Research Infrastructures for Accelerator-Centric Architectures

David Brooks, Mark Hempstead, Mike Lui, Parnian Mokri, Siddharth Nilakantan, Brandon Reagen, Yakun Sophia Shao
# Tutorial Outline

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 am – 9:00 am</td>
<td>Introduction</td>
</tr>
<tr>
<td>9:00 am – 10:00 am</td>
<td>Pre-RTL Simulation Framework: Aladdin</td>
</tr>
<tr>
<td>10:00 am – 10:30 am</td>
<td>Break</td>
</tr>
<tr>
<td>10:30 am – 11:00 am</td>
<td>Workload Characterization Tool: WIICA</td>
</tr>
<tr>
<td>11:00 am – 12:00 pm</td>
<td>CAD &amp; Benchmarks: HLS &amp; MachSuite</td>
</tr>
<tr>
<td>12:00 pm – 2:00 pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>2:00 pm – 3:00 pm</td>
<td>Embedded Keynote Talk: Mark Horowitz (Stanford)</td>
</tr>
<tr>
<td>3:00 pm – 3:30 pm</td>
<td>Accelerator Selection Tool: Sigil</td>
</tr>
<tr>
<td>3:30 pm – 4:00 pm</td>
<td>Break</td>
</tr>
<tr>
<td>4:00 pm – 5:00 pm</td>
<td>Hands-on Exercise</td>
</tr>
</tbody>
</table>
CMOS Technology Scaling

![Graph showing the performance of different technologies over time.](image)
CMOS Technology Scaling

Technological Fallow Period
Potential for Specialized Architectures

[Brodersen and Meng, 2002]
Beyond Homogeneous Parallelism

In Core

Out of Core

Programmability

Composable Accelerators

SIMD/SSE

AESDEC

GPU

Energy Efficiency

Fixed Function

H.264

Fixed Function Mon
Cores, GPUs, and Accelerators: Apple A8 SoC

[Die photo from Chipworks]
[Accelerators annotated by Sophia Shao @ Harvard]
Cores, GPUs, and Accelerators: Apple A8 SoC

Out-of-Core Accelerators

[Die photo from Chipworks]
[Accelerators annotated by Sophia Shao @ Harvard]
Challenges in Accelerators

• Flexibility
  – Fixed-function accelerators are only designed for the target applications.
Composable Customization

Monolithic Hardware Accelerator
Composable Customization

Composed Accelerator with sub-blocks
Composable Customization

Composed Accelerator w/ Architectural Support

Shared Interconnect and Memory Fabric
Composable Customization

Composed Accelerator w/ Architectural Support

Example: “Accelerator Store”
Lyons et al. TACO’12
Composable Customization

Composed Accelerator w/ Architectural Support

Shared Interconnect and Memory Fabric
Composable Customization

Composed Accelerator w/ Architectural Support

Shared Interconnect and Memory Fabric

Composable Accelerators Provide Application Flexibility
Is this general-purpose?

General-purpose workloads are far less regular
Is this general-purpose?

General-purpose workloads are far less regular

Shared Interconnect and Memory Fabric
Is this general-purpose?

General-purpose workloads are far less regular

Microcontrollers (Cortex-M, Siskiyou Peak)

Programmable Fabrics

Shared Interconnect and Memory Fabric

Embed programmability into the accelerator fabric!
Challenges in Accelerators

• Flexibility
  – Fixed-function accelerators are only designed for the target applications.

• Programmability
  – Today’s accelerators are explicitly managed by programmers.
Today’s SoC

OMAP 4 SoC

http://www.anandtech.com/show/4551/motorola-droid-3-review-third-times-a-charm/10
Challenges in Accelerators

• Flexibility
  – Fixed-function accelerators are only designed for the target applications.

• Programmability
  – Today’s accelerators are explicitly managed by programmers.

• Design Cost
  – Accelerator (and RTL) implementation is inherently tedious and time-consuming.