

# 2022 IEEE TCCA Young Computer Architect Award

Yakun Sophia Shao

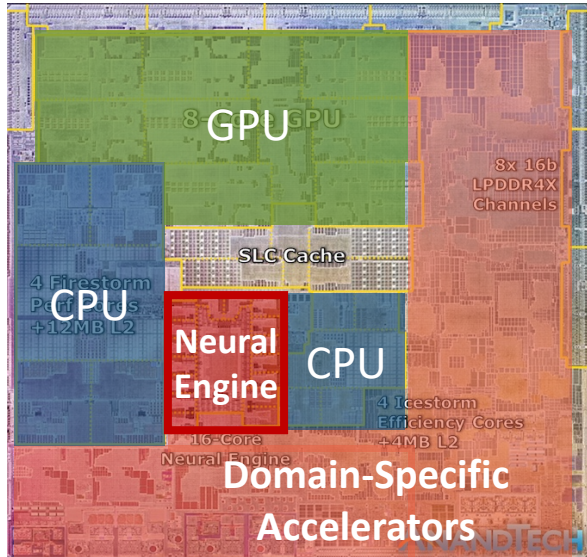
ysshao@berkeley.edu

Electrical Engineering and Computer Sciences



# Domain-Specific Accelerators

- Specialized hardware designed for a domain of applications.



- ❄️ 2% Paper at ISCA
- ❄️ No major development in industry



- ☀️ 30% Paper at ISCA
- ☀️ Wide adoption from edge to cloud



**Research Theme:** Enable **energy-efficient** computing with **specialized hardware**, while improving **flexibility** and **design productivity**.



# Modeling Accelerators



Michael Lyons

“We are unable to evaluate systems with multiple accelerator stores *due to the difficulty of obtaining a large number of distinct accelerators.*”

- Back in 2009, the architecture community still largely focused on **multicore architecture**.
- There was **no standard modeling and simulation flow for accelerators**.

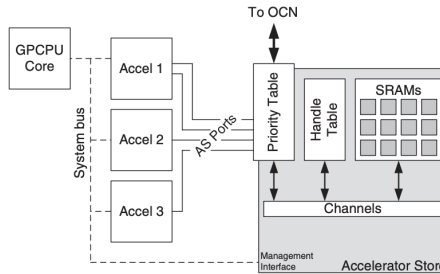
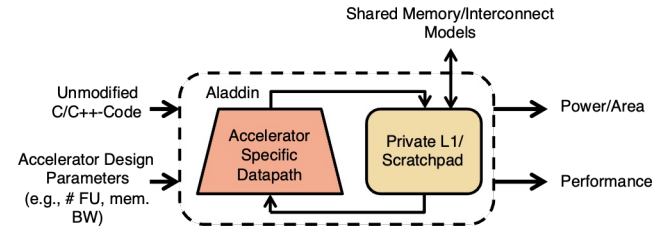


Fig. 4. Accelerator store design.



## Aladdin: A Power-Performance Simulator for Hardware Accelerators



“Accelerator Simulator”  
Design Accelerator-Centric  
System

“Design Assistant”  
Understand Accelerator  
Design Space

- Aladdin [ISCA’2014, **Top Picks**] (<https://github.com/harvard-acc/aladdin>)
- WIICA [ISPASS’2013] (<https://github.com/ysshao/WIICA>)
- MachSuite [IISWC’2014] (<https://github.com/breagen/MachSuite>)
- gem5-Aladdin [MICRO’2016] (<https://github.com/harvard-acc/gem5-aladdin>)
- Timeloop [ISPASS’2019] (<https://github.com/NVlabs/timeloop>)



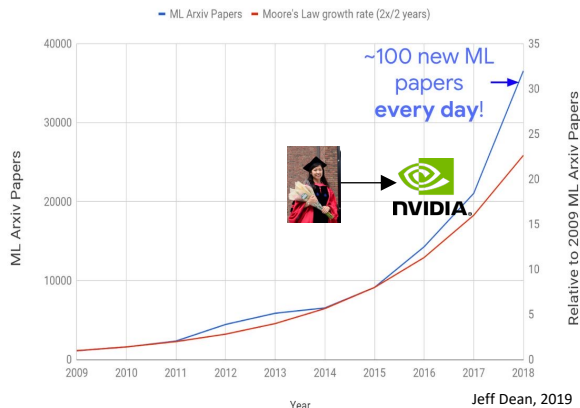
Brandon Reagen



Sam Xi

# Designing Accelerators

Machine Learning Arxiv Papers per Year



Jason Clemons



Rangha Venkatesan



Brian Zimmer

How can we deliver **scalable**  
ML performance with  
**chiplet-based** architecture?

## SIMBA: SCALABLE MCM-BASED ARCHITECTURE

Simba Package and Chiplet

Package and chiplet spec

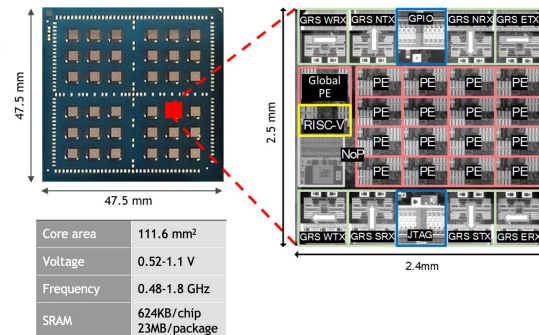
6mm<sup>2</sup> chiplet in TSMC 16nm  
36 chiplets/package

Chip-to-chip interconnect

Ground-Referenced Signaling

Efficient compute tiles

128 TOPS  
0.11 pJ/Op  
8-bit integer datapath



### Architecture

- Simba [MICRO'2019, **Best Paper Award**, CACM Research Highlight]

### Circuit

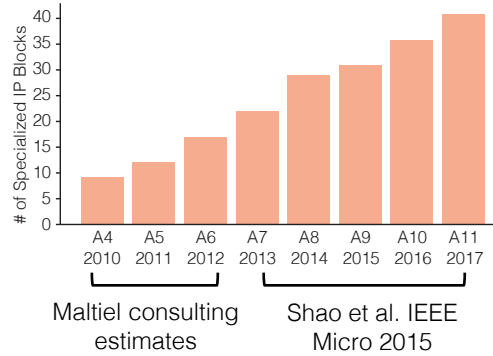
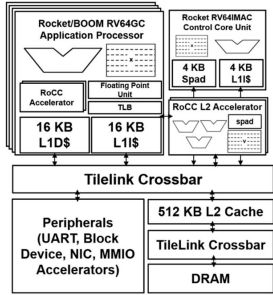
- RC18 [VLSI'2019, JSSC'2020 **Best Paper Award**]

### Methodology

- MAGNet [HotChips'2019, ICCAD'2019] (<https://github.com/NVlabs/matchlib>)

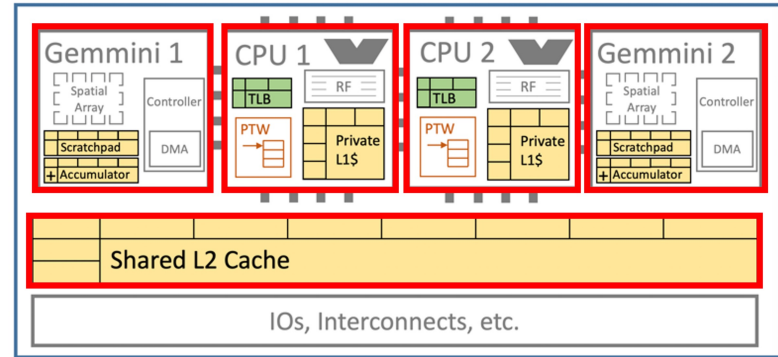


# Integrating Accelerators

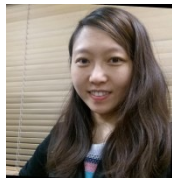


SoC

## Gemmini: Full-System DNN Integration



Hasan Genc

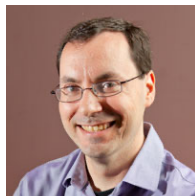
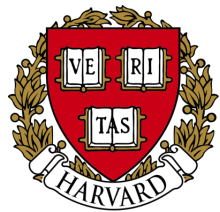


Jenny Huang

How can we better support system-level integration and programming of accelerators?

- Gemmini [DAC'2021, **Best Paper Award**]
  - <https://github.com/ucb-bar/gemmini>
- CoSA [ISCA'2021]
  - <https://github.com/ucb-bar/cosa>

# Thank you!



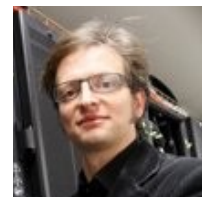
David Brooks



Gu-Yeon Wei



Harvard VLSI-Arch Group



Lieven Eeckhout



Bill Dally



Joel Emer



Brucek Khailany



Steve Keckler



NVIDIA Research



Krste Asanovic



David Patterson



Bora Nikolic



UC Berkeley ADEPT/Slice Lab