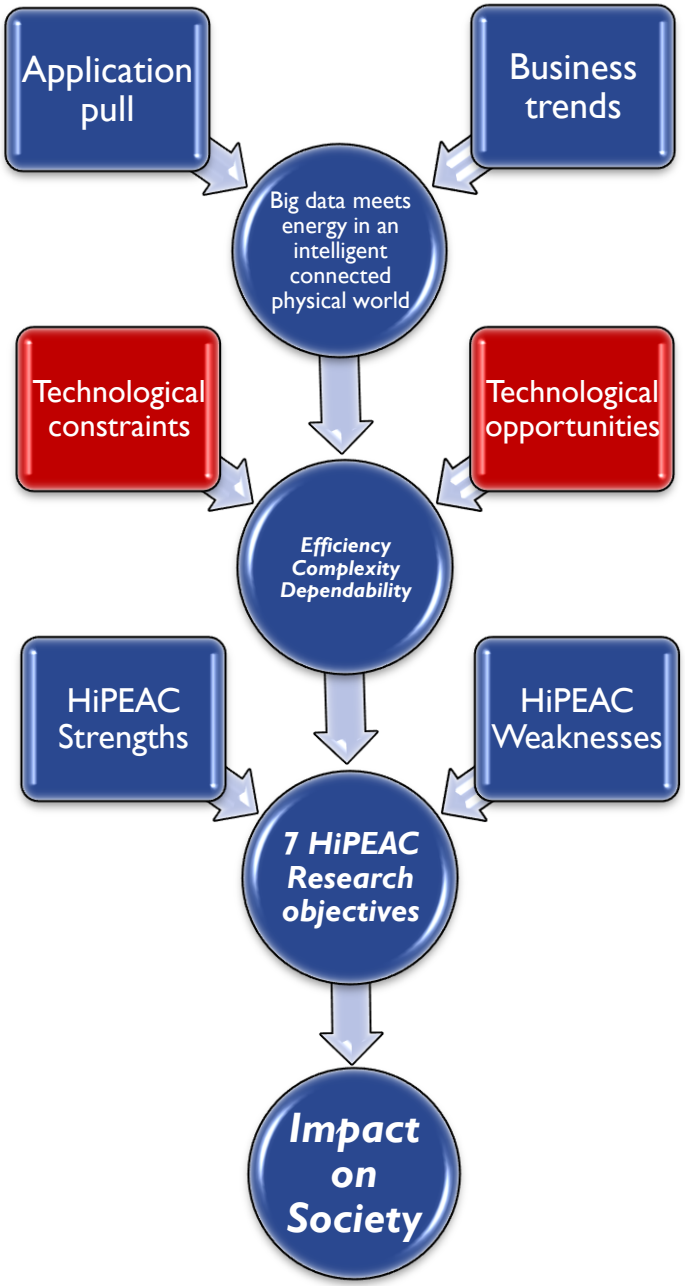
iPad image © Apple, Inc
MP3 player image © J A S P E R@flickr
iPhone image © K!T@flickr

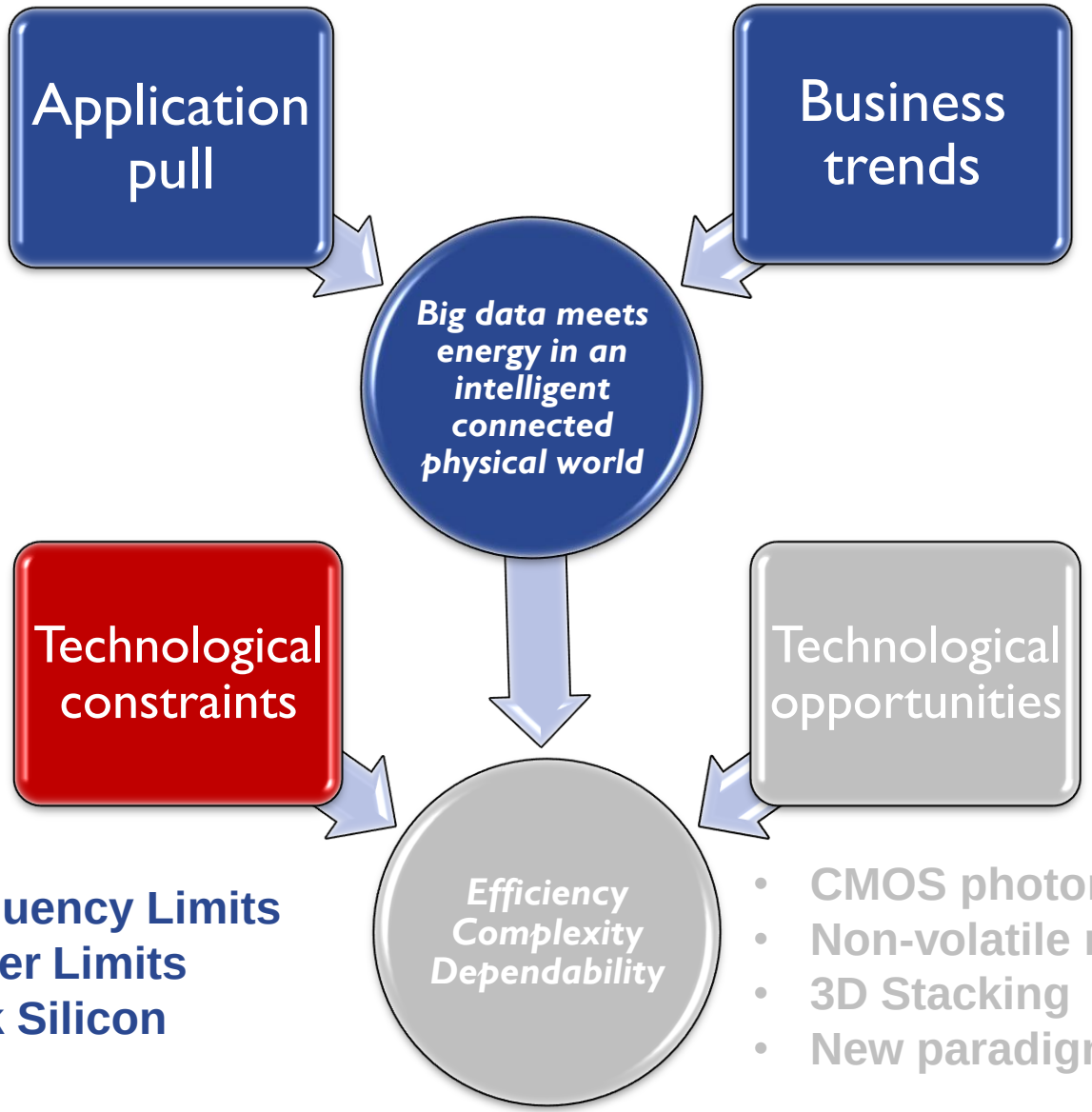
Market forces



- Data Deluge
- Intelligent Processing
- Ubiquitous Communication

- Convergence
- Specialization
- Post-PC Devices


Application pull
Business trends
Big data meets energy in an intelligent connected physical world
Technological constraints
Technological opportunities
Efficiency
Complexity
Dependability
HiPEAC Strengths
HiPEAC Weaknesses
7 HiPEAC Research objectives
Impact on Society

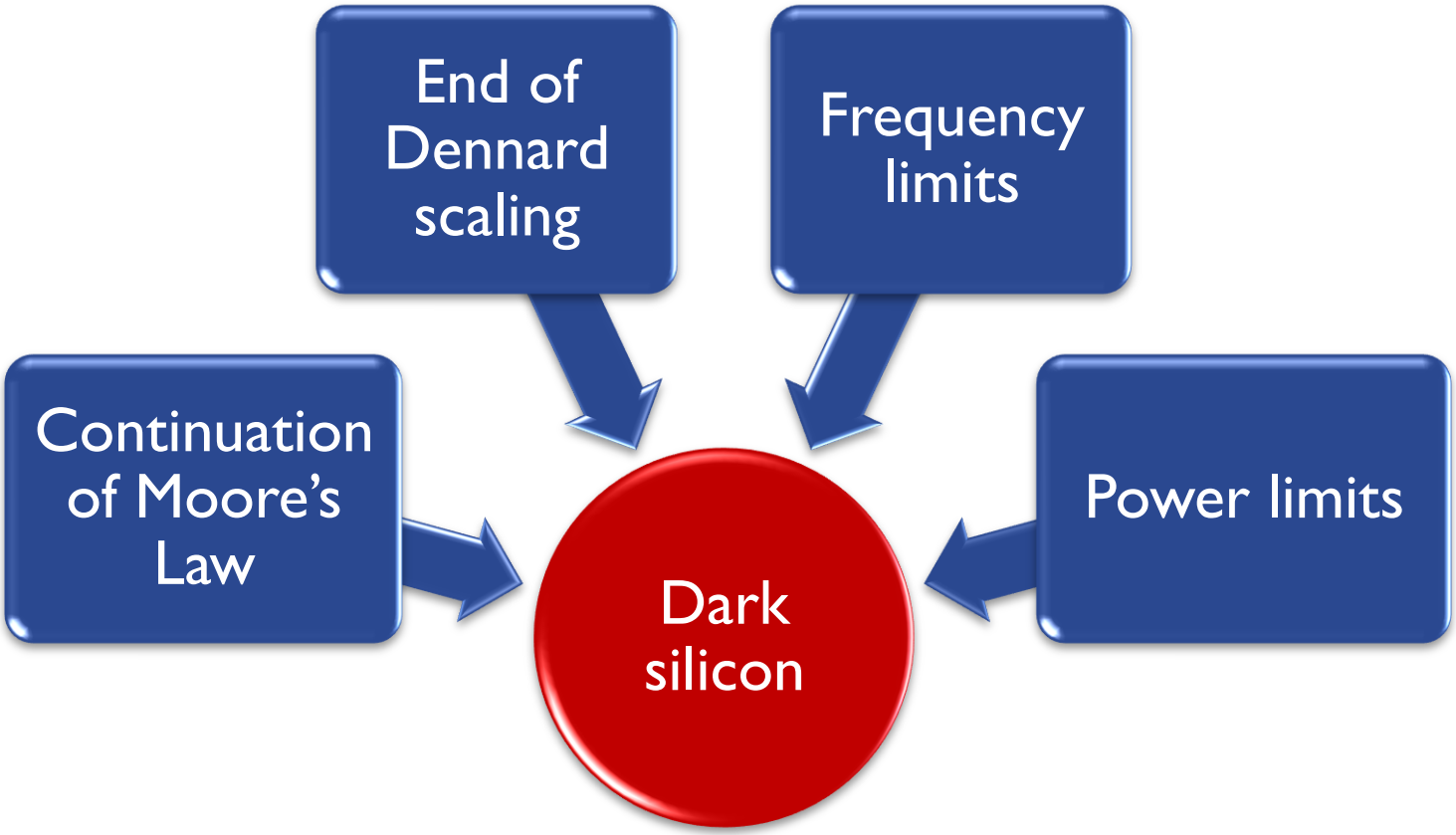


- **Frequency Limits**
- **Power Limits**
- **Dark Silicon**

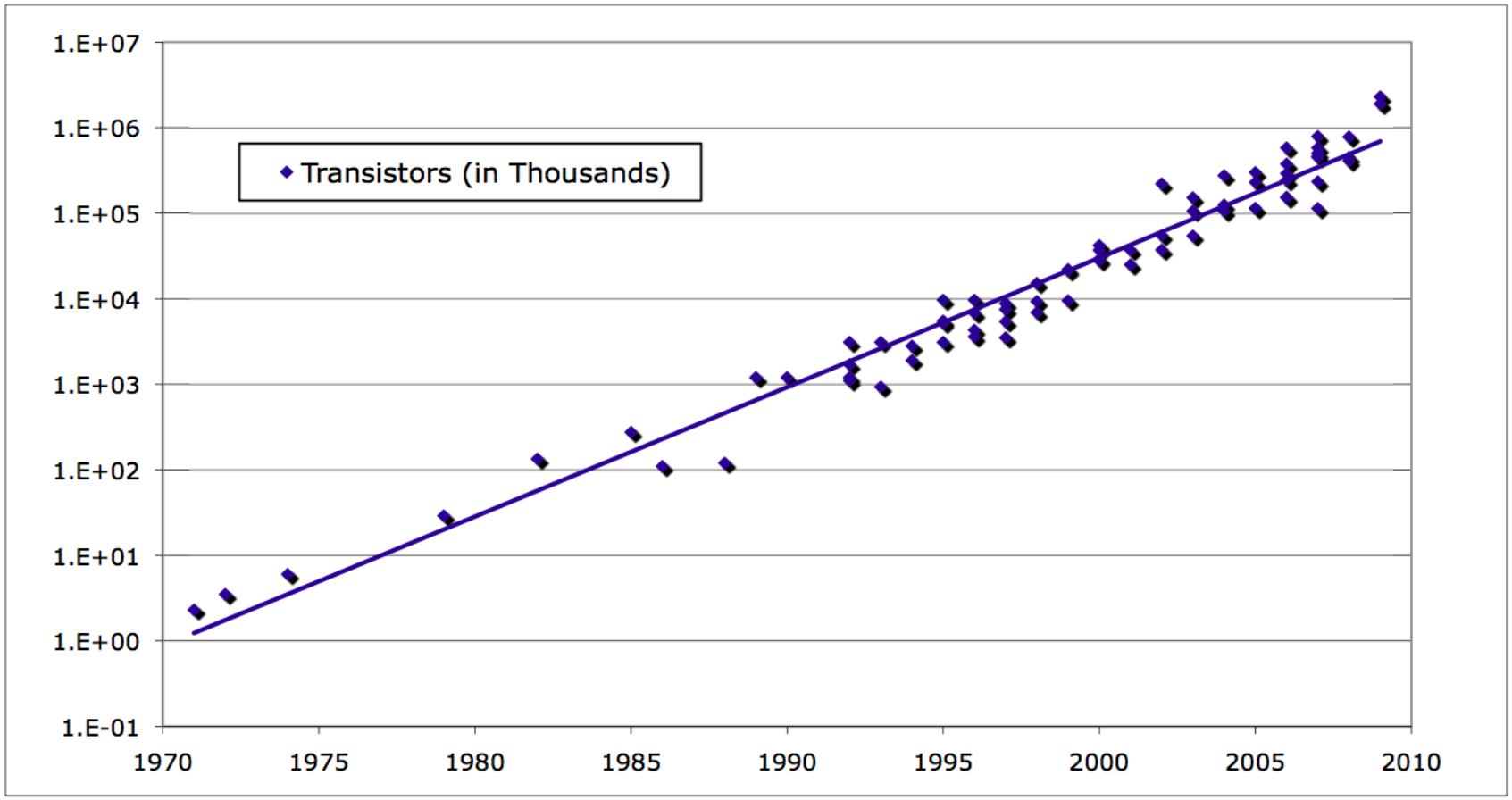
- **CMOS photonics**
- **Non-volatile memories**
- **3D Stacking**
- **New paradigms**

Technological constraints

We are at a turning point



Moore's law: increase in transistor density



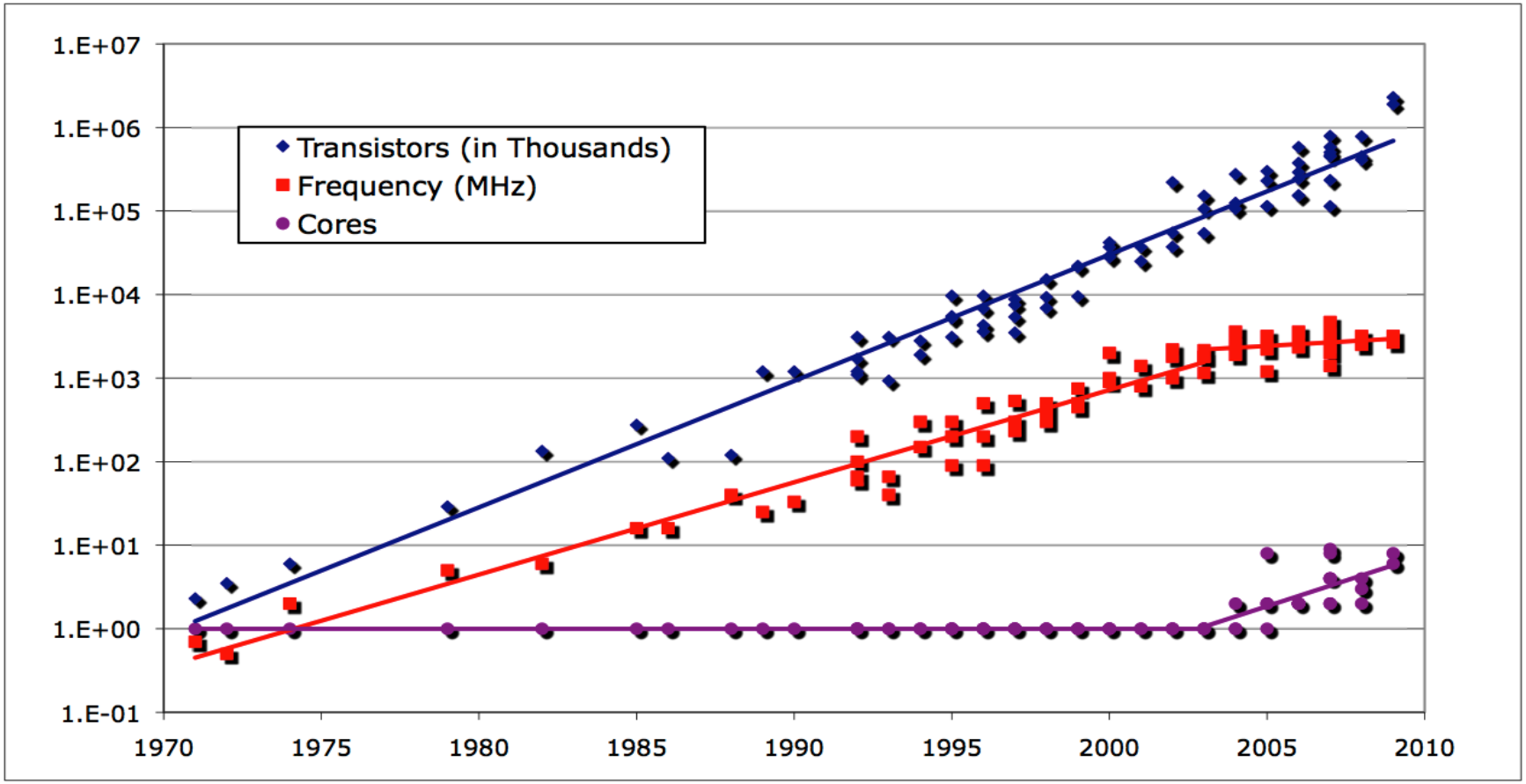
Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović

The end of Dennard Scaling

Parameter (scale factor = a)	Classic Scaling	Current Scaling
Dimensions	1/a	1/a
Voltage	1/a	1
Current	1/a	1/a
Capacitance	1/a	>1/a
Power/Circuit	$1/a^2$	1/a
Power Density	1	a
Delay/Circuit	1/a	~1

Source: Krisztián Flautner "From niche to mainstream: can critical systems make the transition?"

Limited frequency increase ⇒ more cores



Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic¸

Locality and communications management

- In 22 nm, swapping 1 bit in a transistor has an energy cost:

  ~ 1 attojoule (10^{-18} J)

- Moving a 1-bit data on the silicon costs:

  ~1 picojoule/mm (10^{-12} J/mm)

- Moving a data 10^9 per second (1 GHz) in silicon has a cost:

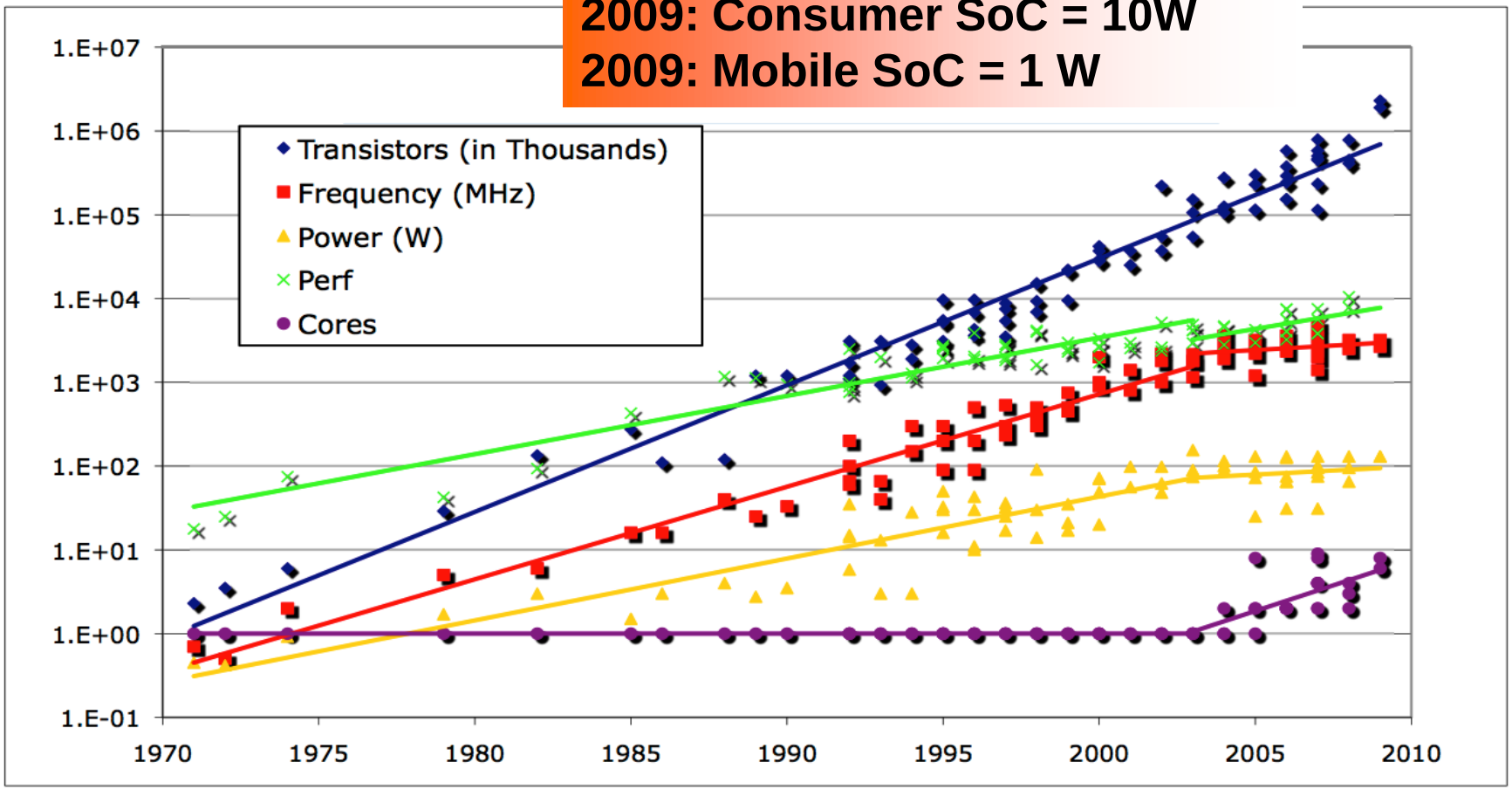
  1 pJ/mm x 10^9 s^{-1} = ~1 milliwatt/mm

- 64 bit bus @ 1 GHz: ~64 milliwatts/mm (with 100% activity)
- For 1 cm of 64 bit bus @ 1 GHz : 0,64 W/cm

- On modern chips, there are about several km of wires on chip, even with low toggle rate, this leads to several Watt/cm^2

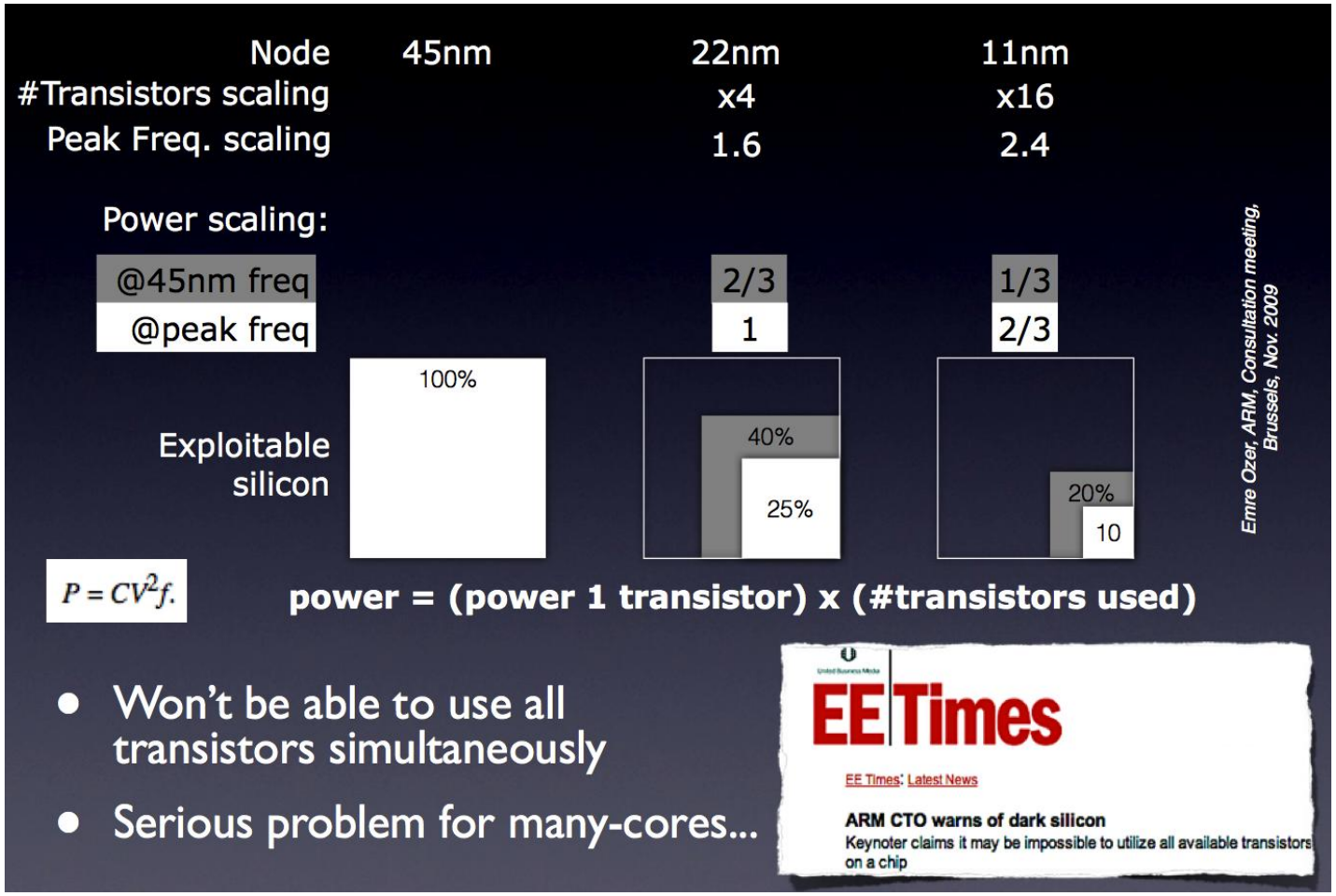
Limitation by power density and dissipation

2009: GP CPU = 130 W (45 nm)
2009: Consumer SoC = 10W
2009: Mobile SoC = 1 W



Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović

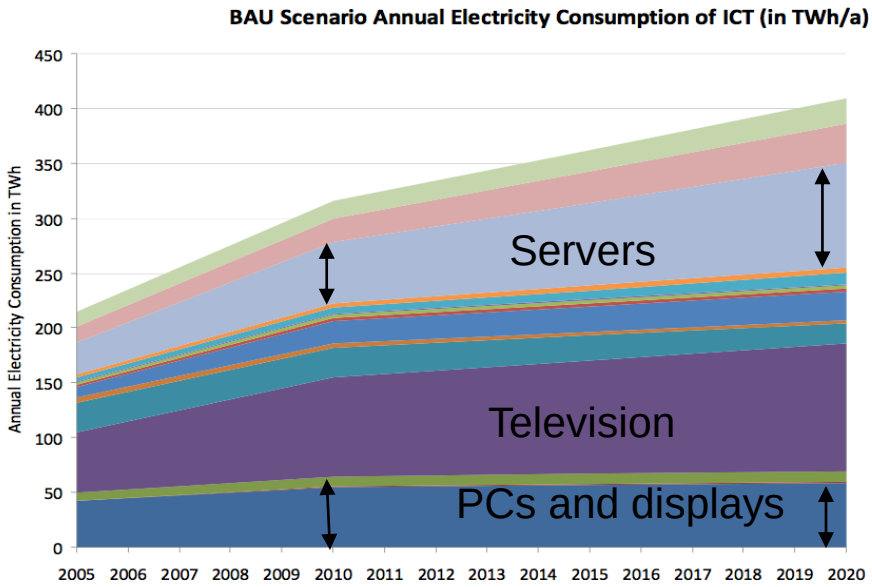
Dark Silicon



Source: Krisztián Flautner "From niche to mainstream: can critical systems make the transition?"

Energy consumption of ICT

Forecast for all ICT domain by the European Commission



"Business as Usual"

300 TWh in 2010

with 70 TWh for the ~10 million Servers

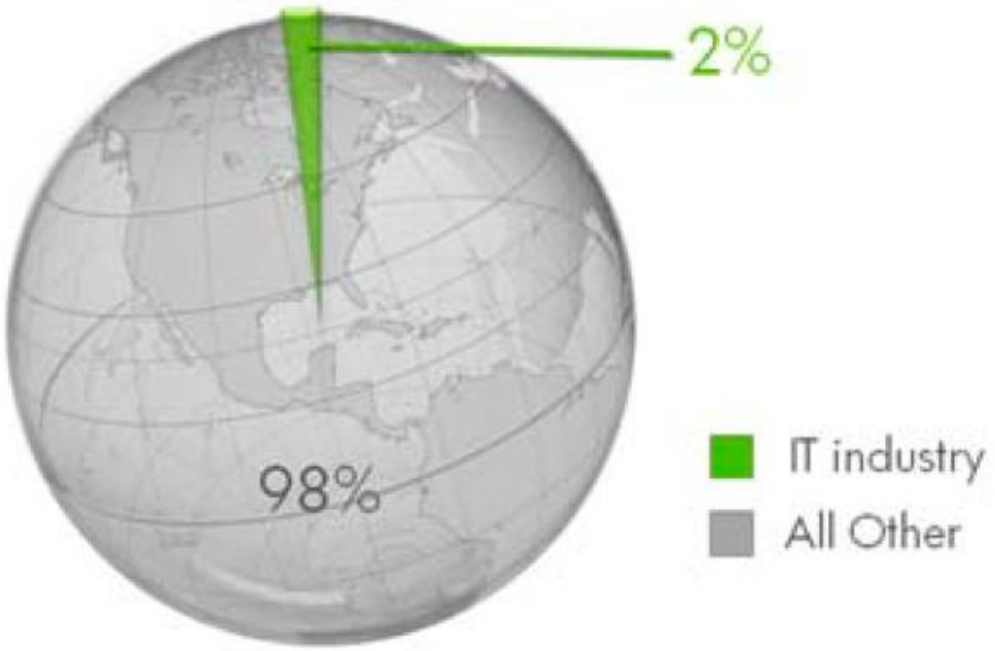
409 TWh in 2020

with100 TWh for Servers

Note: 1 Nuclear Power Plant ~10 TWh

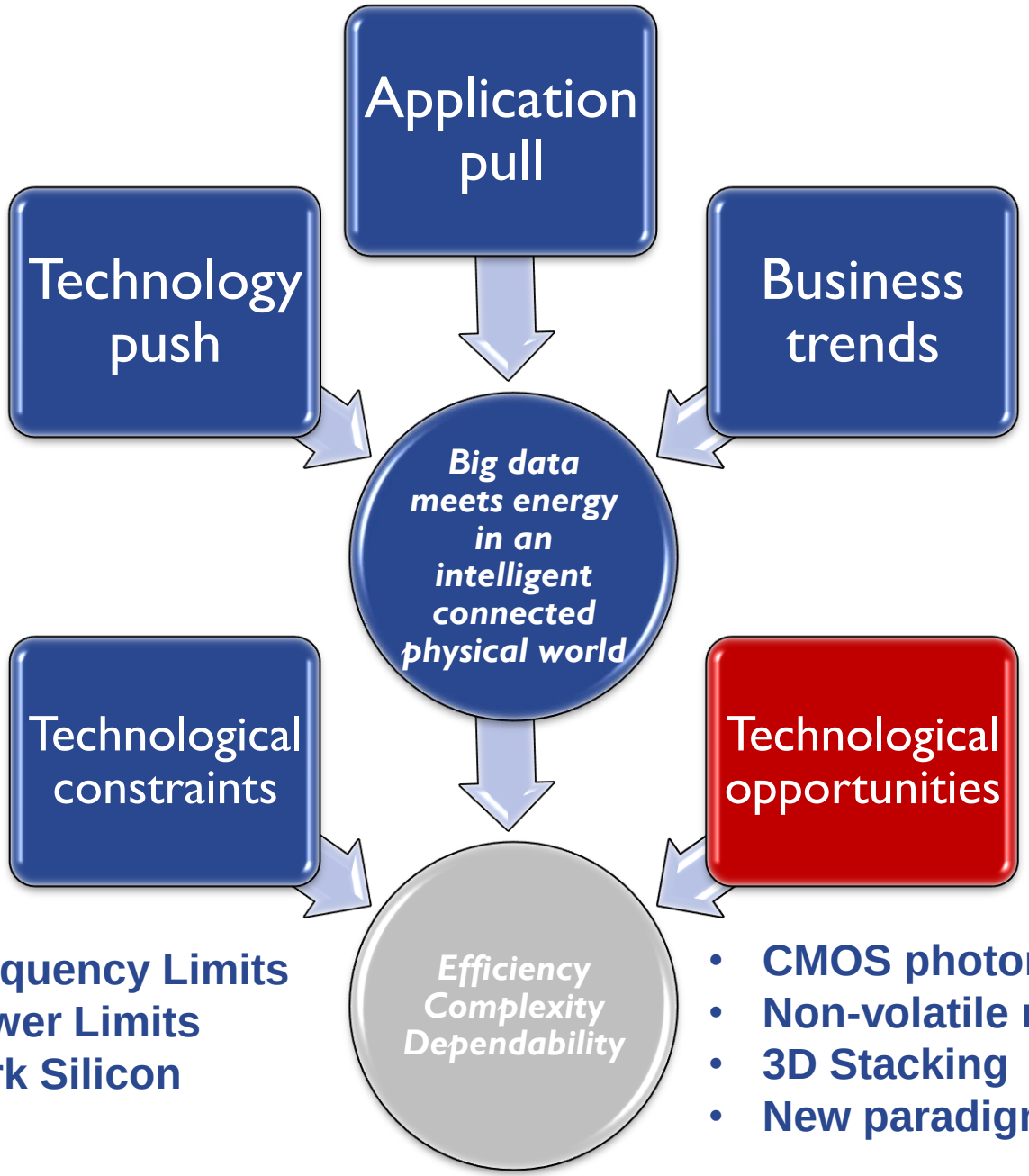
Source: European Commission DG INFSO, Impact of Information and Communication Technologies on Energy Efficiency, final report, 2008
ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/sustainable-growth/ict4ee-final-report_en.pdf

ICT: 2% of global carbon emissions



Source: P. Ranganathan, "Saving the world together, one server at a time…"

ACACES 2011



- **Frequency Limits**
- **Power Limits**
- **Dark Silicon**

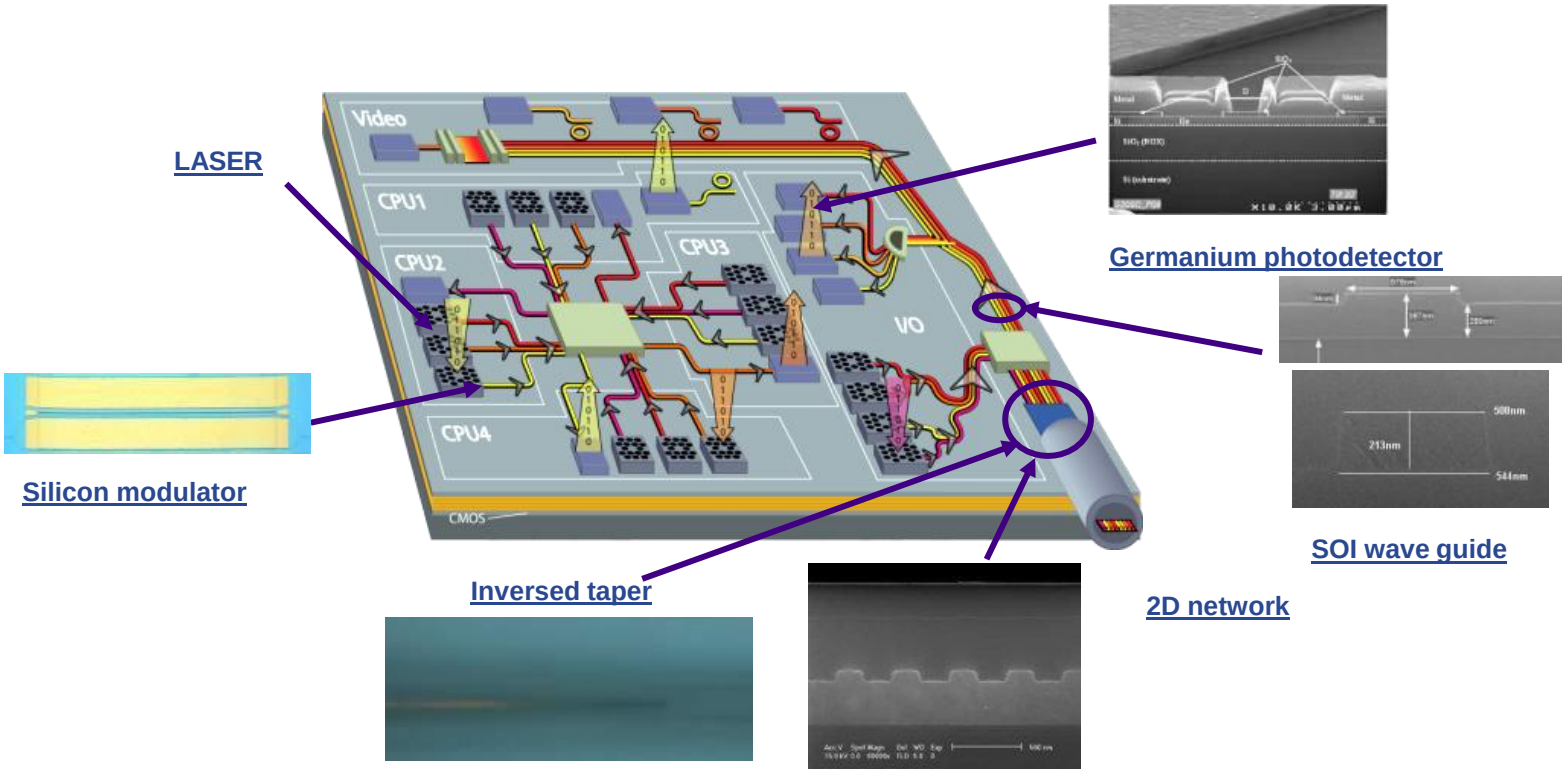
- **CMOS photonics**
- **Non-volatile memories**
- **3D Stacking**
- **New paradigms**

Optical interconnects

CMOS photonics is the integration of a photonic layer with an electronic circuit.

Advantages of CMOS photonics are:

- Use of standard tools and foundry, wafer scale co-integration
- **Lower energy** (~100 fJ/bit), (wire: ~1 pJ/mm)
- High bandwidth (10 Gbps), Low latency (~10 ps/mm)

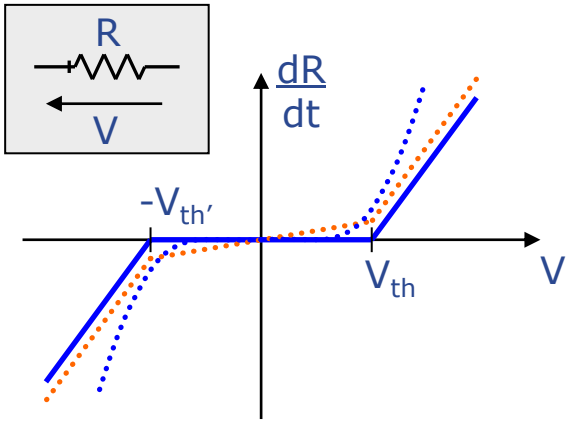


Source: CEA, Ahmed Jerraya

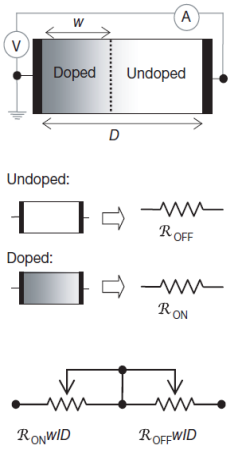
Non-volatile memories….

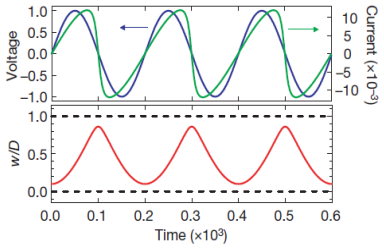
Example: Memristive Devices Principle

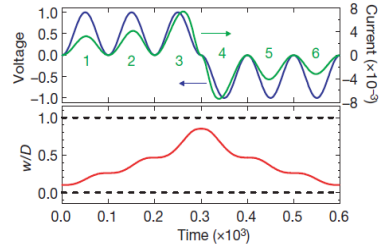
$$v = R(x,i).i \qquad \frac{dx}{dt} = f(x,i)$$

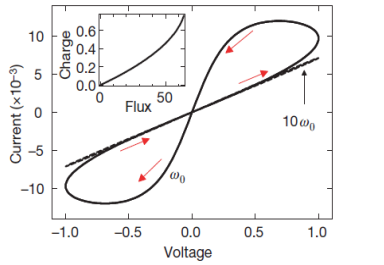


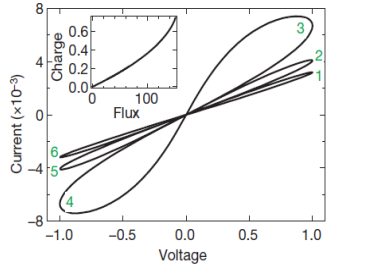
Nonlinear characteristic

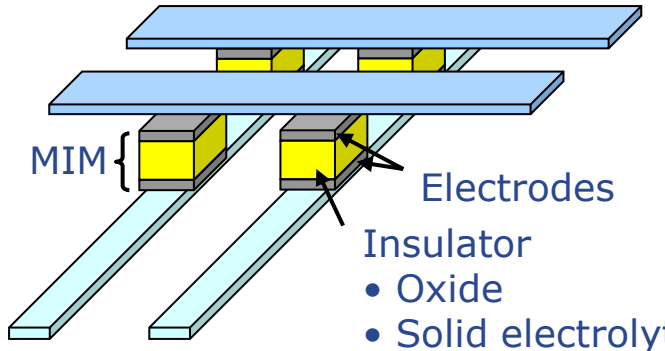


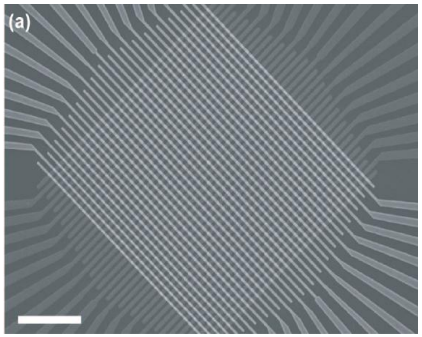










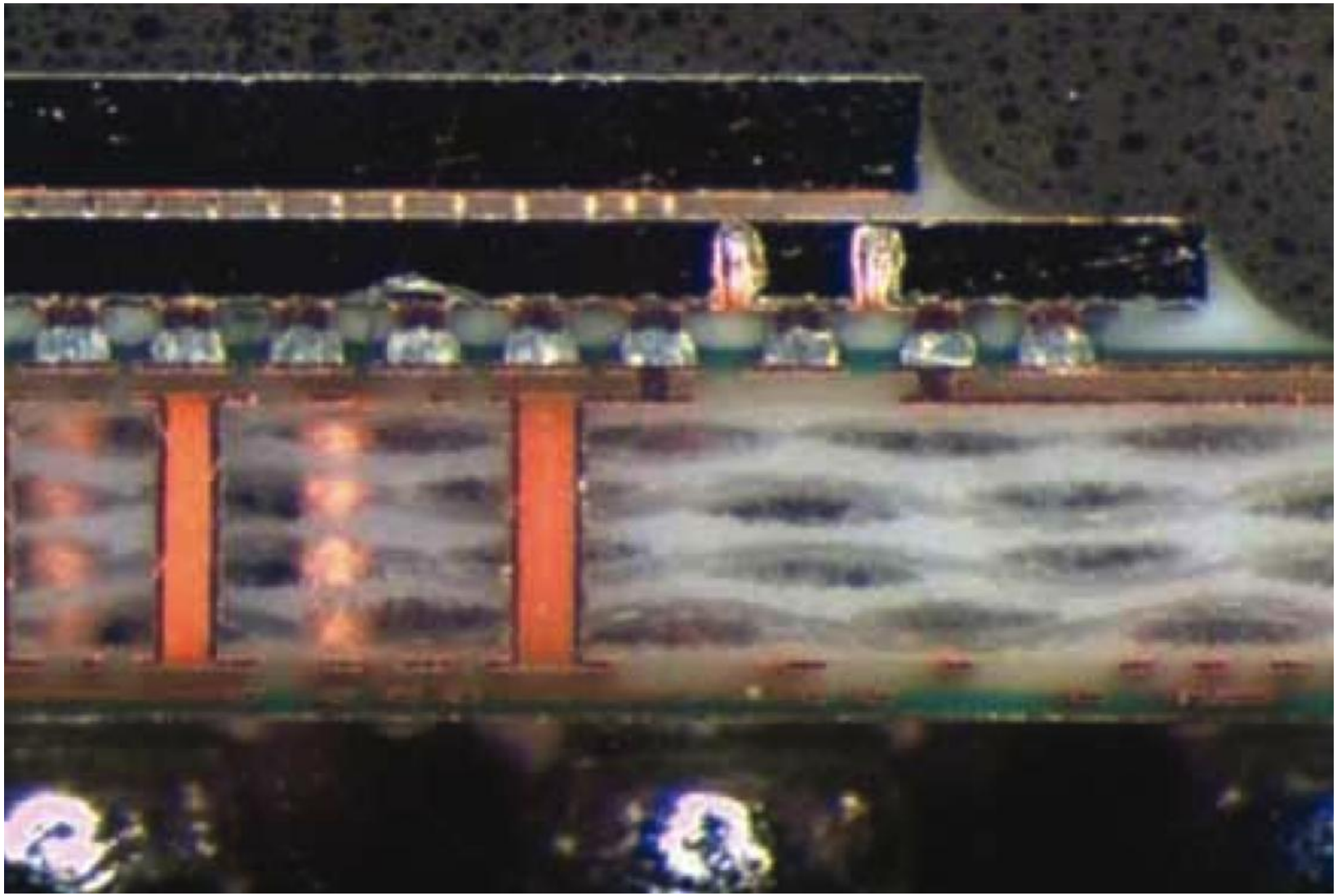


Crossbar
(University of Michigan)

[1] L. Chua and S. Kang, *Proceedings of the IEEE*, 1976
[2] D. Strukov et al., *Nature*, 2008

Source: CEA, C. Gamrat

3D stacking

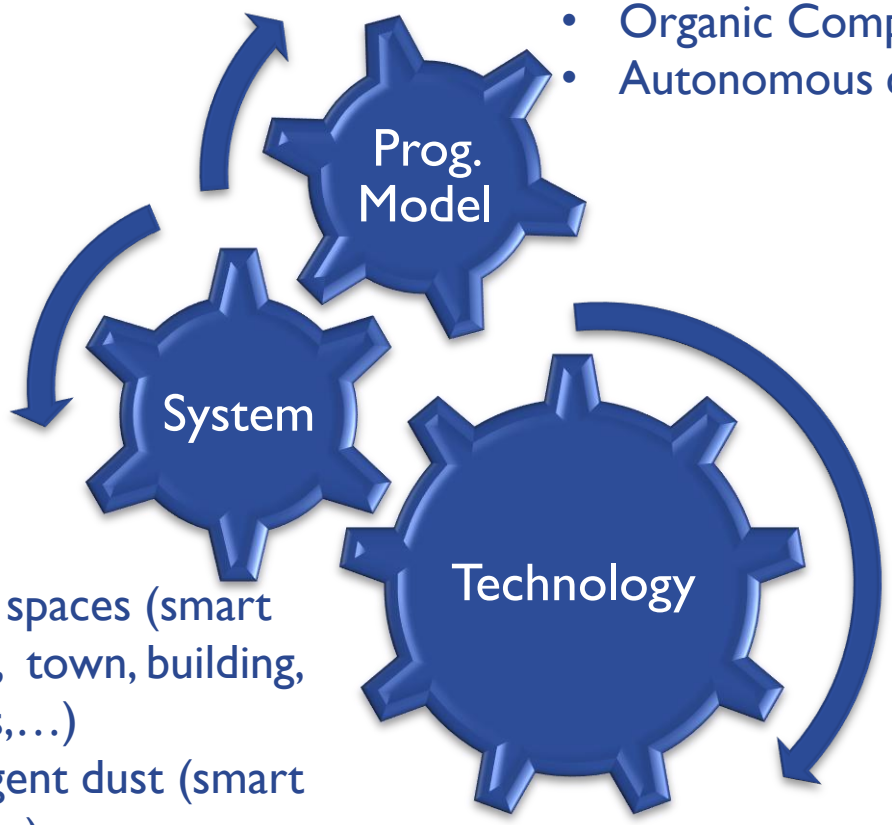


Multiple integration with 3D stacking…

Source: STMicroelectronics & CEA

Technology also drives us to think differently…

- Sub-Symbolic computing
- Stochastic computing
- Biologically inspired computing
- Organic Computing
- Autonomous computing, Self-*



- Smart spaces (smart house, town, building, rooms,…)
- Intelligent dust (smart sensors)

- 3D stacking
- Photonic interconnect
- Non-volatile memories
- Molecular computing
- More-than-Moore
- Spintronics
- Chemical computing
- Biologically inspired cells
- Memristors
- ...
- Also silicon based!

Together…

Electrons for compute

Electrons like to interact; easily moved; interaction needed for compute

+ Ions for storage

Ions like to interact; stay put; good for storage

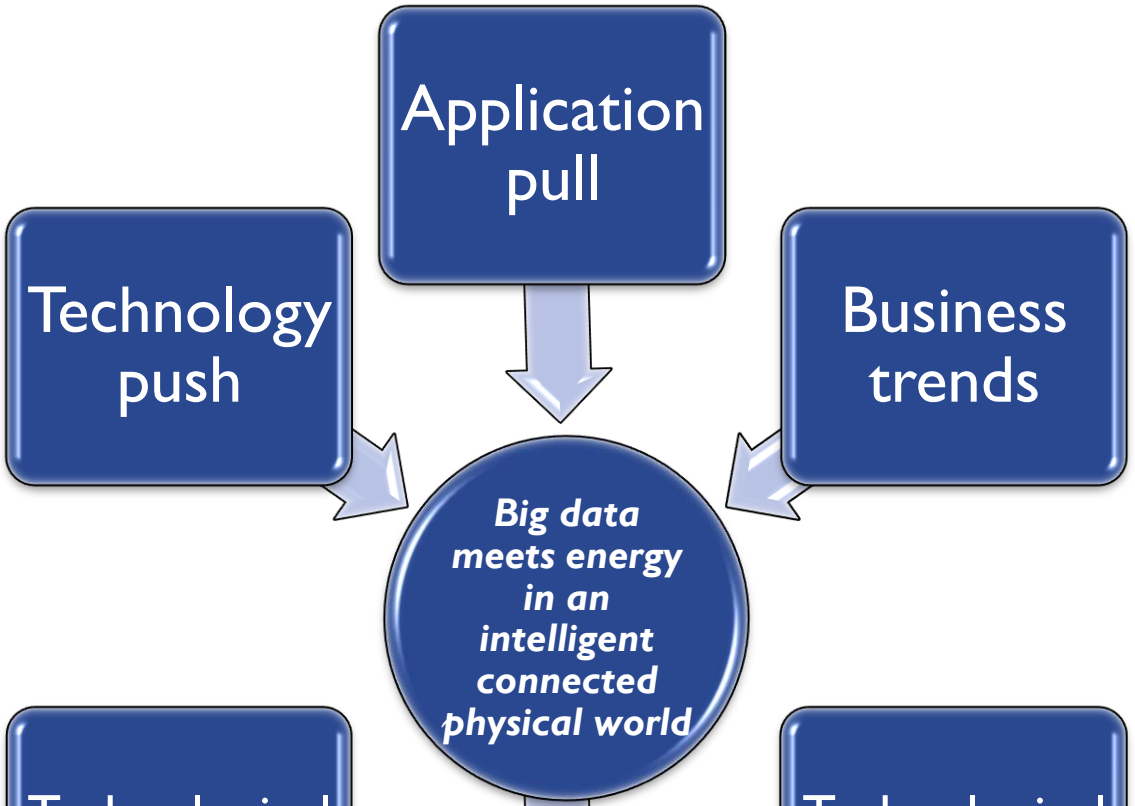
+ Photons to communicate

Photons don't like to interact or stay put; good for long-distances

Courtesy: Jouppi2011

Source: P. Ranganathan, "Saving the world together, one server at a time…"

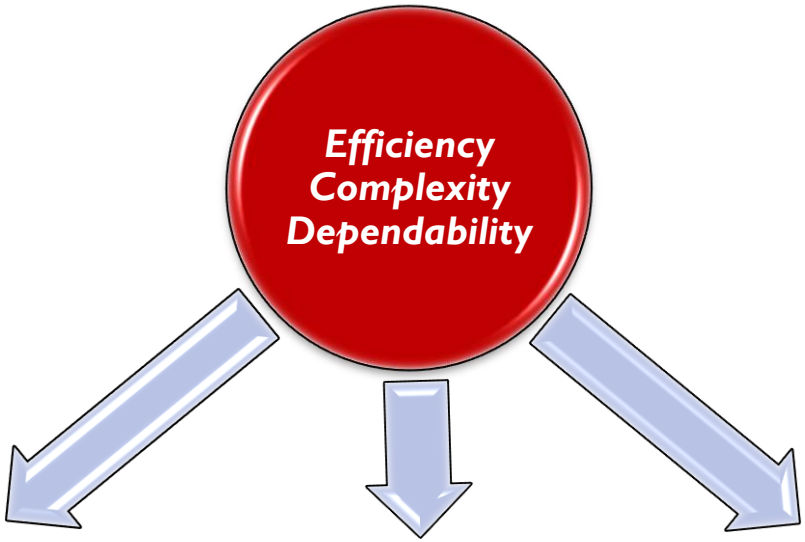
ACACES 2011



- Frequency Limits
- Power Limits
- Dark Silicon

- CMOS photonics
- Non-volatile memories
- 3D Stacking
- New paradigms

Core Computing Systems Challenges



Efficiency	Complexity	Dependability
• Power • Performance	• Heterogeneity • Parallelism	• Reliability • Privacy

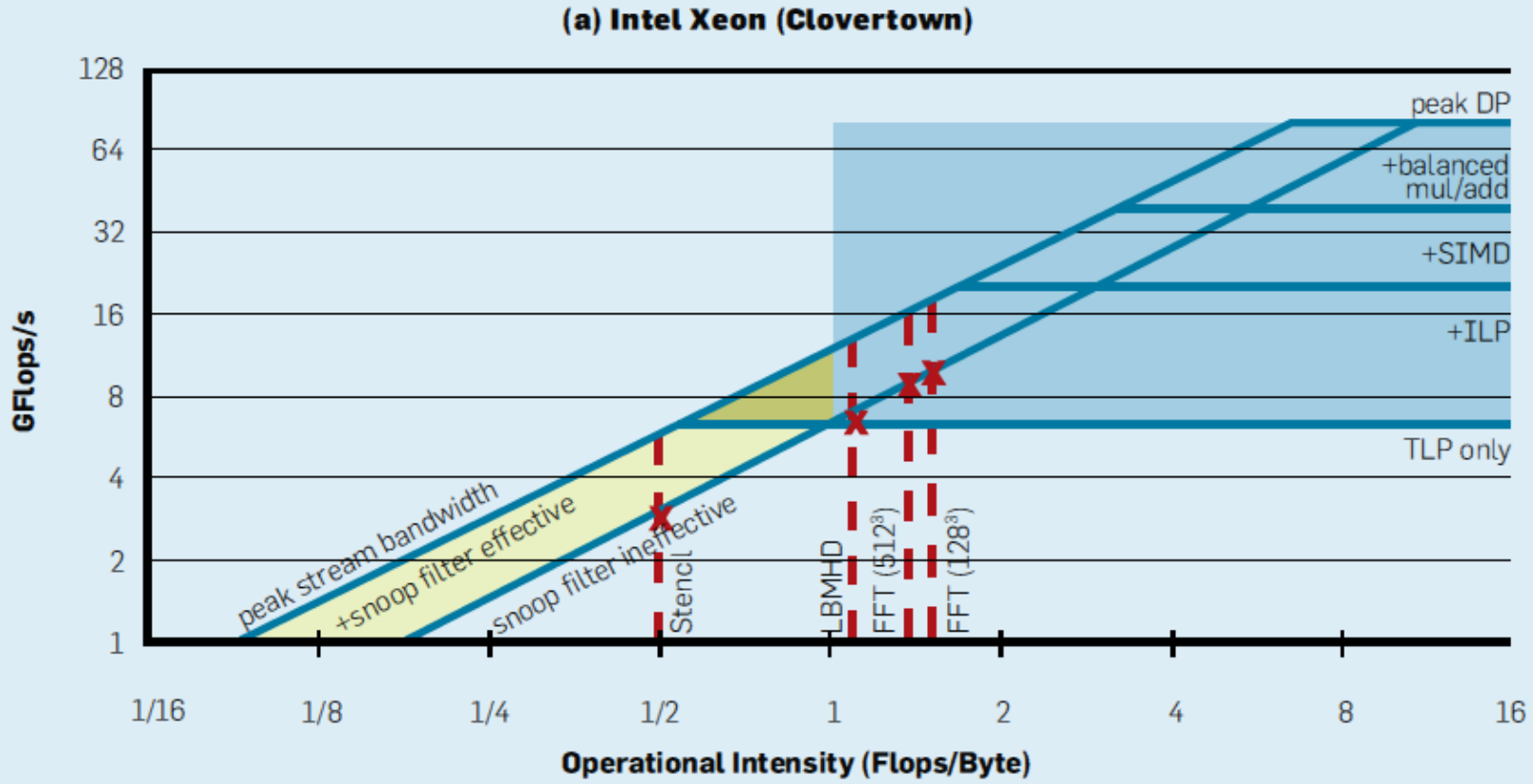
Improving efficiency

Power defines performance

Operations/Watt – Operations/Watt/Euro

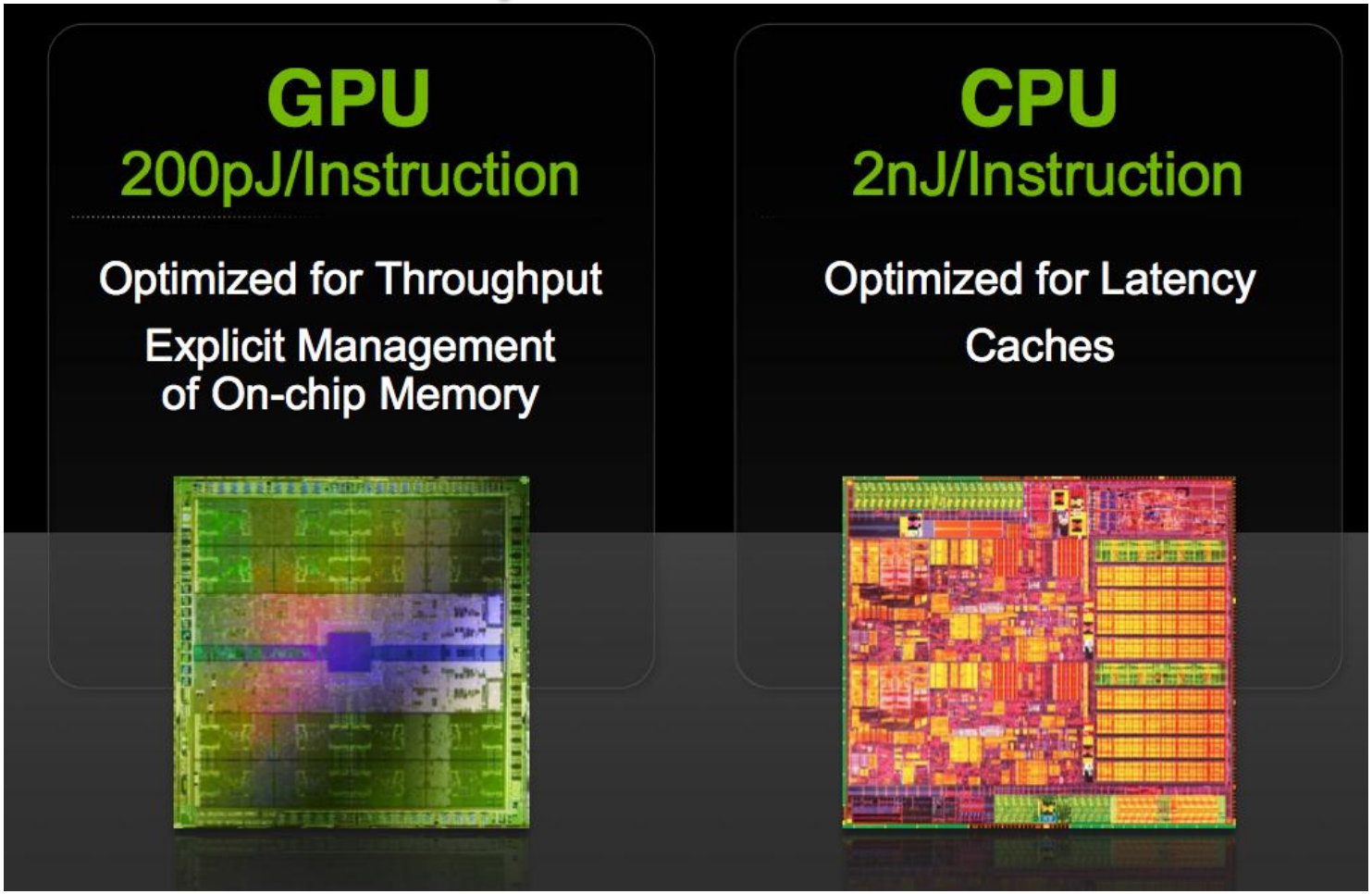
Improving efficiency

Communication defines performance



[Roofline model: Williams, Waterman, Patterson; CACM April 2009]

Parallelism and specialization leads to more efficiency

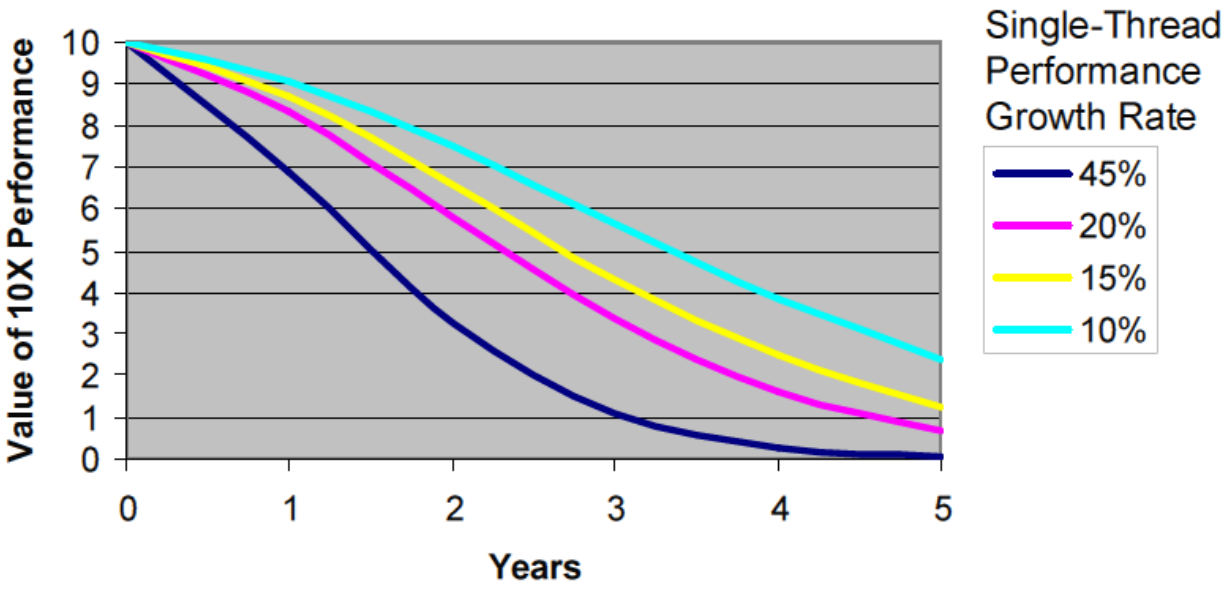


Source: Bill Dally, "To ExaScale and Beyond"

www.nvidia.com/content/PDF/sc_2010/theater/Dally_SC10.pdf

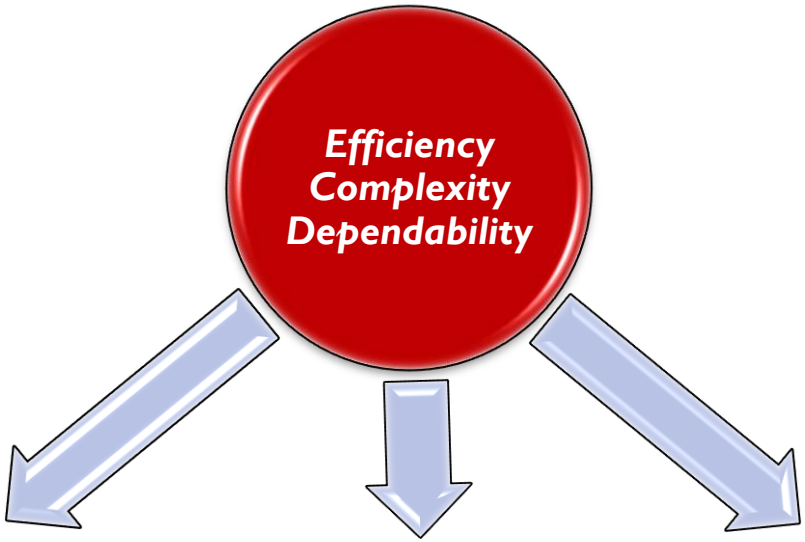
Accelerators become economically more viable

Accelerator / Architecture Longevity (Controversial!)



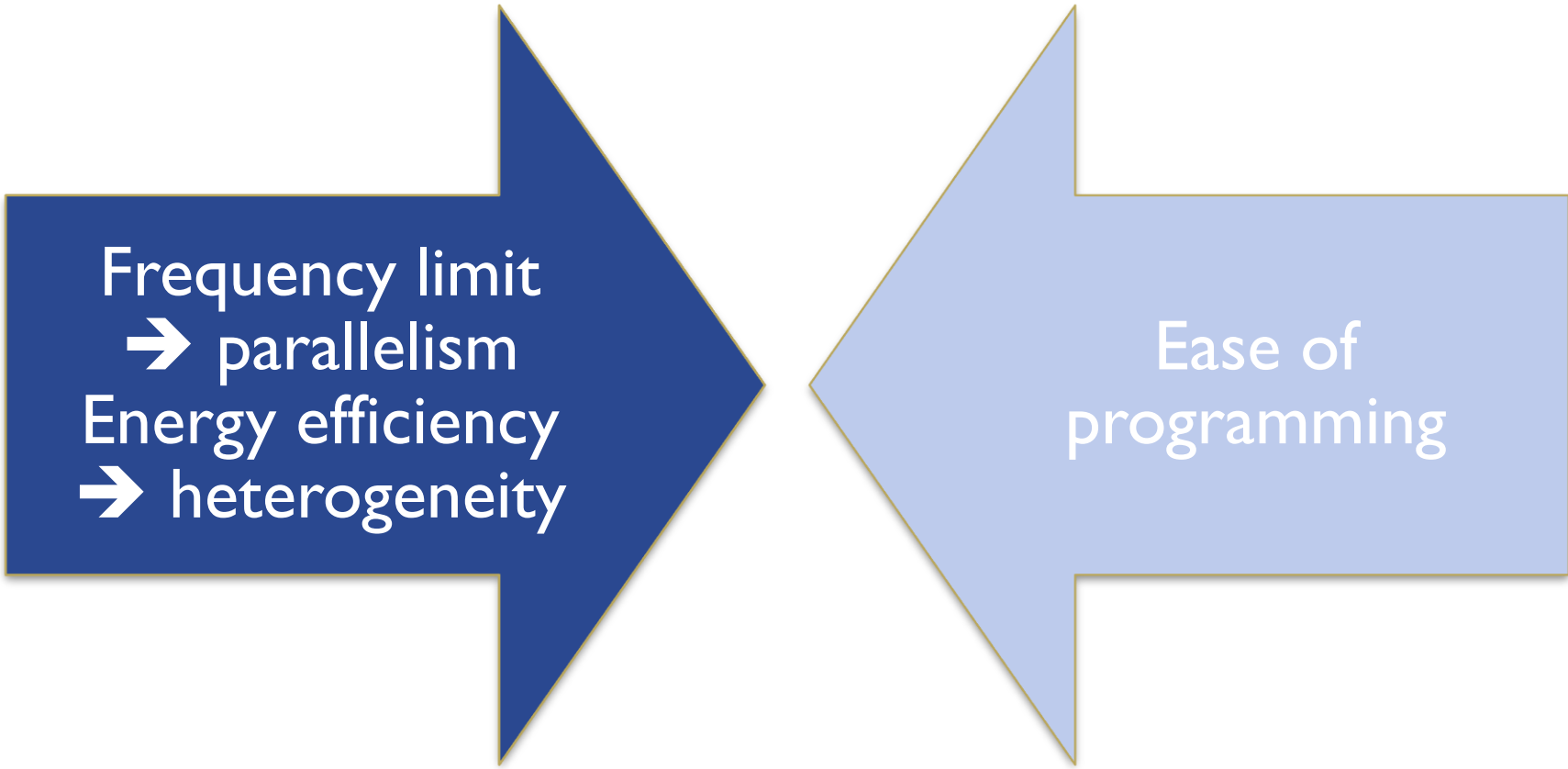
- Frequency growth slowdown ⇨ enhanced accelerator business case:
  - Past (45%): After 3 years – useless.
  - Now (<20%): as much 5 years useful lifetime
- Controversy? *Aggregate* chip performance still at 45%.
- **Has architecture ever really mattered in the long run?**
  - RISC, CISC, vector, SIMD, MIMD, whatever; who cares in 3 years?
  - **Has that changed?**

Core Computing Systems Challenges



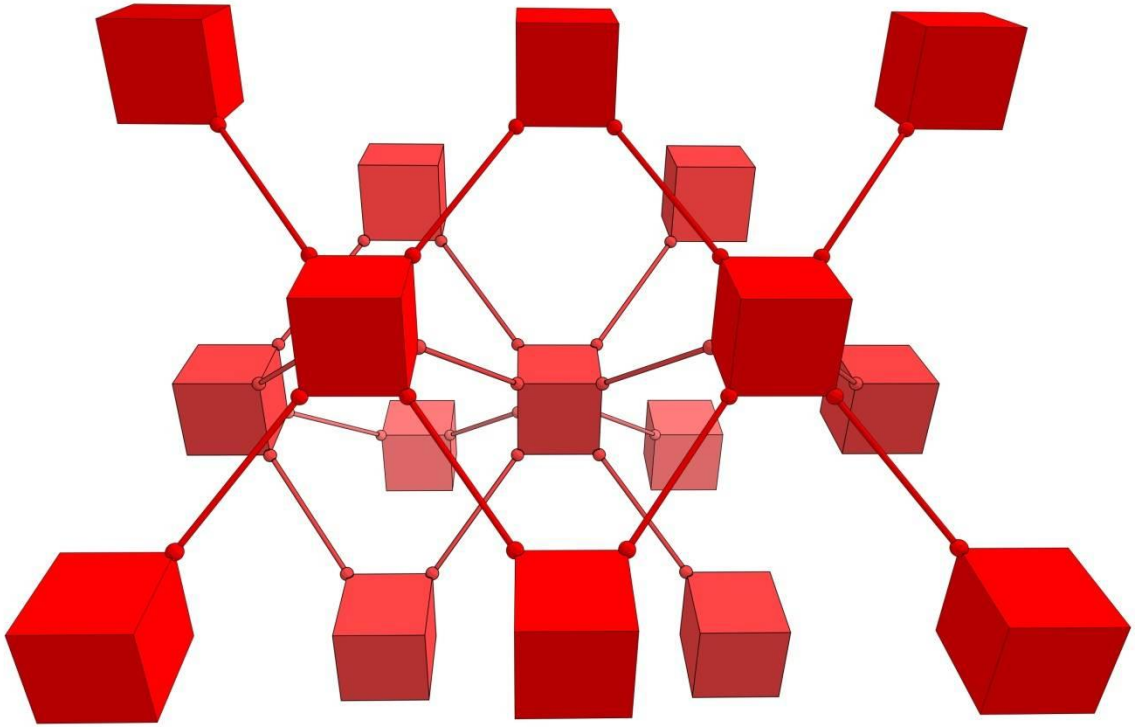
Efficiency	Complexity	Dependability
• Power ⇒ • Performance ⇒	• Heterogeneity • Parallelism	• Reliability • Privacy

Parallelism and specialization are not for free…



Managing complexity

Parallelism seems to be too complex for humans



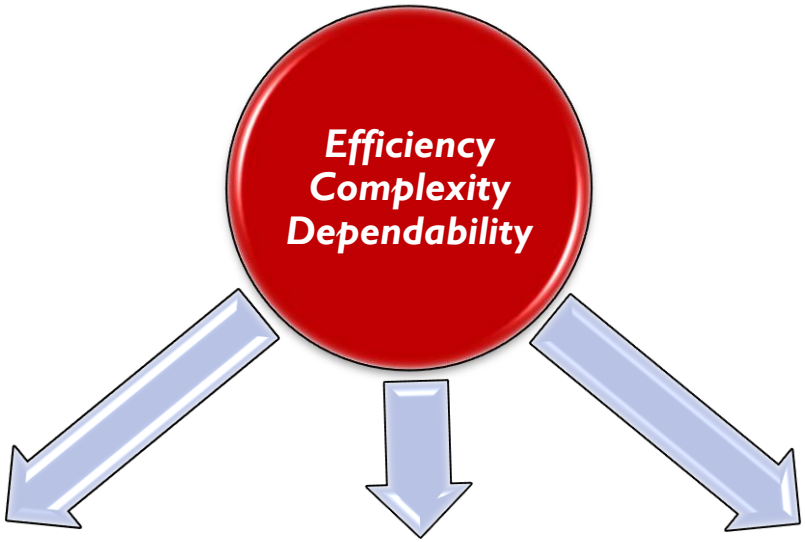
Tools…

Managing complexity

Hardware has become more flexible than software

Portable performance

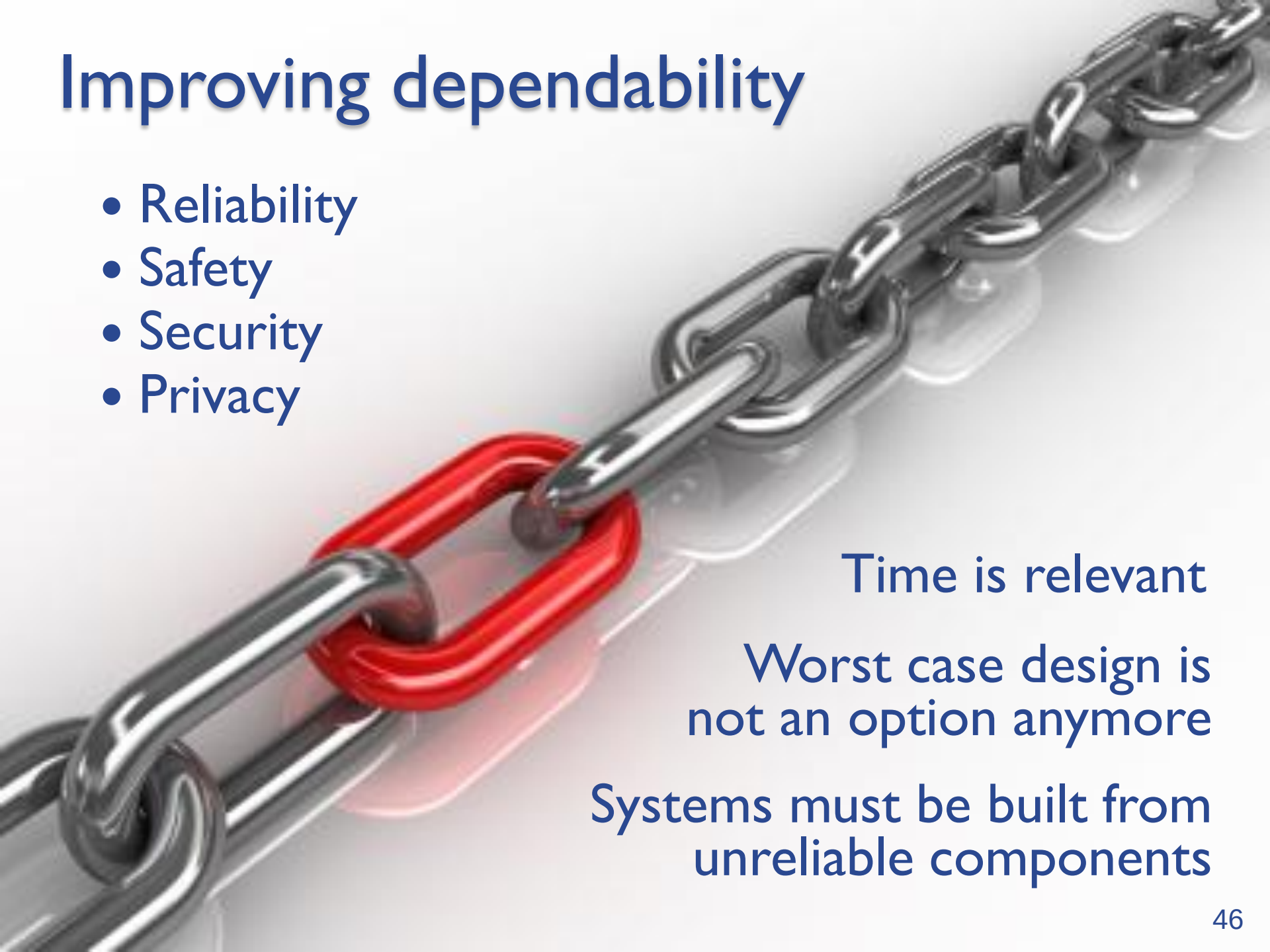
Core Computing Systems Challenges



Efficiency	Complexity	Dependability
• Power ⇒ • Performance ⇒	• Heterogeneity • Parallelism	• Reliability • Privacy

Improving dependability

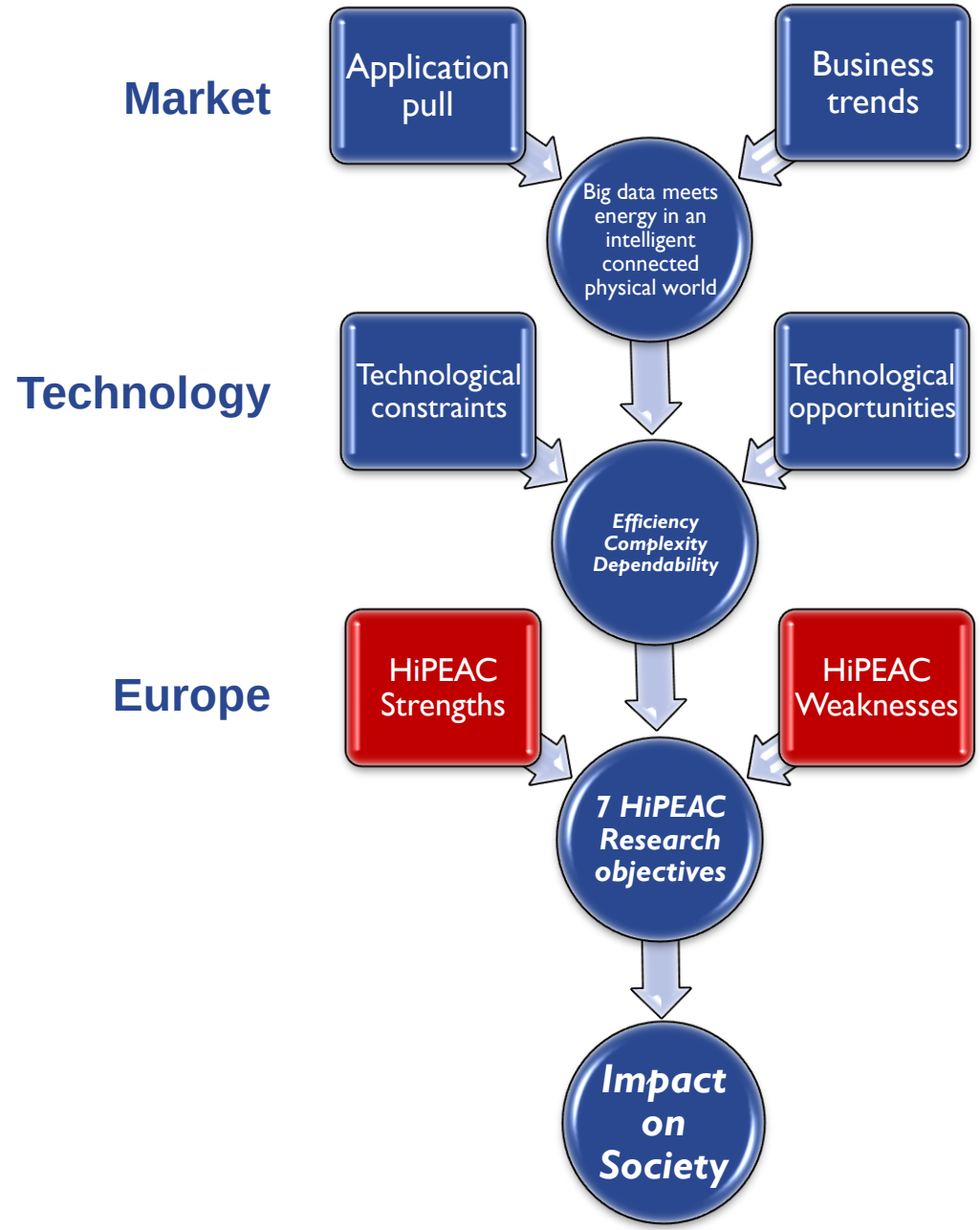
- Reliability
- Safety
- Security
- Privacy



Time is relevant

Worst case design is not an option anymore

Systems must be built from unreliable components


Market
Technology
Europe
Application pull
Business trends
Big data meets energy in an intelligent connected physical world
Technological constraints
Technological opportunities
Efficiency
Complexity
Dependability
HiPEAC Strengths
HiPEAC Weaknesses
7 HiPEAC Research objectives
Impact on Society

SWOT

Strengths

- Strong embedded ecosystem from semiconductor technology to consumer products with global players in all market segments (semiconductors, automotive, aerospace, defense, telecom, …)
- Many innovative SMEs
- Excellent educational system
- Pan-European research centralized Research programs

Weaknesses

- Weak link between academia and industry, especially at the PhD level and brain drain to other countries
- Lack of venture capital, lack of enterpreneurial culture
- Lack of and pan European companies in HiPEAC domain an no general purpose computing company (like HP, Intel, IBM,…), ARM and Bull being the exceptions.
- Relatively low international visibility of computing systems departments, low numbers of foreign students
- The lack of open source tools; lack of affordable CAD tools and validation platforms
- Language barriers

Opportunities

- Challenges like aging population, environment, energy and mobility bring new market opportunities
- Convergence of embedded and GP computing
- Disruptive technology like the cloud, more than Moore also bring new market opportunities
- The recognition of micro- and nano-electronics as a key enabling technology at the EU-level.
- Cultural diversity in Europe creates cultural sensitivity to do business with emerging markets
- Proximity to the Middle East, the Russian Federation and Africa (Huge market opportunity)

Threats

- China might want to compete with ARM
- Financial crisis in the Euro zone

Strengths

- Strong embedded ecosystem from semiconductor technology to consumer products with global players in all market segments (semiconductors, automotive, aerospace, defense, telecom, …)
- Many innovative SMEs
- Excellent educational system
- Pan-European research centralized Research programs

Weaknesses

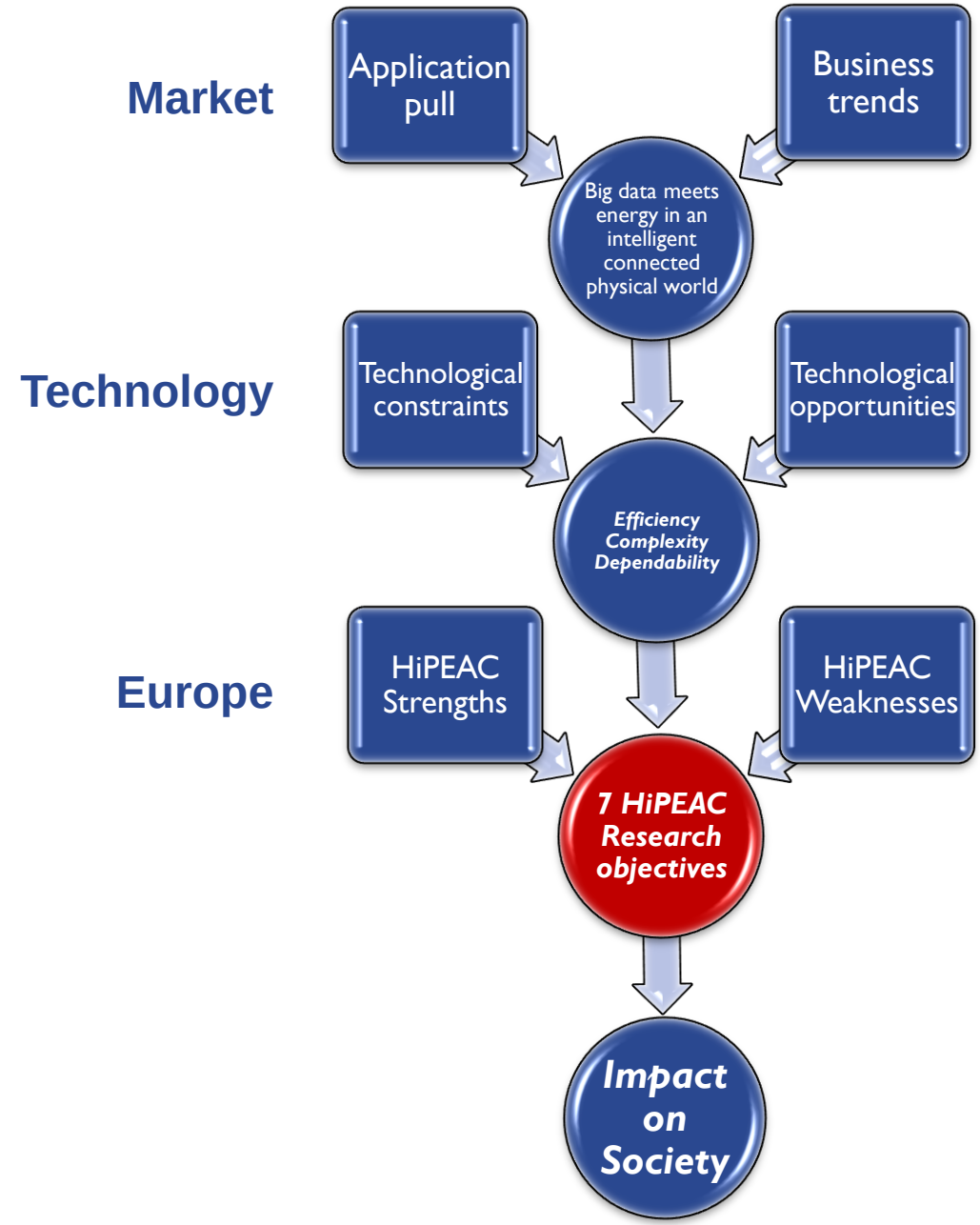
- Weak link between academia and industry, especially at the PhD level and brain drain to other countries
- Lack of venture capital, lack of enterpreneurial culture
- Lack of and pan European companies in HiPEAC domain an no general purpose computing company (like HP, Intel, IBM,…), ARM and Bull being the exceptions.
- Relatively low international visibility of computing systems departments, low numbers of foreign students
- The lack of open source tools; lack of affordable CAD tools and validation platforms
- Language barriers

Opportunities

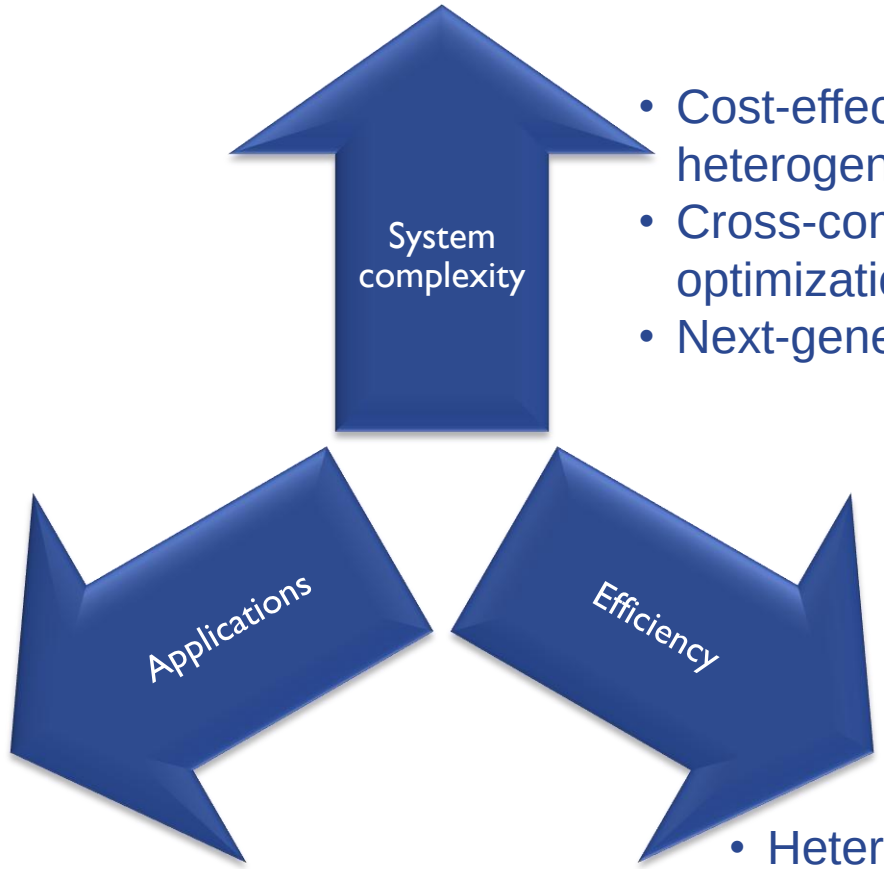
- Challenges like aging population, environment, energy and mobility bring new market opportunities
- Convergence of embedded and GP computing
- Disruptive technology like the cloud, more than Moore also bring new market opportunities
- The recognition of micro- and nano-electronics as a key enabling technology at the EU-level.
- Cultural diversity in Europe creates cultural sensitivity to do business with emerging markets
- Proximity to the Middle East, the Russian Federation and Africa (Huge market opportunity)

Threats

- China might want to compete with ARM
- Financial crisis in the Euro zone


Market
Technology
Europe
Application pull
Business trends
Big data meets energy in an intelligent connected physical world
Technological constraints
Technological opportunities
Efficiency
Complexity
Dependability
HiPEAC Strengths
HiPEAC Weaknesses
7 HiPEAC Research objectives
Impact on Society

Derived HiPEAC Research Objectives



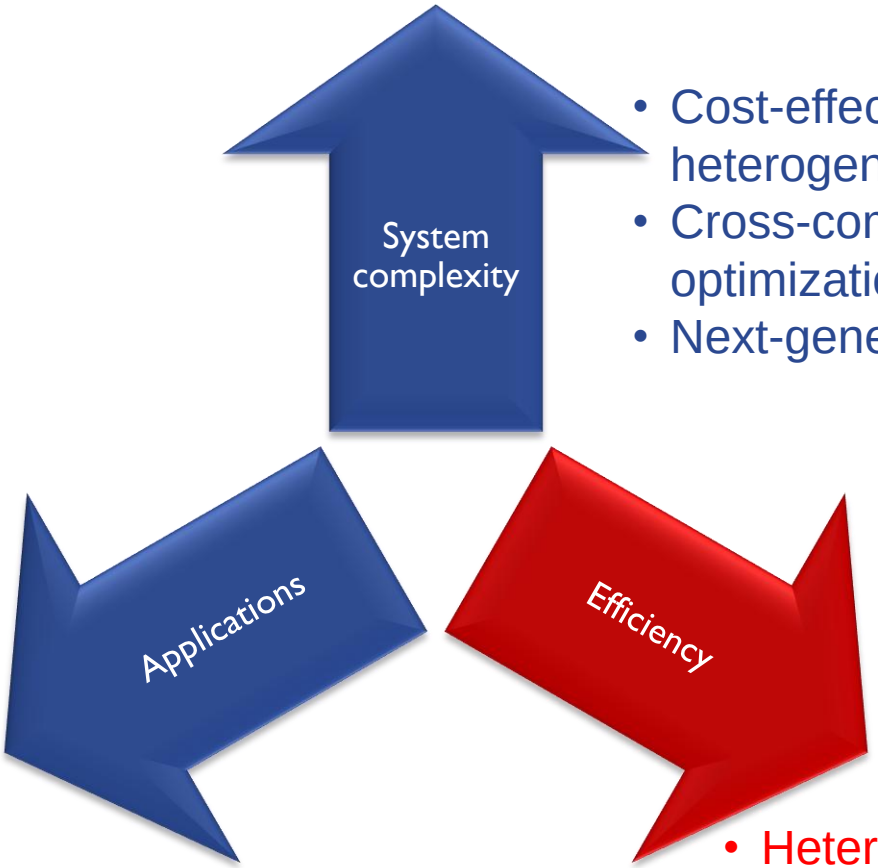
- Cost-effective software for heterogeneous multi-cores
- Cross-component/cross-layer optimization for design integration
- Next-generation processor cores

- Architectures for the Data Deluge
- Reliable systems for Ubiquitous Computing

- Heterogeneous computing systems
- Locality and communications management

Computing systems

Derived HiPEAC Research Objectives

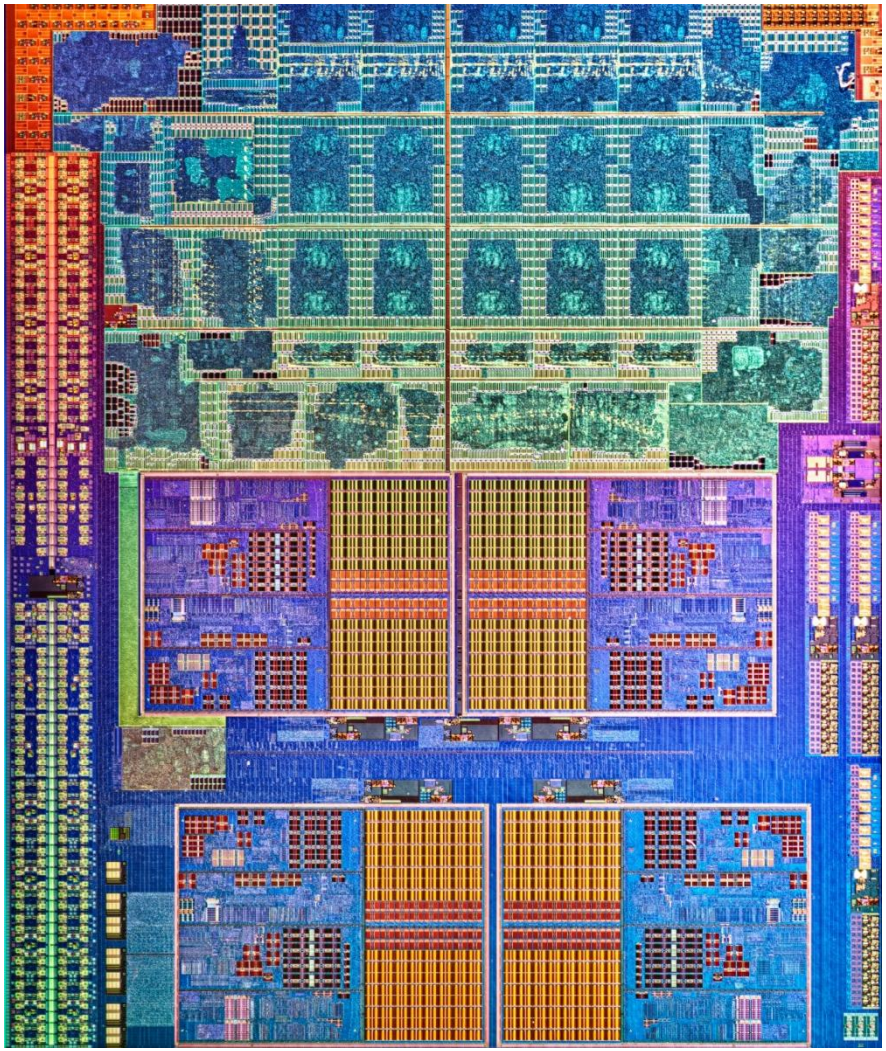


- Cost-effective software for heterogeneous multi-cores
- Cross-component/cross-layer optimization for design integration
- Next-generation processor cores

- Architectures for the Data Deluge
- Reliable systems for Ubiquitous Computing

- Heterogeneous computing systems
- Locality and communications management

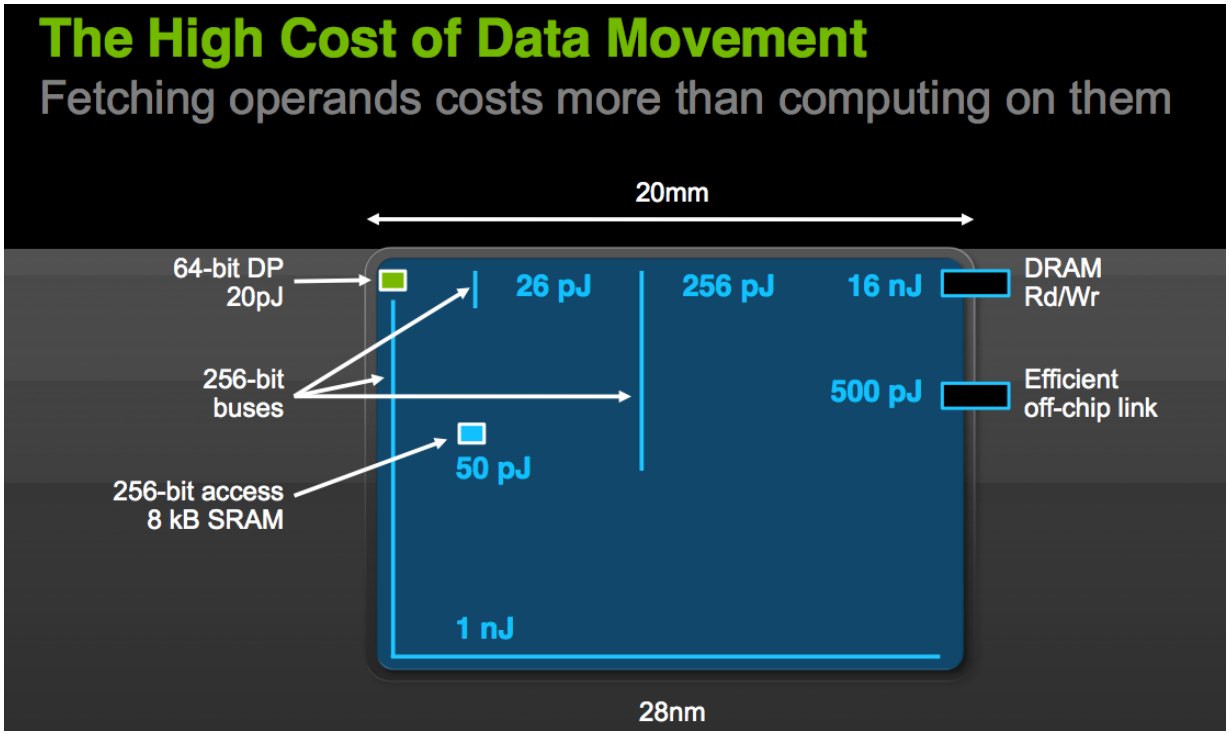
Heterogeneous computing systems



AMD A8-3850 : Llano on the Desktop

- **Short term:**
  Development of efficient accelerators
- **Medium term:**
  Tools for improving design productivity
- **Long term:**
  Automatic porting of legacy applications on parallel and heterogeneous systems

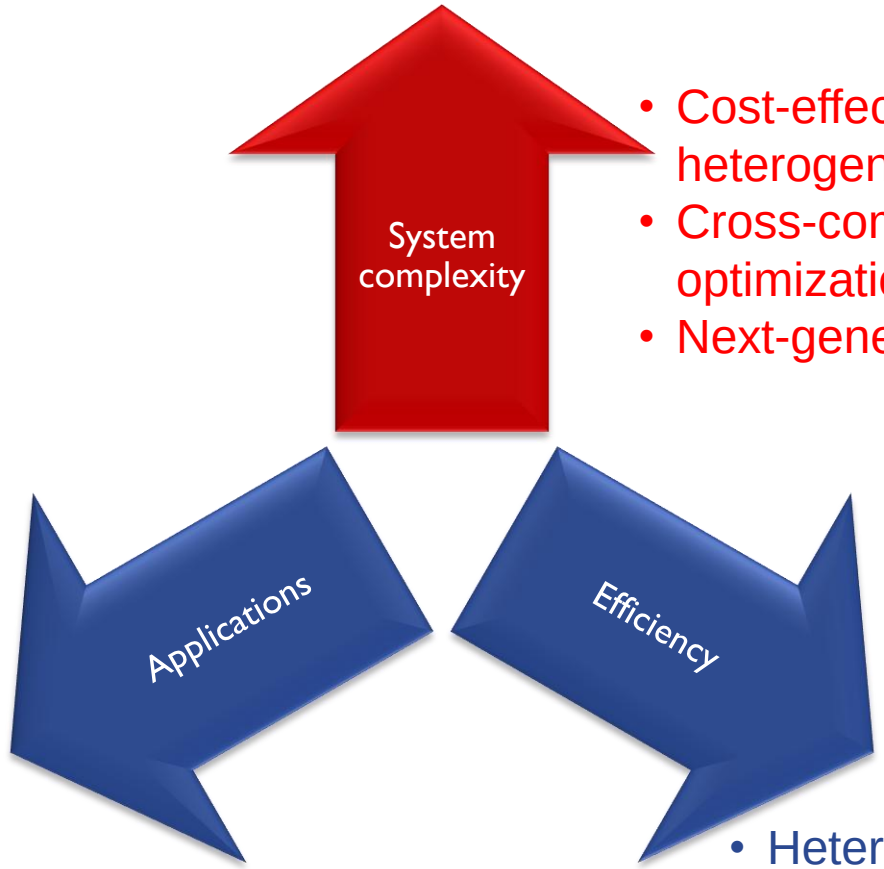
Locality and communications management



Source: Bill Dally, "To ExaScale and Beyond"
www.nvidia.com/content/PDF/sc_2010/theater/Dally_SC10.pdf

- **Short term:** Simple models for manual locality management
- **Medium term:** Automatic static locality management. Architectures with reduced memory hierarchy
- **Long term:** Automatic and dynamic locality management. Use of new storage devices for co-locating storage and processing

Derived HiPEAC Research Objectives



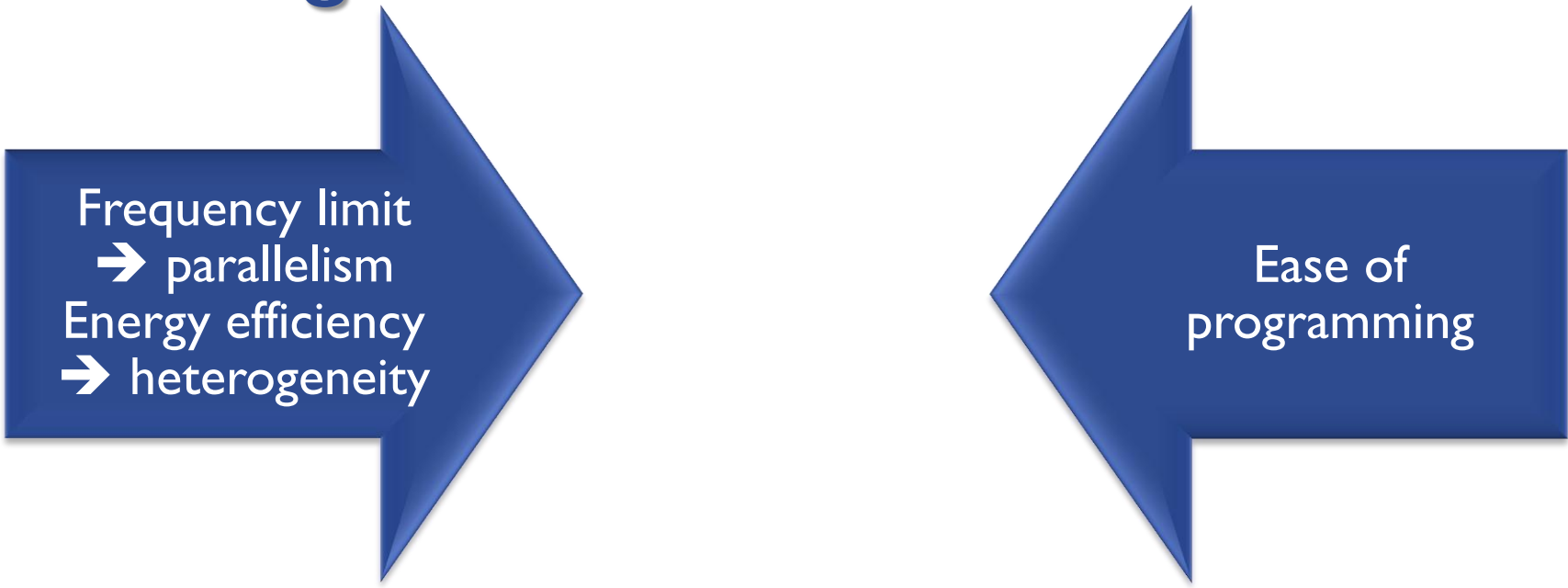
- Cost-effective software for heterogeneous multi-cores
- Cross-component/cross-layer optimization for design integration
- Next-generation processor cores

- Architectures for the Data Deluge
- Reliable systems for Ubiquitous Computing

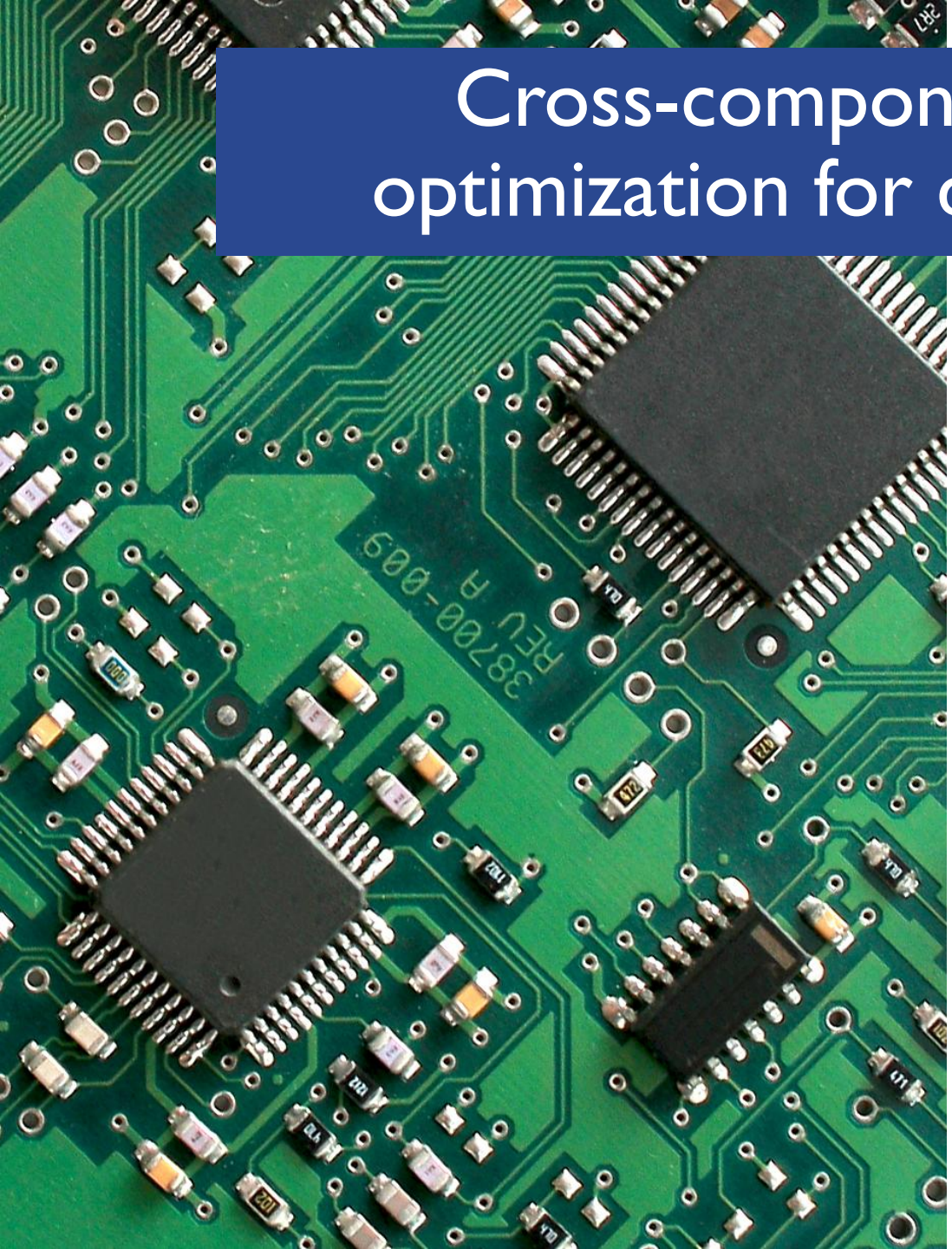
- Heterogeneous computing systems
- Locality and communications management

Computing systems

Cost-effective software for heterogeneous multi-cores



- **Short term:** Systems to enable understanding of existing code and assist with parallelizing. Directive-based compiler tools. Runtime analysis
- **Medium term:** Programming systems with a more integrated runtime and language. Providing object-oriented paradigms across accelerators. Integrated performance/power/timing modeling and optimization
- **Long term:** Full performance portability. Self-adapting software. New programming paradigms

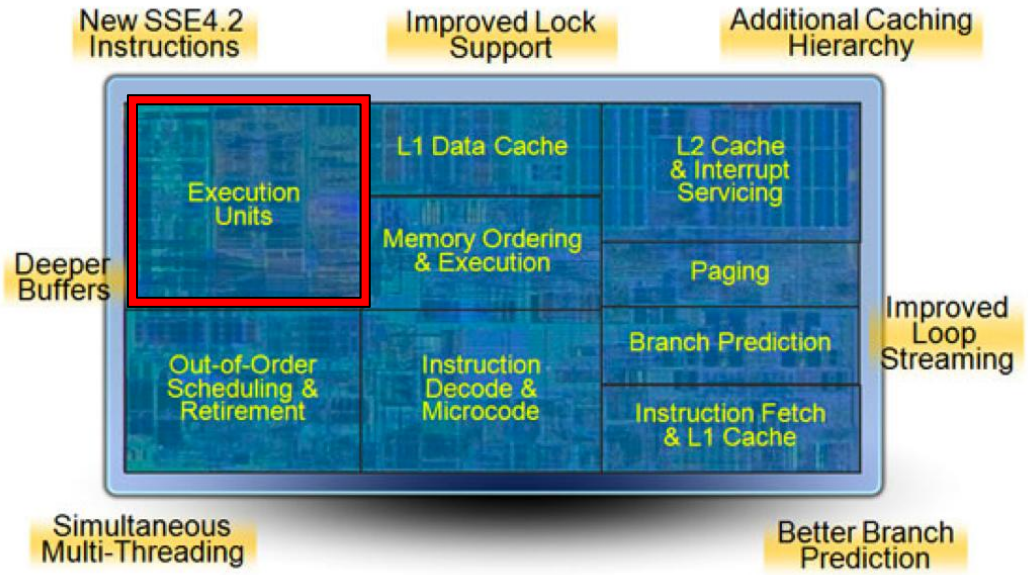


Cross-component/cross-layer optimization for design integration

- **Short term:** Semi-automatic cross component/cross layer static optimization
- **Medium term:** Automatic cross component/cross layer static optimization
- **Long term:** Automatic cross component/cross layer static optimization with dynamic runtime optimization

Next-generation processor cores

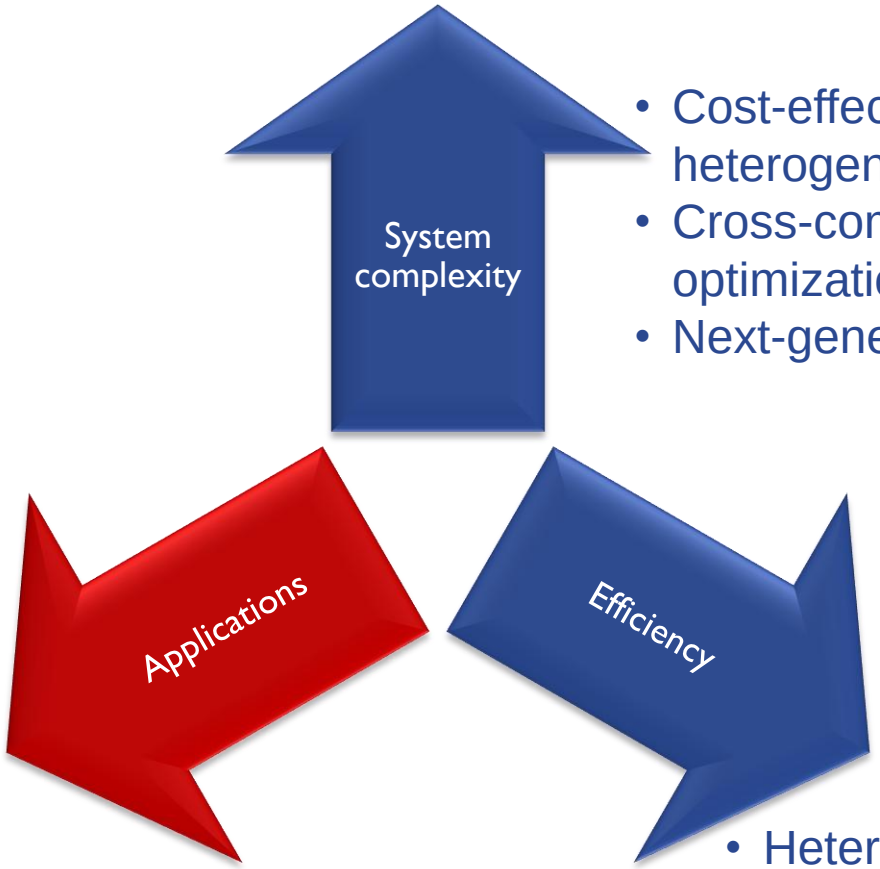
Less than 20% of the area for execution units…



- **Short term:**
  Energy efficient movement within cores, tools to assist in the generation of the hardware of the computing cores and their compilers
- **Medium term:**
  Automated design space exploration tools to propose efficient architectures and compilers
- **Long term:**
  New compute engines with minimized data movement

Source: Dan Connors, "OpenCL and CUDA Programming for Multicore and GPU Architectures" ACACES 2011

Derived HiPEAC Research Objectives

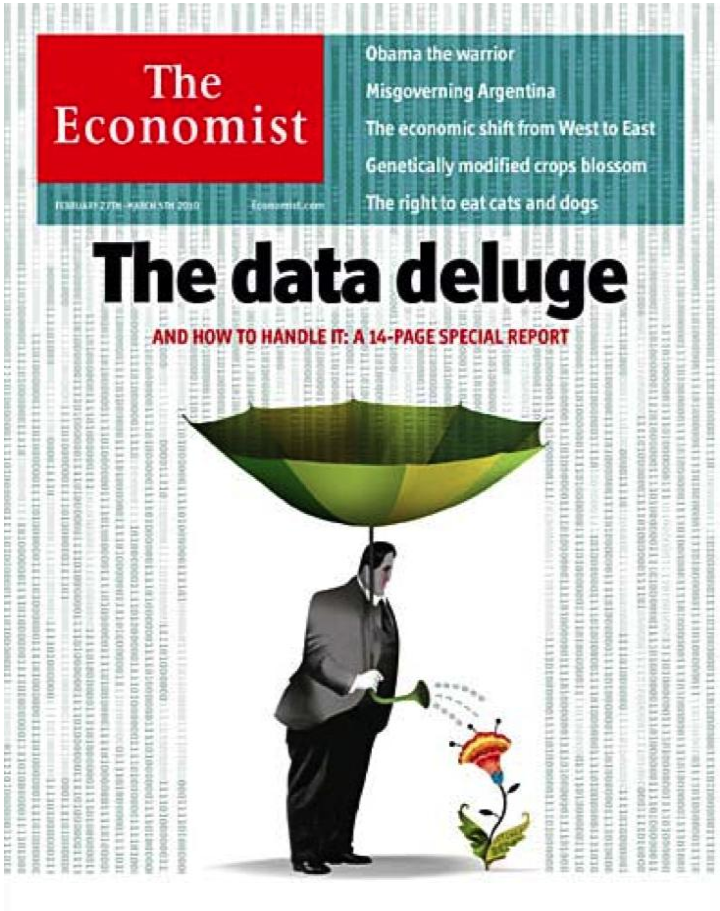


- Cost-effective software for heterogeneous multi-cores
- Cross-component/cross-layer optimization for design integration
- Next-generation processor cores

- Architectures for the Data Deluge
- Reliable systems for Ubiquitous Computing

- Heterogeneous computing systems
- Locality and communications management

Architectures for the Data Deluge



- **Short term:**
  Energy efficient architectures for data processing. Latency analysis tools
- **Medium term:**
  System development tools for minimizing latency and energy. New concepts and processing paradigms for natural data processing
- **Long term:**
  Real-time analysis of data. Accelerators for natural data processing using new computing paradigms

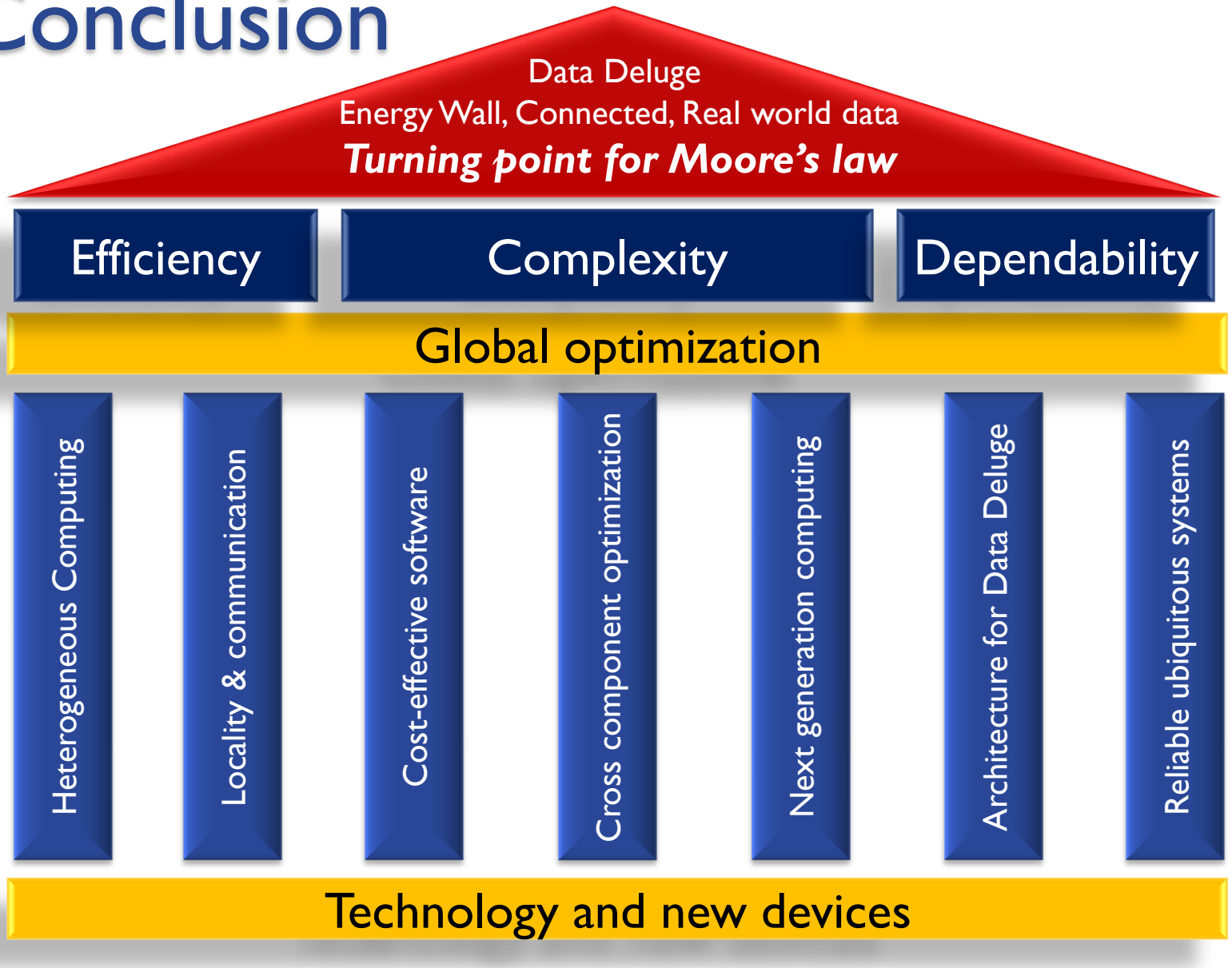
Reliable systems for Ubiquitous Computing

- **Short term:** Manually secured and verified systems
- **Medium term:** Semi-automatically secured and verified systems. First designs with variability and fault tolerance
- **Long term:** Fully automatically secured and verified systems, or correct-by-design tool chains. Self-reconfiguring systems to optimized variability and errors

Detailled HiPEAC Research areas

		Heterogeneous computing systems	Locality and communications management	Software for heterogeneous multi-cores	Cross-component/cross-layer optimization	Next -generation processing cores	Architectures for the Data Deluge	Reliable systems for Ubiquitous Computing
Parallelism and Programming Models								
9.1.1.	Locality Management	x	x	x	x	x	x	x
9.1.2.	Optimizations programmer hints, tuning	x	x	x	x			x
9.1.3.	Runtime Systems and Adaptivity	x	x	x	x		x	x
Architecture								
9.2.1.	Processors, Accelerators, Heterogeneity	x	x			x		x
9.2.2.	Memory Architectures	x	x		x	x	x	
9.2.3.	Interconnection Architectures	x	x		x	x	x	
9.2.4.	Reconfigurability	x	x		x	x		
Compilers								
9.3.1.	Automatic Parallelization		x	x		x		
9.3.2.	Adaptive Compilation			x			x	x
9.3.3.	Intelligent Optimization			x	x	x		x
Systems Software and Tools								
9.4.1.	Virtualization	x		x	x		x	x
9.4.2.	Input, Output, Storage, and Networking		x		x		x	
9.4.3.	Simulation and Design Automation Tools	x			x	x		
9.4.4.	Deterministic Performance Tools	x	x	x	x	x	x	x

Conclusion



Exiting new opportunities are ahead of us!

http://www.hipeac.net/roadmap

HiPEAC vision 2011-2012

http://www.hipeac.net/roadmap