Cross-cutting Panel 3

*What computational models and architectures enable the nanotech-to-applications connection?*

Nanotechnology-Inspired Information Processing Systems of the Future Workshop

Washington D.C.

8/31/16
Panelists

Gert Cauwenberghs
UCSD, Bioengineering
Neuromorphic Computing

Luis Ceze
UW Seattle, CSE
Systems: Architecture, OS
Programming Languages

Lav Varshney
UIUC, ECE
Information and
Systems Theory

Sharad Malik
Princeton EE
System Design Methodology
From Applications to Fabrics

Inference Applications

Virtual Reality

Applications: Several algorithms
Algorithm: Several kernels

Figure source: N. Verma
Picture source: ultravr.org
From Applications to Fabrics

• Relevant application metrics
  • Statistical throughput/latency?
  • Other QoR metrics?
  • Energy per task/result?
  • BW per task/result?
  • Other resource efficiency metrics?
From Applications to Fabrics

• Emerging computation models
  • Statistical?
  • Probabilistic?
  • Neuro based?
  • Integration of hybrid models?
From Applications to Fabrics

- Microarchitecture components
  - Novel implementations of $X[t+1] = f(X[t])$
  - New state elements
  - New functional accelerators
  - New combinations
    - In-memory computing
    - In-sensor computing
From Applications to Fabrics

- Microarchitecture models
  - Error models?
  - Statistical models?
    - Functional
    - Speed
    - Energy
  - Derivation from device models?

Application Metrics
Computational Models
Microarchitecture Models
Device Models
From Applications to Fabrics

- Programming interfaces
  - Balance application specification needs with ease of mapping to microarchitectures
    - Instruction granularity
    - Appropriate intermediate representations
  - Mapping techniques

- Application Metrics
- Computational Models
- Programming Interface
- Microarchitecture Models
- Device Models
From Applications to Fabrics

- Overarching issues
  - Specialization vs. generalization
    - Amortizing platform costs
    - Amdahl’s law like benefit limits
  
- Maximally leveraging existing application to device stacks

Application Metrics
Computational Models
Programming Interface
Microarchitecture Models
Device Models
• Relevant application metrics
  • Statistical throughput/latency?
  • Other QoR metrics?
  • Energy per task/result?
  • BW per task/result?
  • Other resource efficiency metrics?

• Microarchitecture models
  • Error models?
  • Statistical models?
    • Functional
    • Speed
    • Energy
  • Derivation from device models?

• Emerging computation models
  • Statistical computation models?
  • Probabilistic computation models?
  • Integration of hybrid models?

• Microarchitecture components
  Novel implementations of $X[t + 1] = f(X[t])$
  • New state elements
  • New functional accelerators
  • New combinations
    • In-memory computing
    • In-sensor computing

• Overarching issues
  • Specialization vs. generalization
    • Amortizing platform costs
    • Amdahl’s law like benefit limits
  • Maximally leveraging existing application to device stacks

• Programming interfaces
  • Balance application specification needs with ease of mapping to microarchitectures
    • Instruction granularity
    • Appropriate intermediate representations
  • Mapping techniques

• Programming interfaces
  • Balance application specification needs with ease of mapping to microarchitectures
    • Instruction granularity
    • Appropriate intermediate representations
  • Mapping techniques