

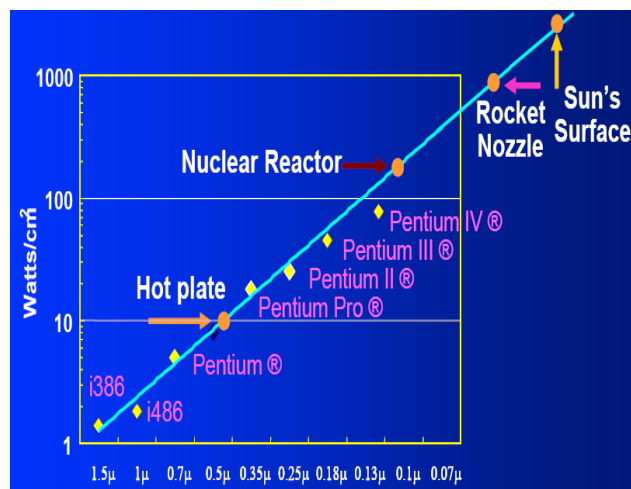
# Energy-Efficient Computing and Applications

Jason Cong  
Chancellor's Professor, UCLA Computer Science Department  
[cong@cs.ucla.edu](mailto:cong@cs.ucla.edu)

Director, Center for Domain-Specific Computing  
[www.cdsc.ucla.edu](http://www.cdsc.ucla.edu)

Co-Director, UCLA/PKU Joint Research Institute in Science and Engineering  
<http://www.pku-jri.ucla.edu/>

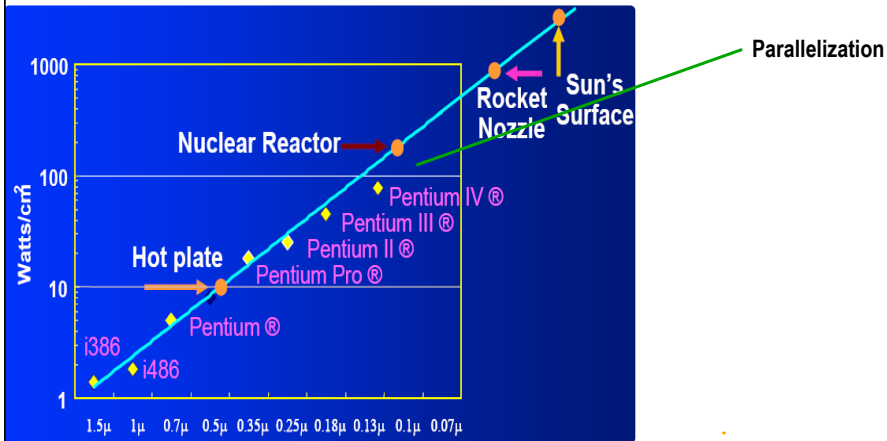
## The Power Barrier ...



Source : Shekhar Borkar, Intel

## Current Solution to Power Barrier

- 10's to 100's cores in a processor
- 1000's to 10,000's servers in a data center



3

## Cost and Energy are Still a Big Issue ...

Hiding in Plain Sight, Google Seeks More Power



Melanie Conner for The New York Times

Google is building two computing centers, top and left, each the size of a football field, in The Dalles, Ore.

By JOHN MARKOFF and SAUL HANSELL  
Published June 14, 2006

THE DALLES, Ore., June 8 — On the banks of the windswept Columbia River, Google is working on a secret weapon in its quest to dominate the next generation of Internet computing. But it is hard to keep a secret when it is a computing center as big as two football fields, with twin cooling plants protruding four stories into the sky.

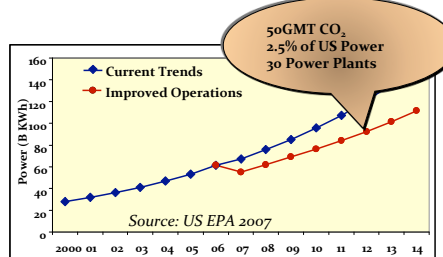
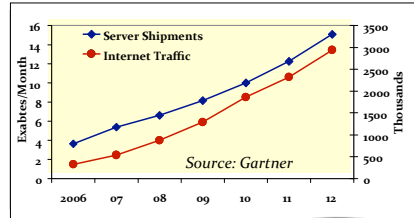
### Cost of computing

- HW acquisition
- Energy bill
- Heat removal
- Space
- ...

4

## Energy Usage of Data Centers

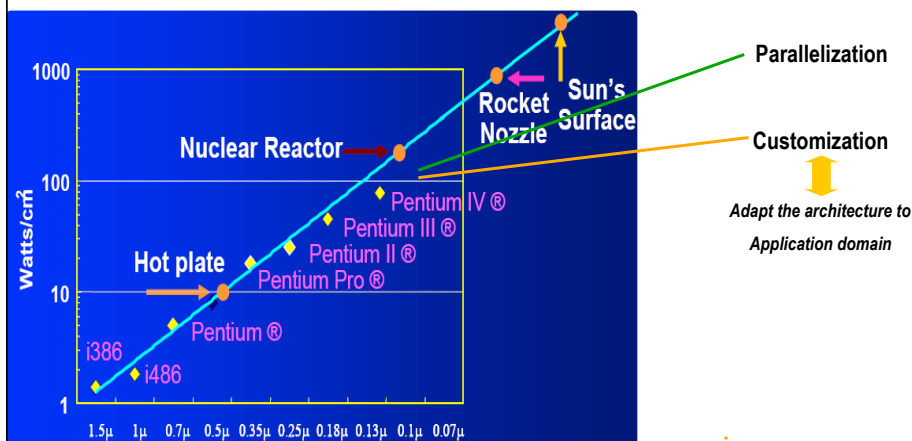
- ◆ Datacenters consume ~ 2% of all US electricity (2006)
- ◆ Data center annual growth (15%) is unsustainable
- ◆ Datacenter power projected to be > 8% of US power by 2020
- ◆ Need a paradigm shift in data center computing to put us on a more sustainable/scalable ICT energy efficiency curve



Courtesy Massoud Pedram

5

## Next Big Opportunity – Customization and Specialization



6

## Justification 1 – Potential of Customization

◆ AES 128bit key 128bit data	◆ Throughput	◆ Power	◆ Figure of Merit (Gb/s/W)
0.18μm CMOS	>3.84 Gbits/sec	350 mW	11 (1/1)
FPGA [1]	1.32 Gbit/sec	490 mW	2.7 (1/4)
ASM StrongARM [2]	31 Mbit/sec	240 mW	0.13 (1/85)
Asm Pentium III [3]	648 Mbits/sec	41.4 W	0.015 (1/800)
C Emb. Sparc [4]	133 Kbits/sec	120 mW	0.0011 (1/10,000)
Java [5] Emb. Sparc	450 bits/sec	120 mW	0.0000037 (1/3,000,000)

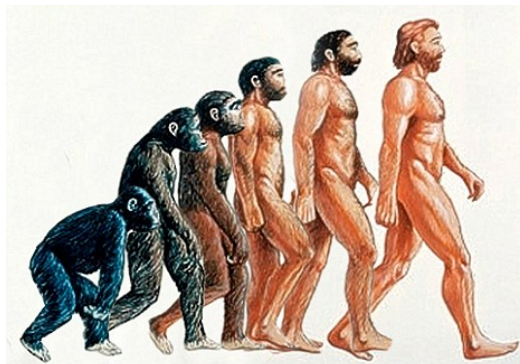
- ◆[1] Amphion CS5230 on Virtex2 + Xilinx Virtex2 Power Estimator
- ◆[2] Dag Arne Osvik: 544 cycles AES – ECB on StrongArm SA-1110
- ◆[3] Helger Lipmaa PIII assembly handcoded + Intel Pentium III (1.13 GHz) Datasheet
- ◆[4] gcc, 1 mW/MHz @ 120 Mhz Sparc – assumes 0.25 u CMOS
- ◆[5] Java on KVM (Sun J2ME, non-JIT) on 1 mW/MHz @ 120 MHz Sparc – assumes 0.25 u CMOS

Source: P. Schaumont, and I. Verbauwhede, "Domain specific codesign for embedded security," IEEE Computer 36(4), 2003.

7

## Justification 2 -- Advance of Civilization

- ◆ For human brain, Moore's Law scaling has long stopped
  - The number neurons and their firing speed did not change significantly
- ◆ Remarkable advancement of civilization via specialization
  - ◆ More advanced societies have higher degree of specialization



8

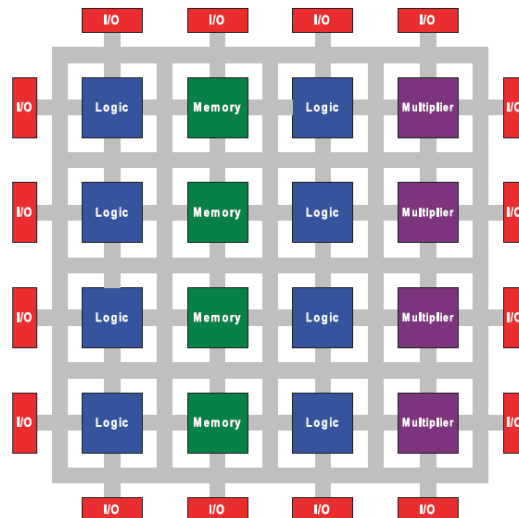
## Goal of CDSC (Center for Domain-Specific Computing)

- ◆ **A general, customizable platform**
  - Can be customized to a wide-range of applications
    - May focus on one or several given domains
  - Can be massively produced with cost efficiency
  - Can be programmed efficiently with novel compilation and runtime systems
- ◆ **Metric of success**
  - A “supercomputer-in-a-box” with 100X performance/power improvement via customization for the intended domain(s)
- ◆ **Supported by the NSF Expeditions in Computing Program (2009)**

9

## Example of Customizable Platforms: FPGAs

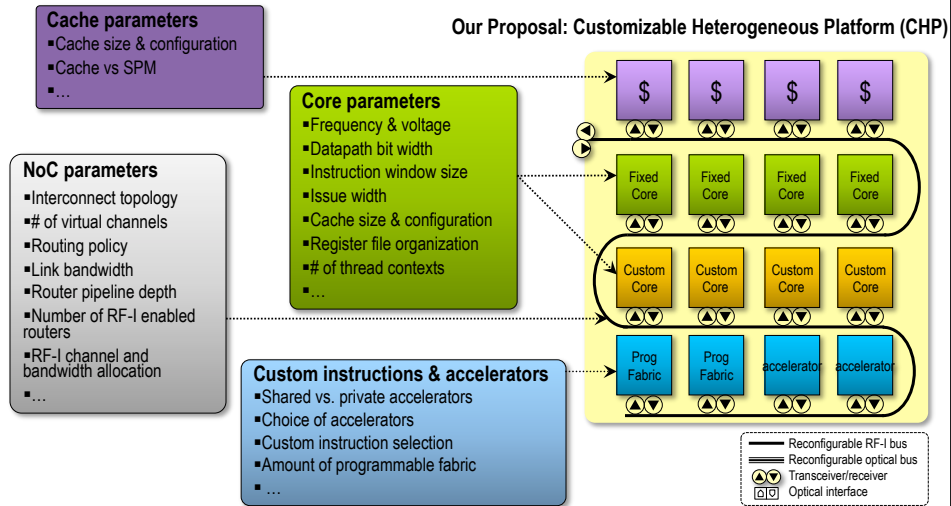
- ◆ **Configurable logic blocks**
- ◆ **Island-style configurable mesh routing**
- ◆ **Dedicated components**
  - Specialization allows optimization
  - Memory/Multiplier
  - I/O, Processor
  - Anything that the FPGA architect wants to put in!



Source: I. Kuon, R. Tessier, J. Rose. FPGA Architecture: Survey and Challenges. 2008.

10

## More Opportunities for Customization to be Explored



**Key questions: Optimal trade-off between efficiency & customizability**  
**Which options to fix at CHP creation? Which to be set by CHP mapper?**

11

## Center for Domain-Specific Computing (CDSC)



Aberle  
(UCLA)



Baraniuk  
(Rice)



Bui  
(UCLA)



Chang  
(UCLA)



Cheng  
(UCSB)



Cong (Director)  
(UCLA)



Palsberg  
(UCLA)



Potkonjak  
(UCLA)



Reinman  
(UCLA)



Sadayappan  
(Ohio-State)



Sarkar  
(Associate Dir)  
(Rice)



Vese  
(UCLA)

12

## ***Examples of Customization***

---

- ◆ Customization of processor cores
- ◆ Customization of on-chip memory
- ◆ Customization of on-chip interconnects

13

## ***Example 1 – Customization of Cores***

---

- ◆ Large cores or small cores?
- ◆ How many each type?

14

## Core spilling – [Cong et al Trans. on Parallel and Distributed Systems 2007]

◆ **CMP systems focus on improving overall throughput**

- Sequential or legacy applications might not see benefits

◆ **Key idea – allow execution to be spilt from one core to next at run-time**

- Simulate increase in register file, instruction queue, ROB and LSQ size
- Allocate cores intelligently to spilling core

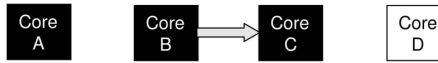
Core A's resources are exhausted.



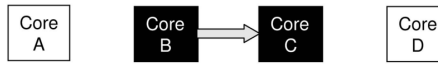
Core A sends 62 compiler visible registers, a 24-entry store buffer, the LSQ entries of any stores that are in-flight, and a PC.



Core B begins fetching from the PC sent by Core A. Core A continues execution, and sends any register or store values that were in-flight at the time of the spill as they complete.



Spilling can continue if B's resources are exhausted.



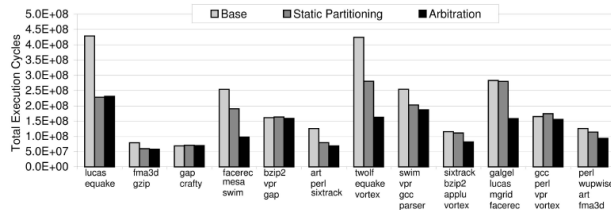
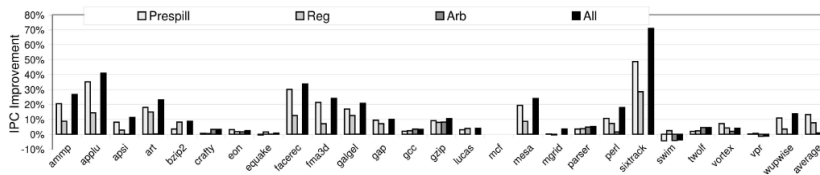
Eventually, all instructions on Core A will commit (unless there is an exception or branch misprediction) and Core A can be released into the pool of idle cores.

15

## Core spilling – [Cong et al Trans. on Parallel and Distributed Systems 2007]

◆ **Results**

- Core spilling achieves more than 50% of the performance of 'ideal' 32-issue core by using 4-issue cores for single applications
- 39% improvement for multiple application workload
- Up to 40% reduction in latency for changing workloads



16



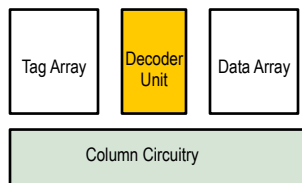
## Example 2: Customization of On-Chip Memory

- ◆ HW controlled cache or SW controlled cache (SPM)?
- ◆ How much to allocate for each type?

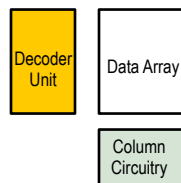
17

## Customizable Hybrid L1 Cache [ISLPED'2011]

- Cache in conjunction with Scratchpad Memory (SPM) in L1
  - Cache: Hardware-controlled
    - Transparent to software: a fast local copy of the global memory address space
  - SPM: Software-controlled
    - Not transparent to software: a separate address space from the global address space
- Customizable
  - Flexibly size the cache and SPM based on the application requirements
    - Cache: dynamic/random access
    - SPM: regular data access pattern



18 (a) Cache memory organization



(b) SPM organization

	Cache	SPM
<b>access time</b>	hit : 1 cycle miss : L cycles	1 cycle
<b>energy</b>	4.57 nJ	1.53 nJ

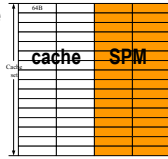
(c) Comparison (2KB)

18

## How to Customize?

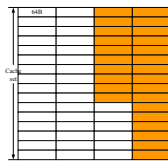
### ◆Way-wise reconfigurable cache

- Configure several ways of cache as SPM
- Column cache [Chiou et.al. DAC'00]



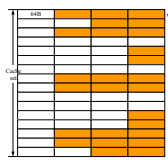
### ◆Block-wise reconfigurable cache

- Virtual local store [Cook et.al. UCB TR'09]
- Unified mapping of SPM blocks onto cache blocks

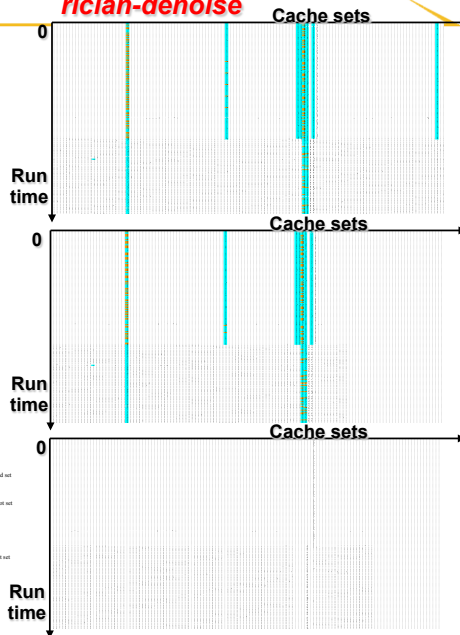


### ◆Adaptive hybrid cache (AH-Cache)

- Dynamically remap SPM blocks from high-demand cache sets to low-demand cache sets.



*rician-denoise*



## Impact of Adaptation

### ◆Design points for comparison:

- Design N:** Non-adaptive hybrid cache, baseline
- Design B:** N + Balanced cache [Zhang, ISCA'06]
- Design Vp:** N + Victim cache [Jouppi, ISCA'90]
- Design Vs:** N + A serially accessed victim cache
- Design R:** Phase-reconfigurable hybrid cache [Zhang, et.al., ISLPED'02]
- Design AH:** AH-Cache
- Design S:** Static optimized hybrid cache (not practical, just check the optimality gap of AH)

### ◆Improvements of AH-Cache

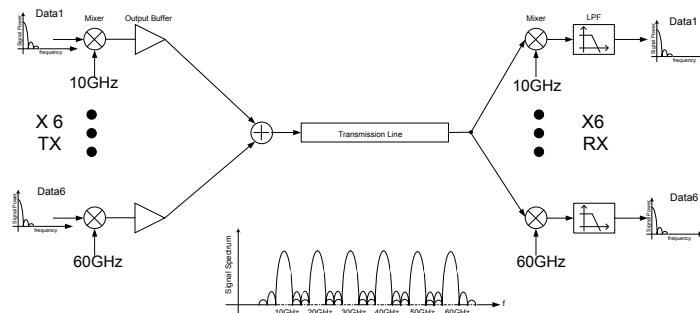
	<i>N</i>	<i>B</i>	<i>Vp</i>	<i>Vs</i>	<i>R</i>	<i>S</i>
Miss rate	52%	19%	22%	22%	33%	-1%
Run-time	18%	3%	4%	8%	12%	~0%
Energy	13%	16%	22%	10%	7%	-1%
ED product	33%	19%	25%	18%	18%	~0%

### Example 3: Customization of On-Chip Interconnects

- ◆ How many wires to include for on-chip communication?
- ◆ Uniform distribution and dedicated connections?

21

### Our Answer : Use of Multiband RF-Interconnect for Customization



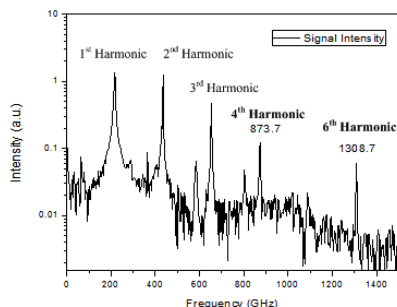
- In TX, each mixer up-converts individual baseband streams into specific frequency band (or channel)
- N different data streams (N=6 in exemplary figure above) may transmit simultaneously on the shared transmission medium to achieve higher aggregate data rates
- In RX, individual signals are down-converted by mixer, and recovered after low-pass filter

22

22

## Terahertz VCO in 65nm CMOS

Measured signal spectrum with uncalibrated power



higher harmonics (4th and 6th harmonics) may be substantially underestimated due to excessive water and oxygen absorption and setup losses at these frequencies.

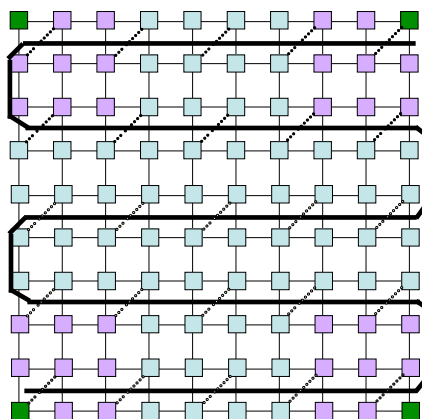
- ◆ Demonstrated an ultra high frequency and low power oscillator structure in CMOS by adding a negative resistance parallel tank, with the fundamental frequency at 217GHz and 16.8 mW DC power consumption.
- ◆ The measured 4<sup>th</sup> and 6<sup>th</sup> harmonics are about 870GHz and 1.3THz, respectively.

“Generating Terahertz Signals in 65nm CMOS with Negative-Resistance Resonator Boosting and Selective Harmonic Suppression”  
Symposium on VLSI Technology and Circuits, June 2010

23

## Mesh Overlaid with RF-I [HPCA '08]

- ◆ 10x10 mesh of pipelined routers
  - NoC runs at 2GHz
  - XY routing
- ◆ 64 4GHz 3-wide processor cores
  - Labeled aqua
  - 8KB L1 Data Cache
  - 8KB L1 Instruction Cache
- ◆ 32 L2 Cache Banks
  - Labeled pink
  - 256KB each
  - Organized as shared NUCA cache
- ◆ 4 Main Memory Interfaces
  - Labeled green
- ◆ RF-I transmission line bundle
  - Black thick line spanning mesh

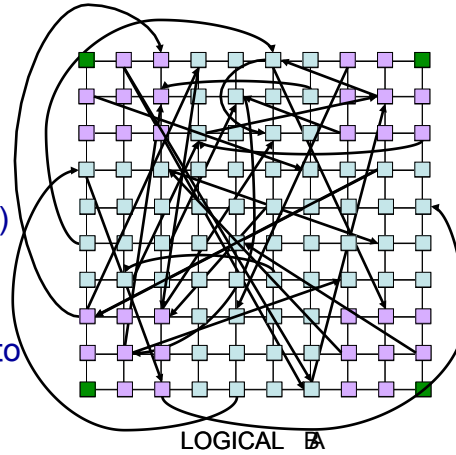


24

24

## RF-I Logical Organization

- Logically:
  - RF-I behaves as set of N express channels
  - Each channel assigned to src, dest router pair (s,d)
- Reconfigured by:
  - remapping shortcuts to match needs of different applications



25

25

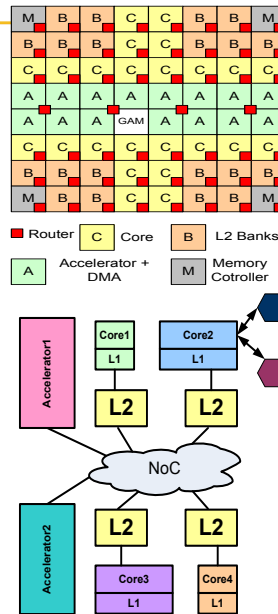
## Specialization is also Important

- ◆ Specialized accelerators used to be considered “wasteful”
  - A story
- ◆ Van Neumann architecture maximizes device reuse
- ◆ Utilization wall
  - 6.5% utilization for a 45 nm chip filled with 64bit operators, assuming a power budget of 80 W [ASPLOS’2010]

26

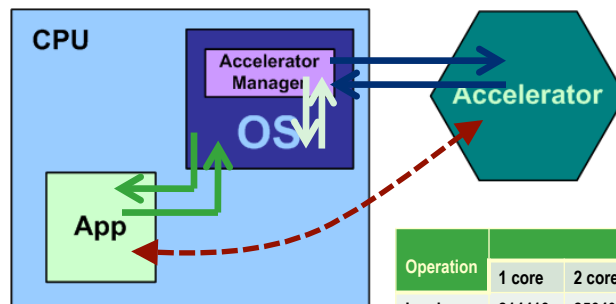
## Our Proposal: Extensive Use of Accelerators [SAW'2011]

- ◆ Proposed solution: extensive use of accelerators (customized or implemented using programmable fabric)
  - Sea of accelerators
- ◆ Type of accelerators:
  - Tightly vs. loosely coupled
- ◆ Benefits
  - Better performance
  - Higher power-efficiency
  - It's ok to be "wasteful"
- ◆ Critical needs:
  - Efficient accelerator management
    - Scheduling
    - Sharing



27

## Using Accelerators with OS Management



Operation	Latency (# Cycles)				
	1 core	2 cores	4 cores	8 cores	16 cores
Invoke	214413	256401	266133	308434	316161
RD/WR	703	725	781	837	885

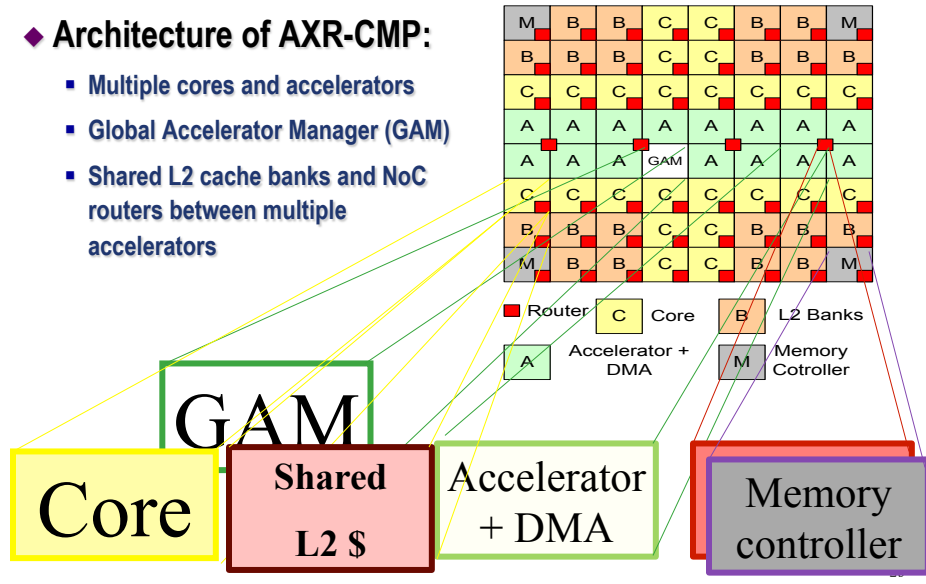
- ◆ Managing accelerator by OS is expensive
- ◆ In an accelerator rich CMP, management should be cheaper both in terms of time and energy

28

## Overall Architecture of AXR-CMP

### Architecture of AXR-CMP:

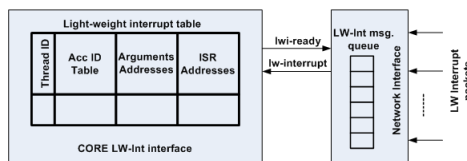
- Multiple cores and accelerators
- Global Accelerator Manager (GAM)
- Shared L2 cache banks and NoC routers between multiple accelerators



## Light-weight Interrupt Support

- To reduce OS interrupt service
  - No need to save context
- Two main components added:
  - A table to store ISR info
  - An interrupt controller to queue and prioritize incoming interrupt packets
- Each thread registers:
  - Address of the ISR and its arguments
  - lw-int source
- Sources of lw-int:
  - GAM responses
    - Accelerator ready
    - Wait time for accelerator
  - Accelerator TLB miss
  - Accelerator task buffer empty
- Limitations:
  - Only can be used when running the same thread which LW interrupt belongs to
  - OS-handled interrupt otherwise

Operation	Latency (# Cycles)				
	1 core	2 cores	4 cores	8 cores	16 cores
Interrupt	16383	20361	24022	26572	28574



<i>lwi-reg x y z</i>	Register service routine <i>y</i> to service interrupts arriving from accelerator <i>x</i> . LWI message packet will be written to <i>z</i> .
<i>lwi-jump</i>	Jump to the service routine for the next pending interrupt. Does nothing if no interrupt is pending.
<i>lwi-ret</i>	Return from an interrupt service routine.

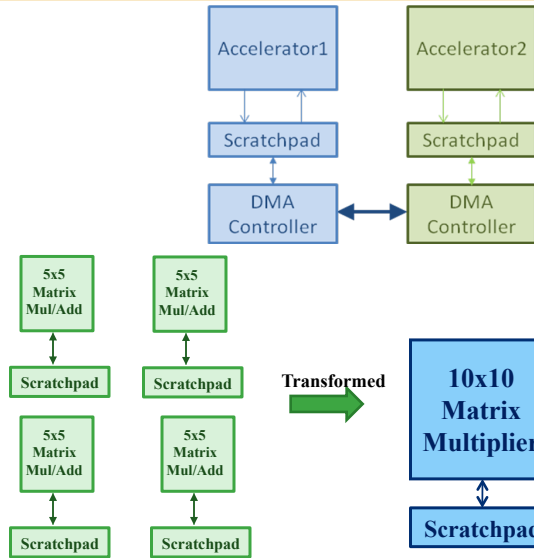
## Accelerator Chaining and Composition

### ◆ Chaining

- To have an efficient accelerator to accelerator communication

### ◆ Composition

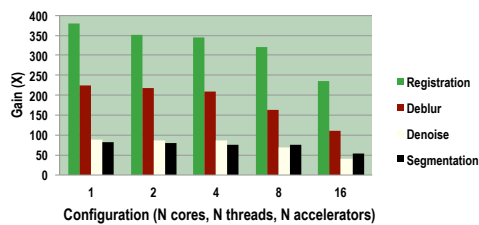
- To create the virtual feeling of having larger accelerators for the applications



31

## Experimental Results – Performance (N cores, N threads, N accelerators)

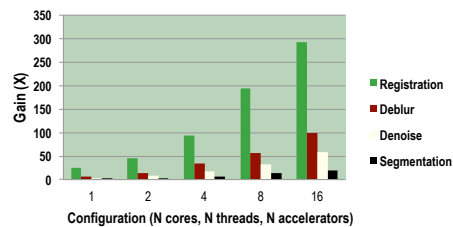
Speedup over SW-Only



Performance improvement over SW only approaches: on average 168X, up to 380X

Performance improvement over OS based approaches: on average 51X, up to 292X

Speedup over OS-based

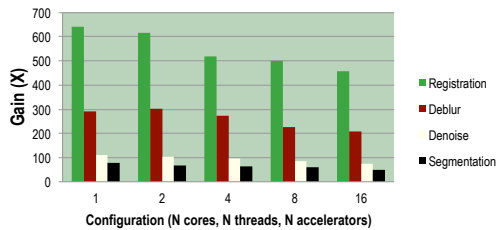


32



## Experimental Results – Energy (N cores, N threads, N accelerators)

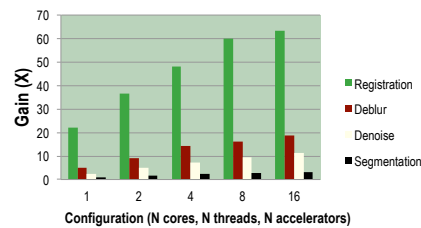
Energy gain over SW-only version



Energy improvement over SW only approaches: on average 241X, up to 641X

Energy improvement over OS based approaches: on average 17 X, up to 63X

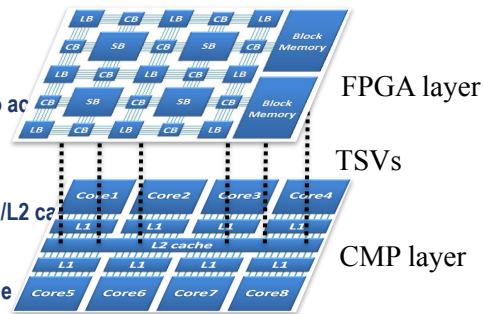
Energy gain over OS-based version



33

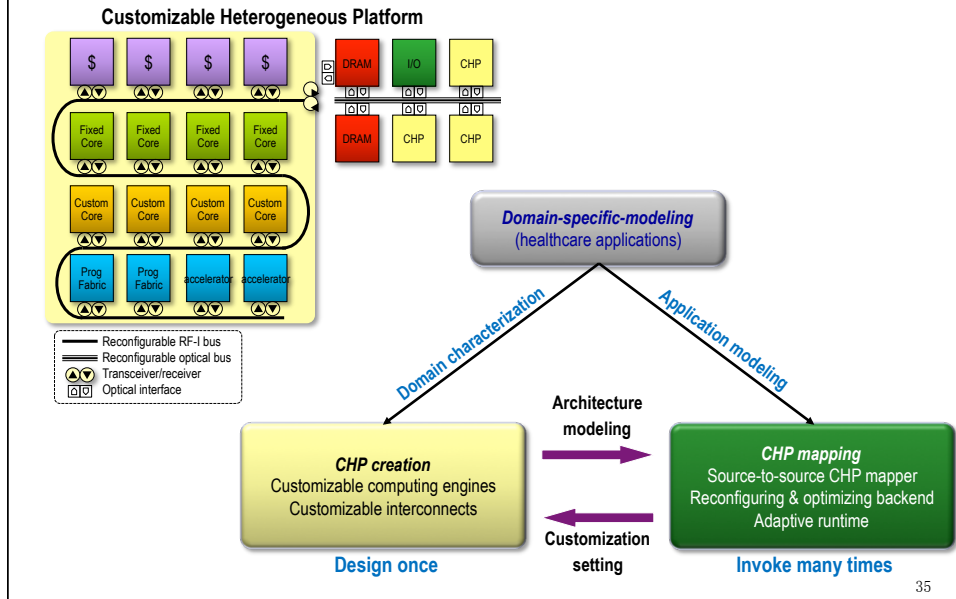
## 3D Integration for Customization or Specialization

- ◆ Vertical integration:
  - CMP layer + customization/acceleration layer
- ◆ Accelerators can directly access caches
  - L1 or L2
- ◆ Low latency
  - 1 cycle traversal across the TSV bundle
  - 2-3 cycles to get from the TSV bundle to ac controller
- ◆ Higher bandwidth
  - Almost equal to the bandwidth of the L1/L2 cache
- ◆ No single bottleneck
  - Each cache can have its own TSV bundle
    - Sharing TSV bundles possible
- ◆ Early results: medical imaging benchmarks [ASAP'2011]
  - > 7x performance gain
  - > 18x energy gain



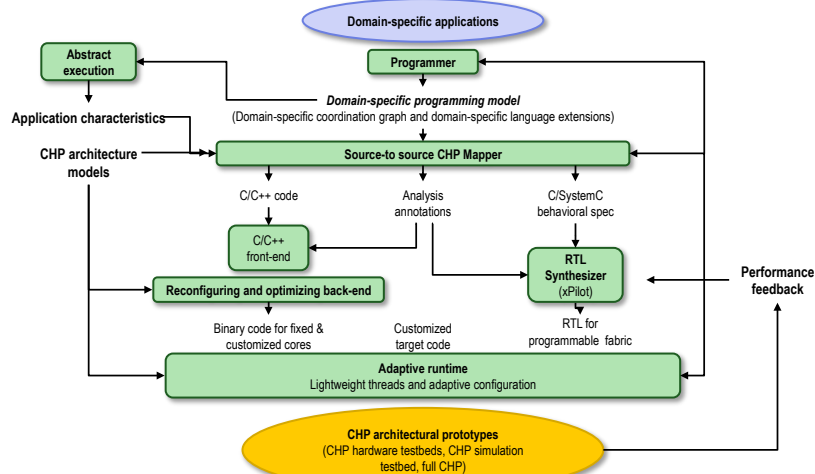
34

## Research Scope in CDSC (Center for Domain-Specific Computing)

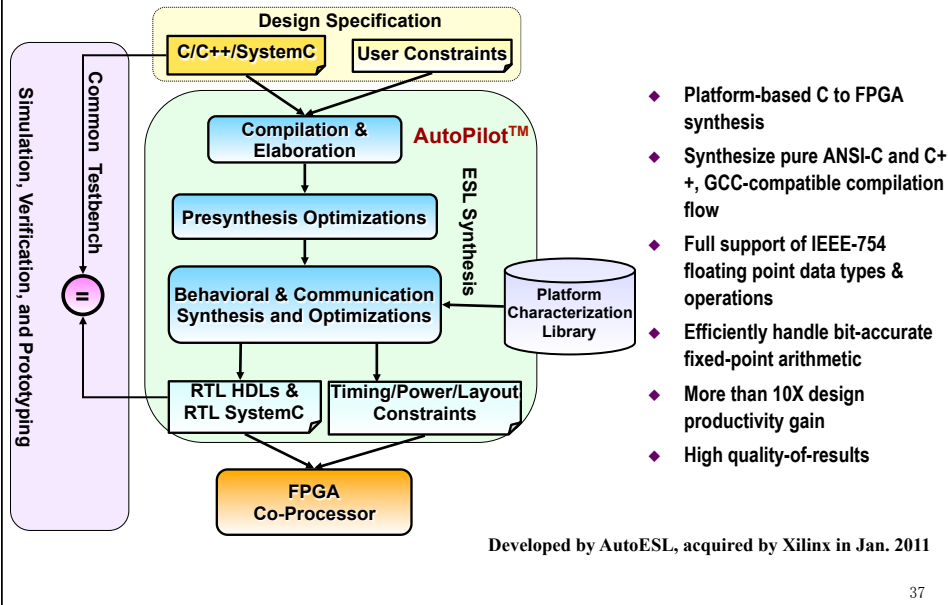


## CHP Mapping – Compilation and Runtime Software Systems for Customization

**Goals: Efficient mapping of domain-specific specification to customizable hardware**  
 – Adapt the CHP to a given application for drastic performance/power efficiency improvement



## AutoPilot Compilation Tool (based UCLA xPilot system)



- ◆ Platform-based C to FPGA synthesis
- ◆ Synthesize pure ANSI-C and C+, GCC-compatible compilation flow
- ◆ Full support of IEEE-754 floating point data types & operations
- ◆ Efficiently handle bit-accurate fixed-point arithmetic
- ◆ More than 10X design productivity gain
- ◆ High quality-of-results

37

## UCLA /PKU Joint Research Institute (JRI) in Science & Engineering

- ◆ Started in 2009
- ◆ Mission and activities
  - Research
    - Promote joint collaborative research
    - Jointly seek research funding from government and private sources
    - Facilitate technology transfer of collaborative research results
  - Education
    - Jointly train students as future leaders with global perspective
    - Host exchange faculty and students
    - Facilitate the sharing of teaching materials
  - Other collaborations
    - Host and/or sponsor international workshops and conferences
    - Homes for UCLA alumni in Beijing, and PKU alumni in So. Cal
- ◆ 74 faculty from UCLA are affiliated
- ◆ 59 faculty from PKU are affiliated

38

## The 1<sup>st</sup> JRI Research Symposium

May 6 Thursday

### Morning Opening Session

*Provost Scott Waugh (UCLA), Provost Jianhua Lin (PKU),  
Nick Entrikin, Jason Cong (UCLA), Xiaoming Li (PKU)*

### Plenary Session

- Next Generation Biofuels,  
*James Liao, Chemical and Biomolecular Engineering, UCLA*
- The Strength of Drug Discovery at Peking University,  
*Zhen Yang, Chemistry and Molecular Engineering, PKU*
- New Algorithms in Information Science,  
*Stanley Osher, Department of Mathematics, UCLA*

### Afternoon Parallel Sessions

- Clean Energy
- Information Technologies
- Life Sciences and Medical Sciences

39

## Friday, May 7, 2010

- ◆ 8:30am - 9am: Breakfast (CNSI Presentation Space, 5th floor)
- ◆ 9am - 10:30am: Theme-wide discussions about possible collaboration topics. (CNSI Presentation Space, 5th floor)
- ◆ 10:45am – 11:30am: Report back to the entire group (CNSI Presentation Space, 5th floor)
- ◆ 11:30am – 12:30 noon: Overview of CNSI (Paul Weiss) and tour of CNSI
- ◆ 12:30noon – 1:30pm: Lunch (CNSI Lobby, 3rd floor)
- ◆ 1:30pm – 5:00pm: Visit various research labs in UCLA for further discussions about research collaboration.

40

## The 2<sup>nd</sup> JRI Research Symposium

April 25 2011, PKU

### Morning Opening Session

President Zhou (PKU), Chancellor Block (UCLA)  
Xiaoming Li (PKU), Jason Cong (UCLA)

### Plenary Session

- Center for Excellence in Neuroscience and Recent Advances in Vision Function  
Professor Nicholas Brecha, UCLA
- Clean Energy – research, applications and impact in China  
Professor Dongxiao Zhang, College of Engineering, PKU
- Recent Advances in Electrical Engineering and Its Applications at UCLA  
Professor Frank Chang, UCLA
- Challenges and Opportunities on Chinese e-Health  
Professor Shan Wang, Renmin Hospital, PKU

41

## The 2<sup>nd</sup> JRI Research Symposium

April 25 2011, PKU

### Afternoon Parallel Sessions

- Clean Energy, Meeting Room 1
- Information Technology, Meeting Room 6
- Life Sciences and Basic Medical Sciences, Meeting Room 7
- Translational Medicine, Meeting Room 4

42

## The 2<sup>nd</sup> JRI Research Symposium

Lab Visit on April 26 2011, PKU

### Wireless Communications Lab (Anpeng Huang)

Morning: Kung Yao (before 11:30am), Frank Chang

Location: Room 2427W, Science Building No.2

### State Key Lab on Advanced Optical Communication Systems and Networks (Anpeng Huang)

Morning: Frank Chang

Location: Room 2427W, Science Building No.2

### Institute of Molecular Medicine (Heping Cheng, Rui-Ping Xiao)

Afternoon: Peipei Ping, Ren Sun

Location: 2<sup>nd</sup> floor, Ying Jie Overseas Exchange Center

### State Key Laboratory of Protein and Plant Gene Research (Li-Ping Wei)

11am: Eleazar Eskin, afternoon: Peipei Ping

Location: Room 608, Life Science Building

### Nano-electronic Devices and Integrate Circuits Research Group (Ru Huang)

10am: Ya-Hong Xie, Jason Cong, Frank Chang

Location: Room 1544, Science Building No.1

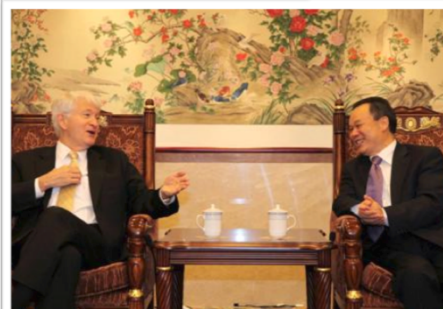
### Prof. Xiaodong Hu's group

Afternoon: Ya-Hong Xie (before 3:30pm)

Location: Room South 118, Physics Building

43

## The 2<sup>nd</sup> JRI Research Symposium



## Center for Energy-efficient and Applications (CECA) at Peking University

Center for Energy-efficient Computing and Applications

**ABOUT US**

**Center for Energy-Efficient Computing and Applications (CECA)** in Peking University is a multi-disciplinary research center in Peking University established in 2010 focused on energy-efficient computing and communication as well as their application. It is a special institute in the School of EECS, and is given considerable flexibility in offering faculty compensation and startup packages, evaluating for faculty advancement, and setting up a world-class research environment, with the goal of recruiting and retaining for the best talents in its research areas internationally. The research area in CECA includes energy efficient computing (including but not limited to energy-efficient computing and communication architectures, compilation and system-level software, systems from embedded to data-center scale) and applications (including but not limited to smart grid, mobile computing, and hardware acceleration of computing-intensive applications, such as bioinformatics and design automation).

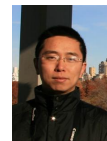
The initial scope of CECA includes the following research directions:

- Novel energy-efficient technology-driven computing architectures and platforms, including exploration of novel architectures enabled by 3D-IC technologies and non-volatile memory technologies.
- Novel energy-efficient system architectures, including the research on customizable domain-specific architectures and associated compilation and runtime management software for customization and virtualization
- Exploration of new energy-efficient applications, such as smart electrical grids, customized computing for bioinformatics, sensor networks and Internet of Things

The lead & director of CECA is Prof. [Jason CONG](#). Dr. Cong received his B.S. degree in computer science from Peking University in 1985, his M.S. and Ph. D. degrees in computer science from the University of Illinois at Urbana-Champaign in 1987 and 1990, respectively. Currently, he is a Chancellor's Professor at the Computer Science Department of University of California, Los Angeles, Associate Vice Provost for Technology and Director of Center for Energy Efficient Computing, funded by NSF, DARPA, and Intel.

## Leadership and Faculty at PKU CECA

- ◆ **Director: Jason Cong, "Thousand Talents Professor"**
  - Chancellor's Professor at UCLA
  - Co-Director of PKU/UCLA Joint Research Institute in Science and Engineering
  - Research interests: synthesis and layout of VLSI circuits, highly scalable VLSI design algorithms and tools, design and synthesis of programmable circuits and systems, computer architectures, and system-on-a-chip designs
- ◆ **Executive Director: Tao Wang, associate research professor**
  - Assistant dean at College of Engineering, PKU (2006)
  - Intel Research Lab in China (2006 – 2010)
  - Research interests: parallel computing, computer architecture, reconfigurable logic and reconfigurable computing
- ◆ **Guojie Luo – assistant professor**
  - Ph.D., UCLA (US), 2011
  - Second Prize in the routability-driven placement contest at ISPD 2011
  - Research Interests: physical design automation, 3-D IC technology, linear & nonlinear Optimizations
- ◆ **Guangyu Sun – assistant professor**
  - Ph.D., Penn State (US), 2011
  - Research Interest: high performance, low power, and high reliability memory system, 3D architecture, process-variation aware architecture, 3D VLSI Design, CAD Tool Development



46

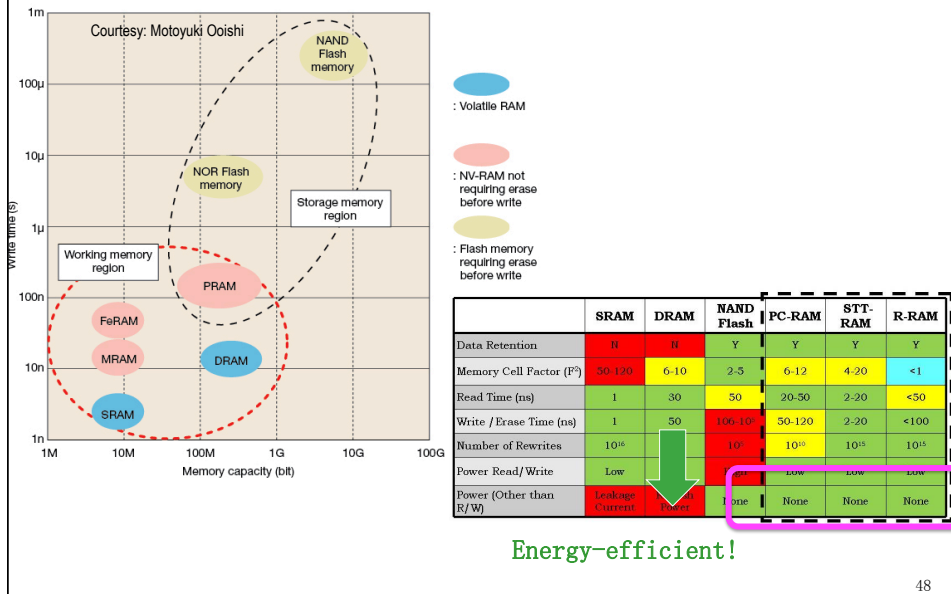
## Adjunct Professors

- ◆ Yuanyuan Zhou, Qualcomm Endowed Chair Professor at UCSD
- ◆ Research interests: operating systems, computer architecture and programming languages
- ◆ Songwu Lu, Professor at UCLA
- ◆ Research interests: wireless networking, cloud computing, mobile systems, wireless network security and Internet security
- ◆ Yuan Xie, Associate Professor at Pennsylvania State University
- ◆ Research interests: VLSI Design, Electronics Design Automation, Computer Architecture, Embedded Systems Design



47

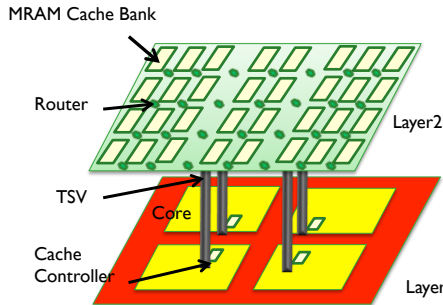
## Heterogeneity and Customization Via Non-Volatile Memory (Sun and Xie, PennState and PKU CECA)



48



## Example: 3D Stacked MRAM L2 Cache for Multi-core



◆ 3D enables cost-effective heterogeneous integration for multi-core with MRAM

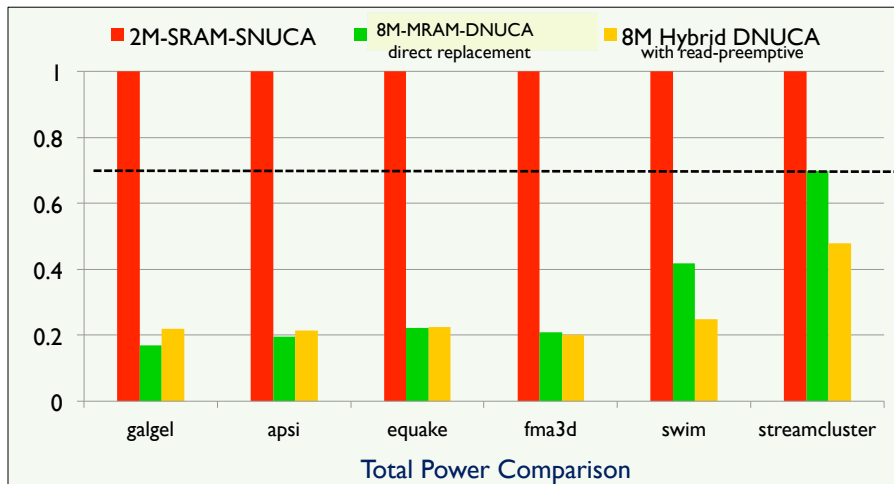
◆ For the same footprint, MRAM-based L2 cache has

- 4 times density capacity (better performance with lower miss rate)
- ~10 times smaller leakage power (lower power)

Cache configurations	Leakage power
2MB SRAM cache	2.09W
8MB MRAM cache	0.26W

“A Novel Architecture of the 3D Stacked MRAM L2 Cache for CMPs”, (Sun et al., HPCA 2009)<sup>49</sup>

## Power Comparison (SRAM vs. MRAM L2 Stacked Cache)

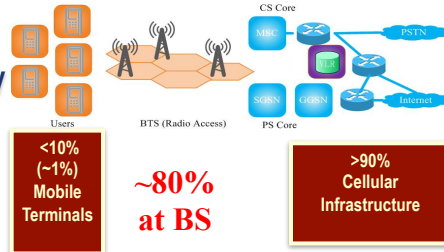


“A Novel Architecture of the 3D Stacked MRAM L2 Cache for CMPs”, (Sun et al., HPCA 2009)<sup>52</sup>

## Toward "Green" 3G Network Infrastructure (Songwu Lu, UCLA and PKU CECA)

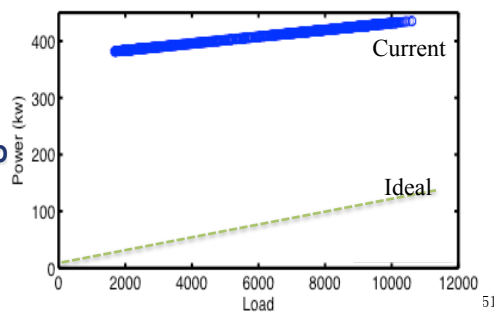
### ◆ Large energy consumption

- (720TWh) ~0.5% world electricity cellular infrastructure [ICT' 09]
- Rising at 16~20% per year



### ◆ Problem: Non-Energy Proportionality (EP)

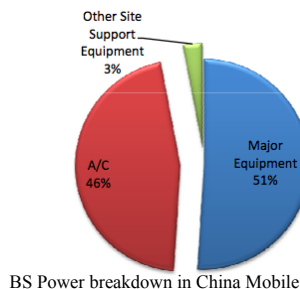
- Energy is not proportional to traffic load



## Root Cause for Energy Non-EP

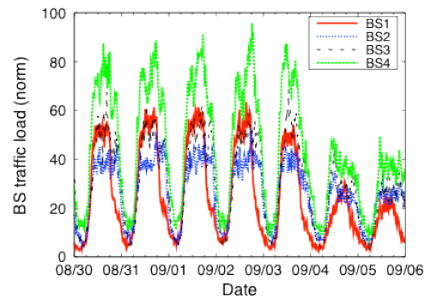
### ◆ Each BS is non-EP

- Cooling power is about 50%



### ◆ Traffic highly dynamic

- Fluctuate over time
- Spatially diverse @ BS



Large energy overhead even for light traffic => non-EP

## Proposed Solution [ACM Mobicom '11]

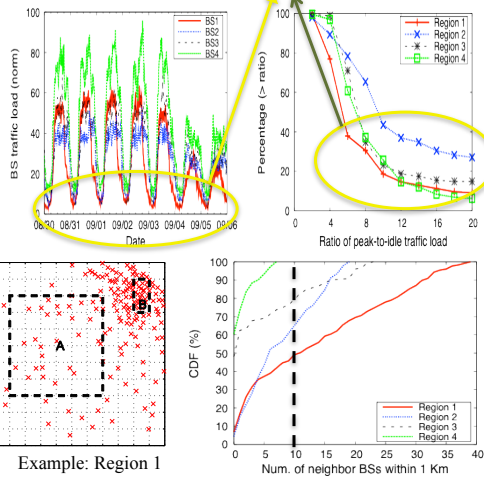
- ◆ Exploit diversity to turn off BSes for system-wide energy efficiency

- Diversity is everywhere
- Diversity helps energy save

- ◆ Significant savings in 4 3G regional networks

- ◆ Using operational 3G traces
- ◆ 23% savings in rural areas
- ◆ 46~53% in dense cities

### Large saving potential for quiet hours



### Rich BS redundancy ensures coverage

## Global Challenges Addressed by International Collaboration

- ◆ UCLA proposes to create a global clean energy research center (CERC-LA)
  - Focus of research: developing smart grid technologies that can be deployed in the United States, China and other global markets
  - Also serving as the international collaboration arm for existing centers

## Research Themes

### ◆ Theme I: Robust Energy Systems

- ◆ Stochastic modeling (including data mining) and control (including pricing and demand response) for grid with deep penetration of renewable energy and electric vehicles;
- ◆ Reliability and cyber-security for software-hardware co-designed systems used in grid and other energy systems
- ◆ Sensing, communication, and control circuits and systems for reliability, robustness, and high voltage and high power

### ◆ Theme II: Nanoscale Energy Engineering

55

## Research Themes

### ◆ Theme I: Robust Energy Systems

### ◆ Theme II: Nanoscale Energy Engineering

- ◆ Materials and devices for energy harvesting/storage
- ◆ Materials and devices for integrated high voltage power electronics and for super conductivity

### ◆ Theme III: Broader Impacts

- ◆ Interaction and integration of grid, electric vehicle and massive public transportation such as high-speed railway, and demonstration and data exchange (between nations) of the above integration
- ◆ IP protection, technology transfer, and environmental, societal and policy issues

56

## International Research Consortia

- ◆ Start with US-China Consortium, and extend globally in the future
- ◆ Led by UCLA ECRC and State Grid of China (SGCC)
  - SGCC, serving 1 billion people in China and no. 7 of Global Fortune 500
- ◆ Founding members include Peking Univ. (PKU) and Fudan Univ. in China

57

## 2010 Summer Research Program of UCLA Students at PKU

Name	Research Title
Robert Cunningham	An Investigation into Gallium-Nitride Type III Semiconductors
Peichi (Justin) Liu	GaN Growth with Low Dislocation Density on Patterned Sapphire Substrate
Drew Morton	Survival of Melanopsin Expressing ipRGCs in Thy1-CFP DBA/2j Mice, a model for Glaucoma
Chunyi Peng	EPS: Find Parking on the Go
Katherine Rosen	Using Combinatorial RNA Interference to Elucidate the Genetic Interactions of Sir-2.1 in <i>Caenorhabditis elegans</i>
Shirley Shui	Functional Study of Critical Proteins, Neuroglobin and DREAM, in Primary Cultures of <i>Mus musculus</i> Neural Cells
Zhanyong Wang	Systematic Meta-analysis of Genetic Susceptibility Underlying Autism
Sheng Wei	FPGA Implementation of HD Video Deblocking Filter Using C-to-Hardware Synthesis
Allison Wong	The Isolation and Identification of Bioactive Constituents of <i>Ephedra sinica</i>

58

## Research at PKU



59

## Life in Beijing

Tiananmen Square



The Great Wall

60

## Life in Beijing



The World Cup

Chinese Food!

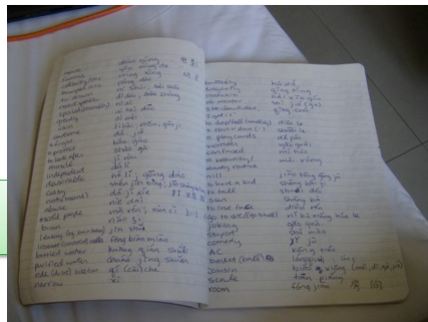


## Chinese Class



Chinese Name

Notes

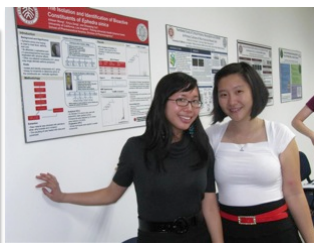
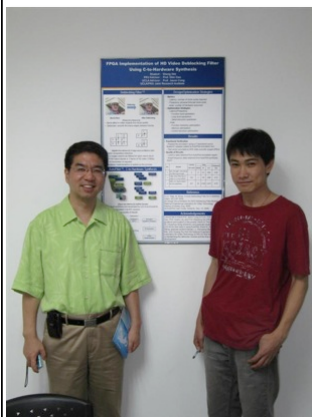


## Presentation at the Closing Ceremony (Aug. 25, 2010)



63

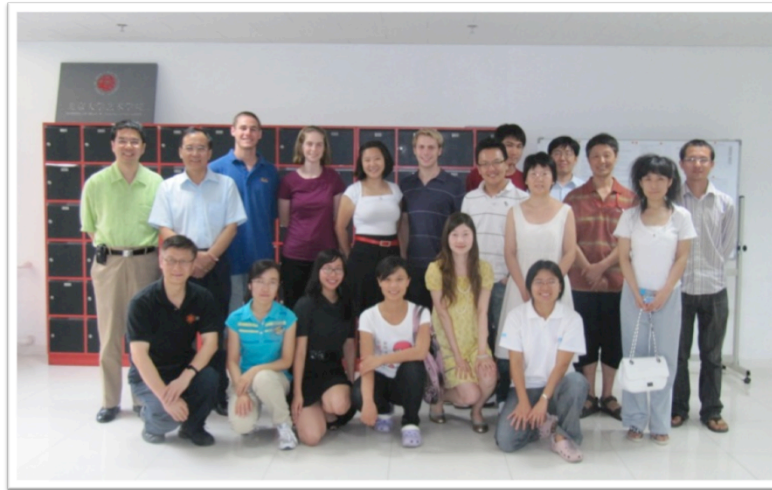
## And the Poster Session ...



64



## A Happy and Fruitful Summer



65

## UCLA Today Article (Nov. 24, 2010)

Home

All Stories

Top Stories

People

Voices

Around Campus

Faculty Bulletin Board

Archives

Contact Info

### Quick Links

Faculty Resources

Human Resources News

UCLA Today → All Stories

Share:

Nov 24, 2010 • By Alison Hewitt

[Print this Page](#) [Mail this Page](#)

### Rickety bikes and cutting-edge research

Over the summer a handful of UCLA science students cruised around Beijing on rickety bikes, learned to read menus in Chinese and teamed up with research labs at Peking University for a one-of-a-kind experience doing cutting-edge research abroad.

The UCLA-Peking University Joint Research Institute (JRI) has begun linking not only students, but also faculty, from both universities with international research opportunities.

Tuesday night (Nov. 23), the students shared details about how their 10 weeks abroad gave them a more global perspective — and a love of “skewer guys,” street-food vendors selling meat skewers from grills on the back of their bikes.

Globalization is making familiarity with foreign countries, particularly China, vital for preparing students for their careers, said UCLA Computer Science Professor Jason Cong, co-director of the



Beijin Alison Wong, right, with her Peking

66

## Feedback from Students

<http://www.pku-jri.ucla.edu/education/exchange-2010-UCLA.asp>

### Feedback from the Students



Shirley Sui,  
Undergraduate student, Medical School

"I learned what it is like to do research in a lab in Beijing - which is not all that different in most ways - and what it is like to live in Beijing. PKU's students are not unlike my peers at UCLA, and conducting research at PKU has rewards and frustrations much like those you would experience in a lab at UCLA. I've learned a lot about through PKU-UCLA JRI's Summer Research opportunity, and I am extremely grateful for this chance to experience both aspects of being a student on the other side of the globe."



Katherine Rosen,  
Undergraduate student, Biochemistry

"I would highly recommend this program to any student who has been dreaming of traveling abroad, but doesn't want to sacrifice a summer's worth of research! Over the summer, I learned what it was like to live in a country where I did not speak the primary language (at all); however, I was able to do so in a supportive environment since the other UCLA students in the program and the students in my lab at PKU spoke English. Through conversations with my lab-mates, I was able to learn about Chinese culture not just from a textbook but from students my age who have lived there their entire lives. Trying to interact with people in shops and restaurants while not speaking the same language was also a humbling experience, but I was able to get around and am pleased to be coming back with the confidence that I could live happily in China if job prospects take me there. In either case, I will definitely be making a return trip!"



Allison Wong,  
Undergraduate student, Biochemistry

"I had an amazing experience, both culturally and with my research. I really appreciated the cultural exchange aspect. I cannot emphasize enough how enriching and valuable I found the language skills and cultural experience to be - that's a type of learning I could not really experience at any research institute in the US."

67

## Applicants to 2011 Summer Research Program at PKU

◆ SCHOOL LEVEL	DEPARTMENT	Count
▪ GRAD 6	Biochemistry	2
▪ JUNIOR 3	Biology	1
▪ SENIOR 21	Biophysics	1
▪ SOPHOMORE 1	Chemical Engineering	2
	Chemistry	1
	Chemistry-Material Science	1
	Civil Engineering	2
	Computer Science	3
	Electrical Engineering	10
	Environmental Science	1
	Molecular, Cell, & Developmental Biology	1
	Physics	1
	Physiological Science	4
	Statistics	1
▪ Grand Total 31	Grand Total	31

68

## 2011 Summer Research Program at PKU

Student Name	Major	Level	PKU Advisor
Tammy Chang	Electrical Engineering	SENIOR	Anpeng Huang
David Cohen	Electrical Engineering	SENIOR	Hongbin Zha
Yasaman Demehri	Physiological Science	SENIOR	Rui-ping Xiao
Xin Guan	Chemistry and Biochemistry	SENIOR	Kai Wu
Hien Huynh	Electrical Engineering	SENIOR	Shao-qing Cai
Niloufar Iranmanesh	Physiological Science	SENIOR	Don Zhang
Brandon Lanthier	Civil Engineering	SENIOR	Zhen Yang
Ian McRae	Chemical Engineering	JUNIOR	Shao-qing Cai
Arefeh Oroujhi	Physiological Science	SENIOR	Gui-qiang Wang
Jamie Tran	Molecular, Cell, & Developmental Bio	SENIOR	John Wang
Daniel Wen	Electrical Engineering	JUNIOR	Gang Huang
Vincent Tse	Physiological Science	SENIOR	Qi-huang Gong
Bingjun Xiao	Electrical Engineering	GRAD	Ru Huang
Defeng Xu	Computer Science	GRAD	Xiaoming Li

69

## 2011 UCLA Students for Summer Research Program at PKU (with Chinese language instructors)



(Left to Right) Susan Jain (Confucius Institute), Defeng Xu, Hien Huynh, You Lu (Instructor), Daniel Wen, Tammy Chang, Niloufar Iranmanesh, Daphne Lee (Instructor), Arefeh Oroujhi, David Cohen, Brandon Lanthier, Yasaman Demehri, Vincent Tse, Xin Guan, Jamie Tran, Ian McRae, Bingjun Xiao, Xiaojie Ma (Instructor), Lailai Zhuang (Confucius Institute)

70

## ***Concluding Remarks***

---

- ◆ **Despite of end of scaling, there is plenty of opportunity with customization and specialization for energy efficient computing**
- ◆ **Green IT should be a global effort**
- ◆ **Green IT enables many other green applications**

71

## ***Acknowledgements***

---

- ◆ **Support from National Science Foundation**
  - Expeditions in Computing Program
- ◆ **CDSC research team**
  - Thanks to all my co-PIs in four universities – UCLA, Rice, Ohio-State, and UC Santa Barbara
- ◆ **Support from Peking University**

72