21st Century Computer Architecture

Mark D. Hill, Univ. of Wisconsin-Madison

6/2016 @ Architecture 2030 @ ISCA

1. **Whitepaper Content (tell a story)**
2. Process, Impact & Computing Community Consortium (CCC)

21st Century Computer Architecture

A CCC community white paper

http://cra.org/ccc/docs/init/21stcenturyarchitecturewhitepaper.pdf

- Information & Commun. Tech's Impact
- Semiconductor Technology's Challenges
- Computer Architecture's Future
- Pre-Competitive Research Justified

Was ***21st Century Computer Architecture*** NEW (even in 2012)??

Our Task: Telling a "New" Story

- Was ***21st Century Computer Architecture*** NEW?
  - **No**
  - **Resulted in more $50M in funding in USA**

In Communication of the China Computer Federation (CCF)

译文 中国计算机学会通讯 第8卷 第12期 2012年12月

21世纪计算机体系结构

计算机体系结构共同体白皮书*

关键词：计算机体系结构

译者：计算机体系结构国家重点实验室（筹）

引言与概要

信息通信技术（information and communication technology，ICT）从医疗、教育、科技、商业、政府、国防和娱乐等各个领域改变着我们的世界。如今，人们已经差不多忘记了20年前人们找资料的第一步是设法到达图书馆，10年前社交网络还主要指

持续地让单个芯片在功耗与成本几乎维持不变的情况下（Dennard scaling，丹纳德微缩）容纳了越来越多的晶体管（摩尔定律）。而计算机体系结构发明了各种创新技术来利用快速增长的晶体管资源，扩展了处理器的性能，缓解了内存系统带来的性能损失。在过去几十年，两者结合的效果让信息通信

Our Task: Telling a "New" Story

- Was ***21st Century Computer Architecture*** NEW?
  - **No**
  - **Resulted in more $50M in funding in USA**
- **Why?**
  - New to **others**
  - Told a story that mattered to **others**

- Develop a New-to-Others Message as a Story
  - Why Important to Others?
  - Why Now?
  - How Might Research Make a Difference?

- This is our task now – not developing ideas new to us

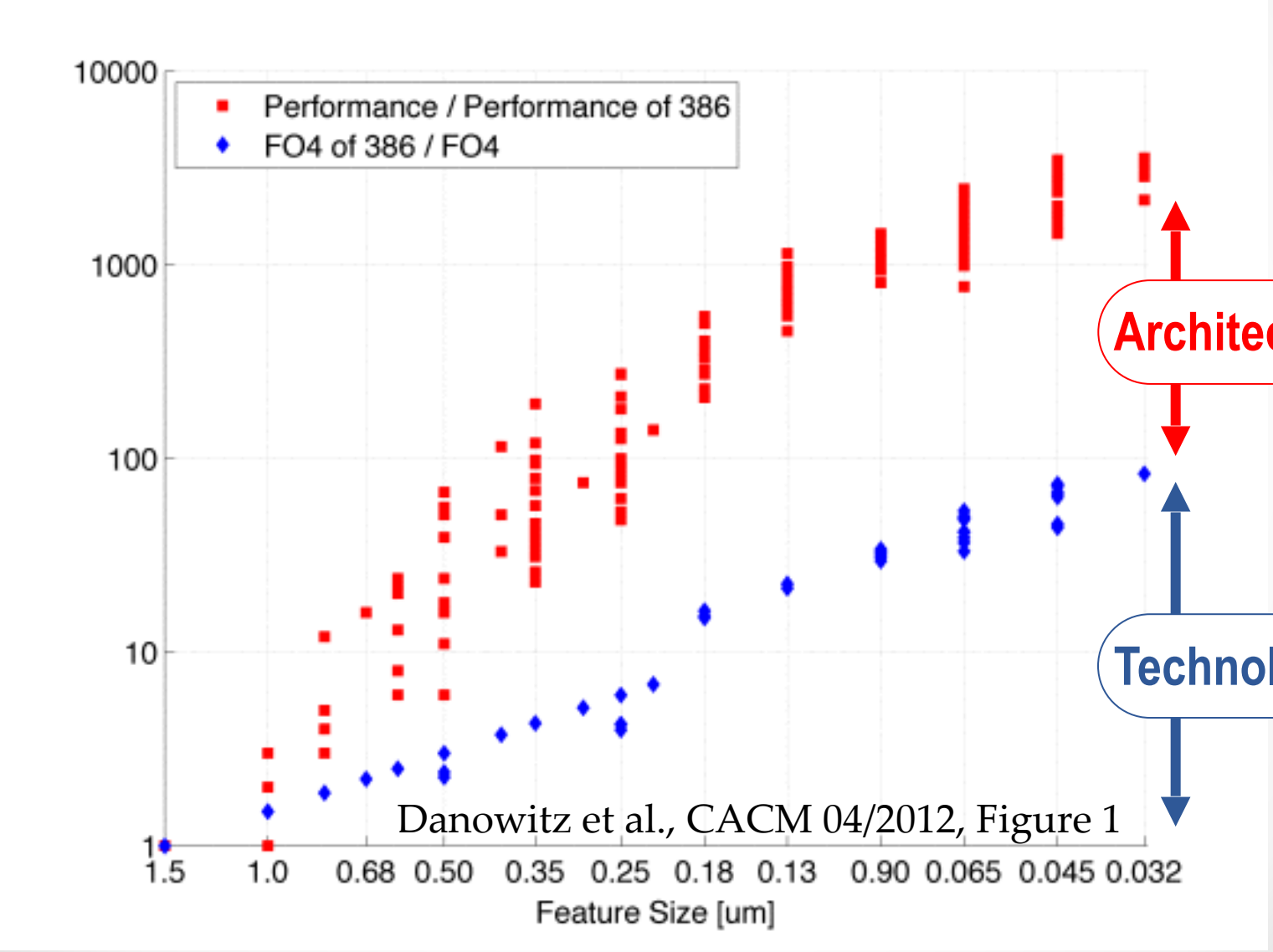
20th Century ICT Set Up

- Information & Communication Technology (ICT) Has Changed Our World
  - <long list omitted>

- Required innovations in algorithms, applications, programming languages, … , & system software

- Key (invisible) enablers (cost-)performance gains
  - Semiconductor technology ("Moore's Law")
  - Computer architecture (~80x per Danowitz et al.)

Enablers: Technology + Architecture



Danowitz et al., CACM 04/2012, Figure 1

21st Century ICT Promises More

Data-centric personalized health care

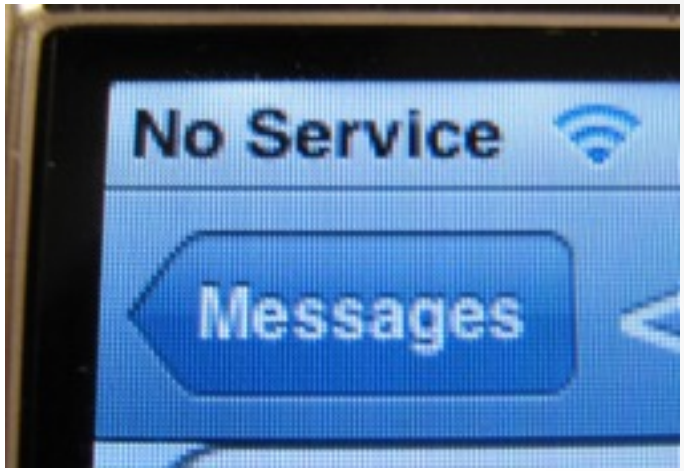
Computation-driven scientific discovery

Human network analysis

Much more: known & unknown

21st Century App Characteristics

BIG DATA



ALWAYS ONLINE

SECURE/PRIVATE

Whither enablers of future (cost-)performance gains?

Technology's Challenges 1/2

Late 20th Century	The New Reality
Moore's Law — 2× transistors/chip	**Transistor count still 2× BUT…**
Dennard Scaling — ~constant power/chip	**Gone.** Can't repeatedly double power/chip

Technology's Challenges 2/2

Late 20th Century	The New Reality
Moore's Law — 2× transistors/chip	**Transistor count still 2× BUT…**
Dennard Scaling — ~constant power/chip	**Gone.** Can't repeatedly double power/chip
Modest (hidden) transistor unreliability	**Increasing transistor unreliability** can't be hidden
Focus on computation over communication	**Communication (energy)** more expensive than computation
1-time costs amortized via mass market	**One-time cost** much worse & want **specialized** platforms

How should architects step up as technology falters?

21st Century Comp Architecture

20th Century	21st Century	
Single-chip in generic computer		
Performance via invisible instr.-level parallelism		
Predictable technologies: CMOS, DRAM, & disks		

21st Century Comp Architecture

20th Century	21st Century	
Single-chip in generic computer	**Architecture as Infrastructure:** Spanning sensors to clouds Performance + security, privacy, availability, programmability, …	
Performance via invisible instr.-level parallelism		
Predictable technologies: CMOS, DRAM, & disks		

21st Century Comp Architecture

20th Century	2	
Single-chip in generic computer		
Performance via invisible instr.-level parallelism		
Predictable technologies: CMOS, DRAM, & disks		

21st Century Comp Architecture

20th Century	21st Century	
Single-chip in generic computer		
Performance via invisible instr.-level parallelism		
Predictable technologies: CMOS, DRAM, & disks		

21st Century Comp Architecture

20th Century	21st Century	
Single-chip in stand-alone computer	**Architecture as Infrastructure:** Spanning sensors to clouds Performance + security, privacy, availability, programmability, …	**Cross-Cutting:** Break current layers with new interfaces
Performance via invisible instr.-level parallelism	**Energy First** • Parallelism • Specialization • Cross-layer design	
Predictable technologies: CMOS, DRAM, & disks	**New technologies** (non-volatile memory, near-threshold, 3D, photonics, …) Rethink: memory & storage, reliability, communication	

What Research Exactly?

- Research areas in white paper (& backup slides)
  1. Architecture as Infrastructure: Spanning Sensors to Clouds
  2. Energy First
  3. Technology Impacts on Architecture
  4. Cross-Cutting Issues & Interfaces

- Much more research developed by future PIs!

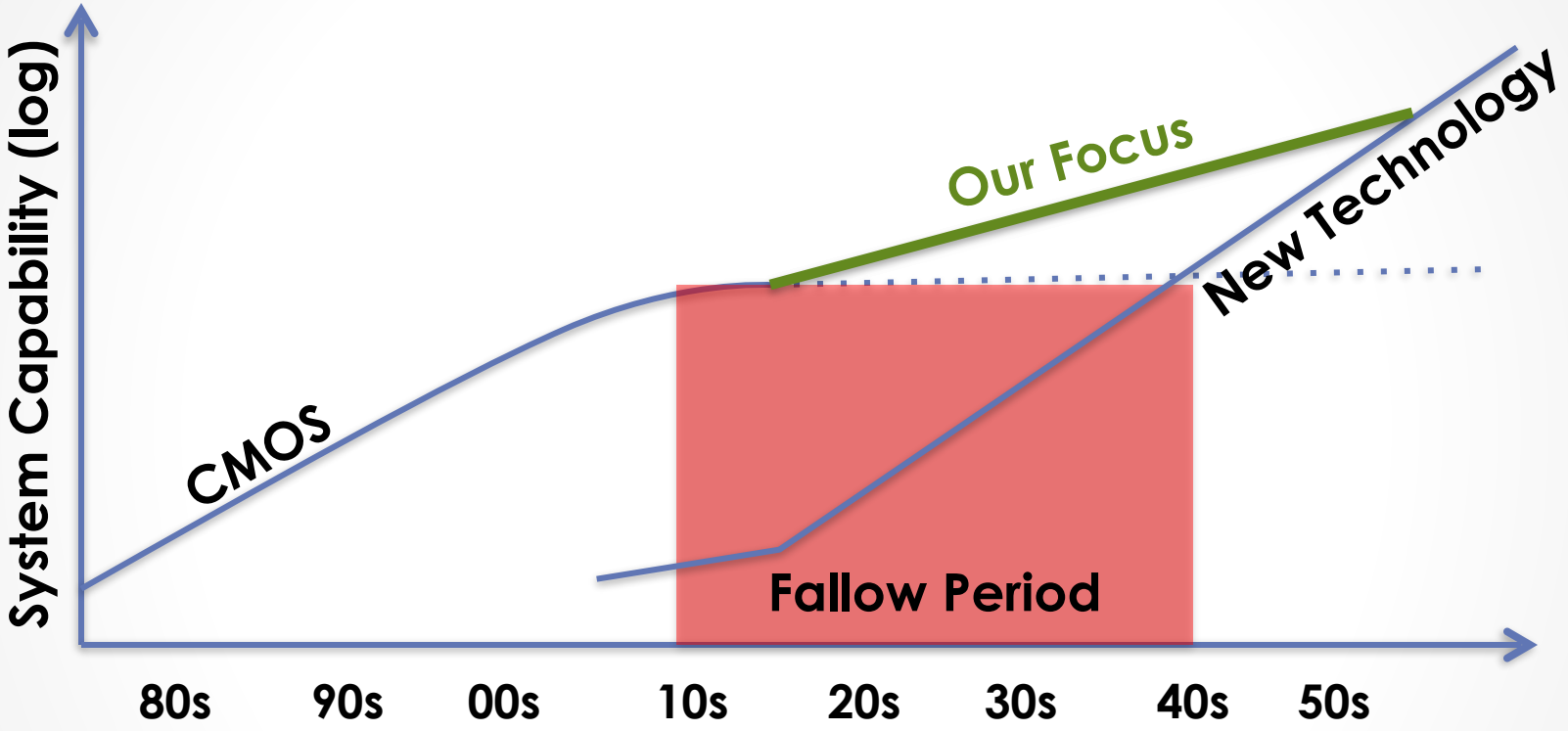
Pre-Competitive Research Justified

- **Retain (cost-)performance enabler to ICT revolution**

- Successful companies cannot do this by themselves
  - Lack needed long-term focus
  - Don't want to pay for what benefits all
  - Resist transcending interfaces that define their products

- Corroborates
  - Future of Computing Performance: Game Over or Next Level?, National Academy Press, 2011
  - DARPA/ISAT Workshop Advancing Computer Systems without Technology Progress with outbrief http://www.cs.wisc.edu/~markhill/papers/isat2012_ACSWTP.pdf

"Timeline" from DARPA ISAT



Source: Advancing Computer Systems without Technology Progress, ISAT Outbrief (http://www.cs.wisc.edu/~markhill/papers/isat2012_ACSWTP.pdf)
Mark D. Hill and Christos Kozyrakis, DARPA/ISAT Workshop, March 26-27, 2012.

21st Century Computer Architecture

Mark D. Hill, Univ. of Wisconsin-Madison

6/2016 @ Architecture 2030 @ ISCA

1. Whitepaper Content (tell a story)
2. **Process, Impact & Computing Community Consortium (CCC)**

White Paper Participants

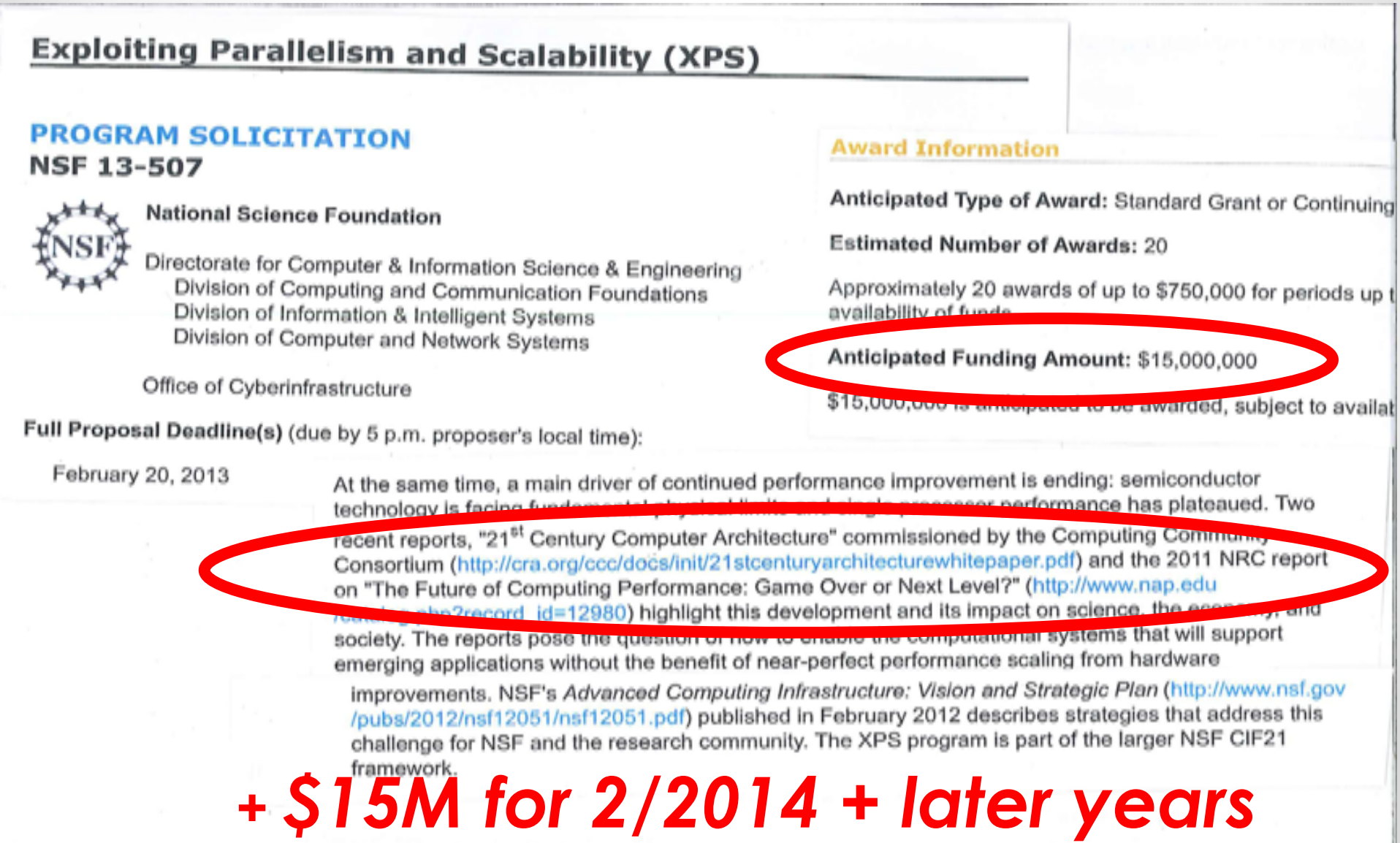
Sarita Adve, U Illinois *
David H. Albonesi, Cornell U
David Brooks, Harvard U
Luis Ceze, U Washington *
Sandhya Dwarkadas, U Rochester
Joel Emer, Intel/MIT
Babak Falsafi, EPFL
Antonio Gonzalez, Intel/UPC
Mark D. Hill, U Wisconsin *,**
Mary Jane Irwin, Penn State U *
David Kaeli, Northeastern U *
Stephen W. Keckler, NVIDIA/U Texas
Christos Kozyrakis, Stanford U
Alvin Lebeck, Duke U
Milo Martin, U Pennsylvania
José F. Martínez, Cornell U
Margaret Martonosi, Princeton U *
Kunle Olukotun, Stanford U
Mark Oskin, U Washington
Li-Shiuan Peh, M.I.T.
Milos Prvulovic, Georgia Tech
Steven K. Reinhardt, AMD
Michael Schulte, AMD/U Wisconsin
Simha Sethumadhavan, Columbia U
Guri Sohi, U Wisconsin
Daniel Sorin, Duke U
Josep Torrellas, U Illinois *
Thomas F. Wenisch, U Michigan *
David Wood, U Wisconsin *
Katherine Yelick, UC Berkeley/LBNL *

Thanks of CCC, Erwin Gianchandani & Ed Lazowska for guidance and Jim Larus & Jeannette Wing for feedback

White Paper Process

- Late March 2012
  - CCC contacts coordinator & forms group
- April 2012
  - Brainstorm (meetings/online doc)
  - Read related docs (PCAST, NRC Game Over, ACAR1/2, …)
  - Use online doc for intro & outline then parallel sections
  - Rotated authors to revise sections
- May 2012
  - Brainstorm list of researcher in/out of comp. architecture
  - Solicit researcher feedback/endorsement
  - Do distributed revision & redo of intro
  - Release May 25 to CCC & via email
- Later
  - CCC & NSF Outbriefs
  - HPCA/PPoPP/CGO/ICS Keynotes
  - ASPLOS & ISCA-Workshop Panels

$15M NSF XPS 2/2013



Exploiting Parallelism and Scalability (XPS)

PROGRAM SOLICITATION
NSF 13-507

National Science Foundation

Directorate for Computer & Information Science & Engineering
Division of Computing and Communication Foundations
Division of Information & Intelligent Systems
Division of Computer and Network Systems

Office of Cyberinfrastructure

Full Proposal Deadline(s) (due by 5 p.m. proposer's local time):

February 20, 2013

Award Information

Anticipated Type of Award: Standard Grant or Continuing

Estimated Number of Awards: 20

Approximately 20 awards of up to $750,000 for periods up t
availability of funds

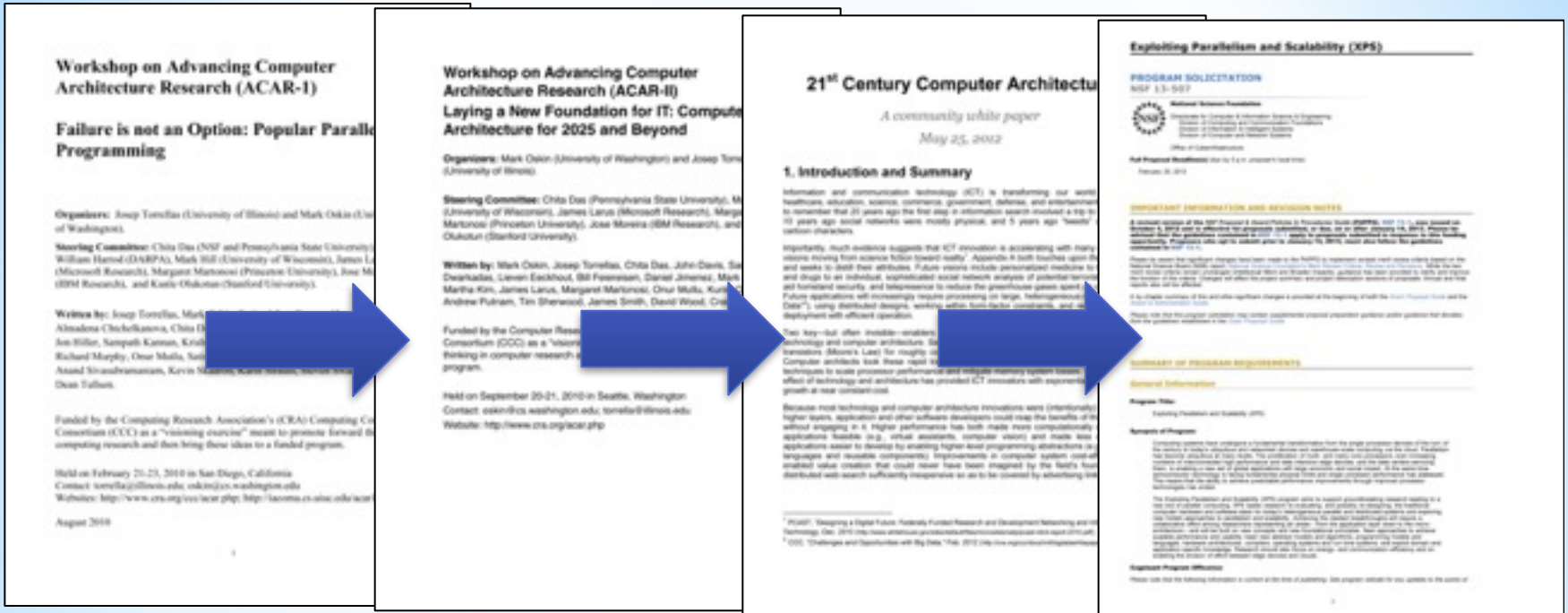
Anticipated Funding Amount: $15,000,000

$15,000,000 is anticipated to be awarded, subject to availat

At the same time, a main driver of continued performance improvement is ending: semiconductor technology is facing fundamental physical limits and single processor performance has plateaued. Two recent reports, "21st Century Computer Architecture" commissioned by the Computing Community Consortium (http://cra.org/ccc/docs/init/21stcenturyarchitecturewhitepaper.pdf) and the 2011 NRC report on "The Future of Computing Performance: Game Over or Next Level?" (http://www.nap.edu /catalog.php?record_id=12980) highlight this development and its impact on science, the economy, and society. The reports pose the question of how to enable the computational systems that will support emerging applications without the benefit of near-perfect performance scaling from hardware improvements. NSF's *Advanced Computing Infrastructure: Vision and Strategic Plan* (http://www.nsf.gov /pubs/2012/nsf12051/nsf12051.pdf) published in February 2012 describes strategies that address this challenge for NSF and the research community. The XPS program is part of the larger NSF CIF21 framework.

+ $15M for 2/2014 + later years

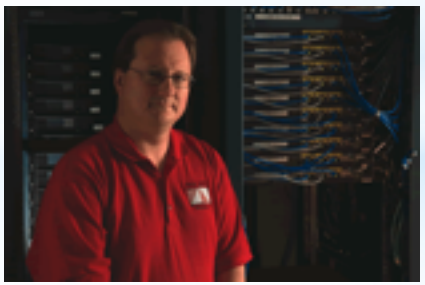
Catalyzing and Enabling: Architecture



2010 2010 2012 2013

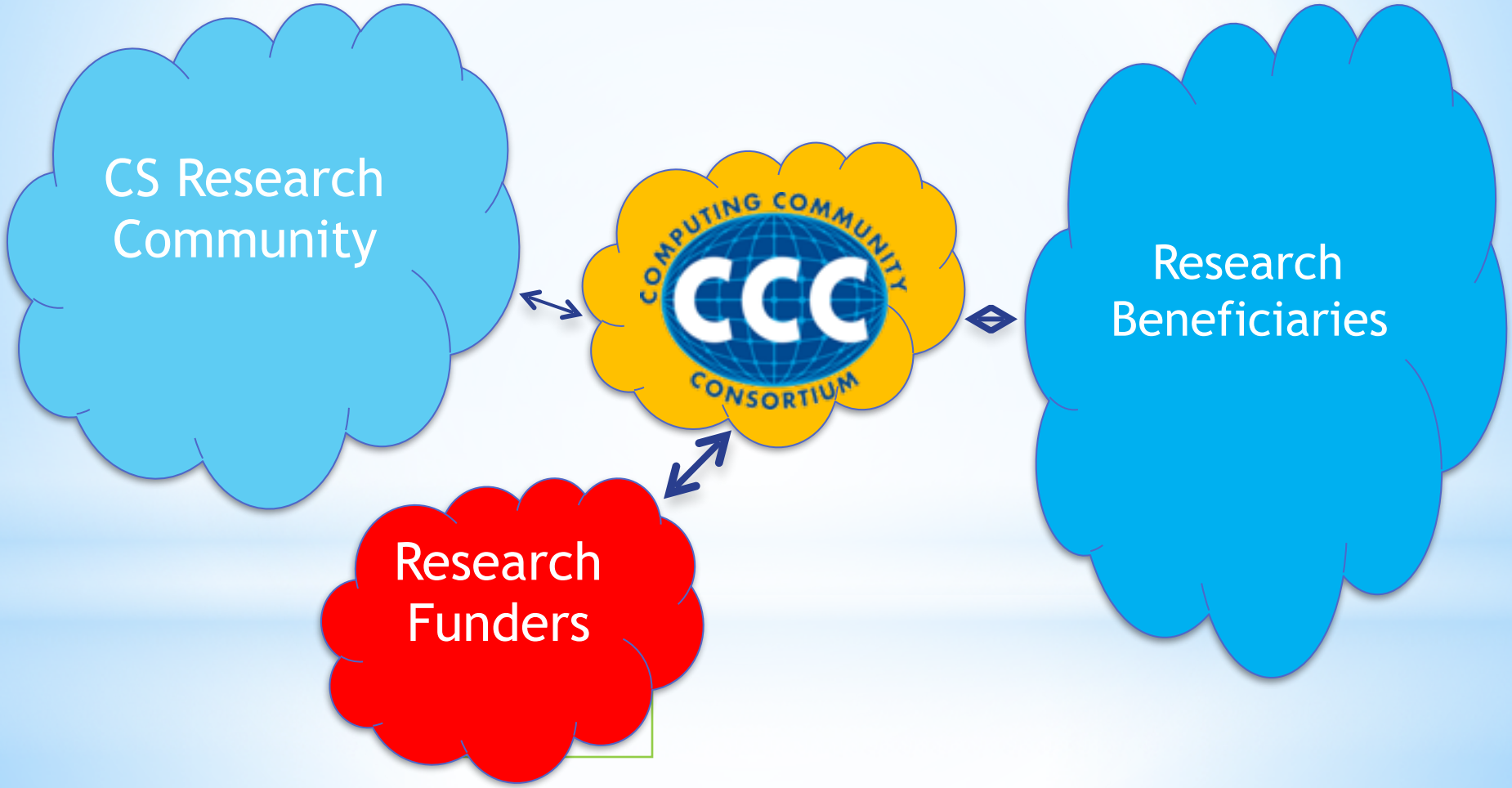
Josep Torrellas
UIUC

Mark Oskin
Washington



Mark Hill
Wisconsin

Computing Community Consortium



CCC & You

- CCC
  - Chair Greg Hager → Beth Mynatt
  - Vice Chair Beth Mynatt → **Mark Hill**
  - ~20 Council Members
  - Standing committee of Computing Research Association
  - CRA established via NSF cooperative agreement

- You
  - Be a rain-maker & give forward
  - White papers
  - Visioning
  - And more

Our Task: Telling a "New" Story

- Was ***21st Century Computer Architecture*** NEW?
  - **No**
  - **Resulted in more $50M in funding**
- Why?
  - New to others
  - Told a story that mattered to others

- Develop a New-to-Others Message as a Story
  - Why Important to Others?
  - Why Now?
  - How Might Research Make a Difference?

- This is our task now – not developing ideas new to us