A 10-core SoC with 20 Fine-Grain Power Domains for Energy-Proportional Data-Parallel Processing over a Wide Voltage and Temperature Range

Thomas Benz, Luca Bertaccini, Florian Zaruba, Fabian Schuiki
Frank K. Gürkaynak, Luca Benini
Integrated Systems Laboratory (IIS)
ETH Zürich, Switzerland
{tbenz, lbertaccini, zarubaf, fschuiki, kgf, lbenini}@iis.ee.ethz.ch
Outline

- Introduction
- Architecture
  - Overview
  - ISA Extensions
  - Power Gating
- Use cases
  - Datacenter
  - Extreme Edge
- Chip Results
- Comparison with SoA
- Conclusion
Introduction

- Energy efficiency: dominant factor for next-gen systems
  - Scaling does not improve leakage

- Leakage is an increasing problem
  - Increased silicon area + heterogeneity
  - Higher integration density
  - High amount of dark silicon

- Leakage: no contribution to useful work
  - Wasted energy
  - Reduction of useful TDP

- A concern across the compute continuum
  - Datacenter
  - Extreme edge

Equi-area chip: power 1.24x up node-to-node!
Challenges & Contributions

- **Challenges**
  - **Fine-grain power gating** in manycore @ low overhead (area, timing) is **non-trivial**
  - Usually: go for **coarse-grain** domains
    - limits **energy proportionality**

- **Contributions**
  - Manycore architecture with ISA extensions
    - Tuned for **energy efficiency**
    - Optimized **μ-architecture** to minimizes state
  - First **fine-grain** power-manged RISC-V multi-core chip
    - Approach energy proportionality
    - Implemented and validated over a **wide freq. – temp. range**
    - Present **sub-10ns** wake-up time of power domains
## Architecture Overview

- **Always on domain**
  - Microcontroller
    - Governor
    - Management responsibilities
  - Peripherals

- **Cluster**
  - 8+1 Snitch Cores
    - Single-stage RISC-V
  - Core-Complex: core + FPU + IPU
  - 64kiB local scratchpad memory
  - ISA extensions
    - SSR
    - FREP
  - Fine-grain power gating
Snitch – highly power-manageable core

- Snitch Core Complex
  - Simple single-stage Snitch core
  - Control core
  - Low amount of logic
  - IPU, FPU functional units

84% of the area of a Core Complex is power gated
Snitch – highly power-manageable core

- **Snitch Core Complex**
  - Simple single-stage Snitch core
    - **Control core**
    - **Low** amount of logic
  - IPU, FPU functional units

- **Goal:** Utilize functional units $>80\%$
  - **Superscalar out-of-order** cores
    - Large amount of **state & deep pipeline** stages in func. units
    - **Highly adverse to power management!**
  - Our solution: **simple** RISC-V ISA extensions
    - **SSR:** streaming semantic registers
    - **FREP:** hardware loop
    - Only **minimal** architectural state in power gated region
    - Algorithm using SSR, FREP: only **temporary** data
    - Optimized microarchitecture: only 4 pipeline stages
Power Gating Granularity

- **AoD Domain**
  - Governor Core
  - Not power gated
  - Focus on cluster

- **Cluster domain**
  - **Coarse-grain**
  - Entire Cluster

- **Functional Units**
  - **Fine-grain**
  - Individual FPU domains
  - Individual IPU domains
Power Control

- Power control module
  - Memory-mapped register interface
  - Finite state machine
  - **Programmable** sequence
  - Only 11.4 kGE

- Header power gates
  - Mother – Daughter
  - Reduce peak inrush current

- Isolation
  - Prevent hardware from injecting wrong transfers
Power Control Sequencing

- 4-stage process
  - Timig configurable
  - **Single-write** power toggle

- Mother – Daughter delay
  - Most critical
  - Mitigate spike in inrush current

- Power toggle speed
  - 3 AoD clock cycles
  - ~6 cluster clock cycles
  - **Sub-10 ns**
Use Case: Datacenter

- Cluster-centric application

- 2 usecases for power gating:
  - IPU / FPU workloads are not mixed
    - Gate unused unit type
  - Memory-bound regimes
    - Gate stalling units

- Control is done by data movement core
  - Already used for data orchestration
  - Insight in xPU utilization
  - Decentralized, scalable control
Use Case: Datacenter - Results

- 0.9V, 75°C, running at 850 MHz

- Running xPU workloads
  - Variable arithmetic intensity

- FPU Workloads
  - Power gate IPU units: 6.5% power reduction
  - Gate stalling units
    - In fully memory-bound region (I): up to 13.1%

- IPU Workloads
  - Power gate FPU units: 8.0% power reduction
  - Gate stalling units
    - In fully memory-bound region (I): up to 14.0%
    - Higher relative gain: IPUs consume less power
Study: Manticore

- **Manticore Architecture***
  - 1024 cores on a chiplet
  - Organized in 128 clusters
  - HBM2 memory interface
  - 4 chiplets: Manticore System

- **AXPY FPU workload**
  - Power gate all IPU units
  - 1/12 SP FLOP / Byte
  - 65 cluster fully gated
  - 63 cluster: 1 FPU active

- **41.4% power reduction**
  - Coarse- and fine-grain power gating
  - 15.18W to 8.96W

Use Case: Extreme Edge

- Microcontroller
  - Self-contained unit
  - 1 management core: AoD **Governor**
  - 8+1 core GP compute accelerator

- Use case for power gating
  - Sporadic need for **massive** compute
  - **Gate cluster fully** during **idle**
  - Need to reduce power-on transition time
  - Near-threshold for dynamic efficiency
    - Worsens **ratio** of leakage
    - Slower clock: **less cycles** for transition
Use Case: Extreme Edge - Results

- **Nominal operating point**
  - Edge node
  - 0.6V, 25°C. Near-threshold
  - Up to 42% power reduction

- **Extended temperature range**
  - Automotive applications e.g.
  - 0.6V, 65°C
  - Up to 65% power reduction

- Near-threshold operation **remains energy efficient** also for **advanced nodes** and **high** operating temperature
Chip Results

<table>
<thead>
<tr>
<th>Technology</th>
<th>GF 22nm FDSOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip Area</td>
<td>1.56 mm²</td>
</tr>
<tr>
<td>VDD Range</td>
<td>0.6V – 0.9V</td>
</tr>
<tr>
<td>Memory size</td>
<td>64kIB L1, 24kIB L2</td>
</tr>
<tr>
<td>Logic Transistors</td>
<td>6 MGE</td>
</tr>
<tr>
<td>Frequency Range</td>
<td>32kHZ – 950MHz</td>
</tr>
<tr>
<td># Controllable Power Domains</td>
<td>18 fine-, 1 coarse-grain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power Domain</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always on Domain</td>
<td>1.52 MGE</td>
</tr>
<tr>
<td>Cluster</td>
<td>2.69 MGE</td>
</tr>
<tr>
<td>FPU</td>
<td>57.0 – 58.7 kGE</td>
</tr>
<tr>
<td>IPU</td>
<td>31.3 – 32.2 kGE</td>
</tr>
</tbody>
</table>
## Comparison With SoA

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Application</strong></td>
<td>IoT</td>
<td>IoT</td>
<td>IoT/HPC</td>
<td>HPC</td>
<td>HPC</td>
</tr>
<tr>
<td><strong>Technology</strong></td>
<td>28nm FSOI</td>
<td>22nm FDSOI</td>
<td>22nm FDSOI</td>
<td>28nm FDSOI</td>
<td>7nm FinFET</td>
</tr>
<tr>
<td><strong>Die Size</strong></td>
<td>4.5 mm²</td>
<td>12 mm²</td>
<td>1.56 mm²</td>
<td>7.84 mm²</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cores/ISA</strong></td>
<td>32b Async RISC 32b / RISC-V</td>
<td>1 + 9 cores / 32b RISC-V</td>
<td>1 + 9 cores / 32b RISC-V</td>
<td>2 cores / 64b RISC-V</td>
<td>48 + 4 cores / Armv8-A</td>
</tr>
<tr>
<td><strong>Accelerator</strong></td>
<td>ML / Crypto</td>
<td>4 shared FPUs / HWCE</td>
<td>8 FPUs / 8 IPUs</td>
<td>Hwacha Vector</td>
<td>SVE / 512 bit SIMD</td>
</tr>
<tr>
<td><strong>Maximum Frequency</strong></td>
<td>350 MHz</td>
<td>450 MHz</td>
<td>910 MHz</td>
<td>475 MHz</td>
<td>2.2 GHz</td>
</tr>
<tr>
<td><strong>Voltage Range</strong></td>
<td>0.45V - 0.9V</td>
<td>0.5V - 0.8V</td>
<td>0.6V - 0.9V</td>
<td>0.55V - 1.1V</td>
<td>-</td>
</tr>
<tr>
<td><strong>On-chip SRAM (State Retention)</strong></td>
<td>464kB</td>
<td>128kB (L1) / 1600kB s.r. (L2)</td>
<td>64kB (L1) / 24kB s.r. (L2)</td>
<td>256kB</td>
<td>32MB (L2 Cache)</td>
</tr>
<tr>
<td><strong>Power Management</strong></td>
<td>Clock &amp; Power Gating</td>
<td>Clock &amp; Power Gating</td>
<td>Clock &amp; Power Gating</td>
<td>DVFS &amp; Body-Biasing</td>
<td>Power Gating &amp; DVFS</td>
</tr>
<tr>
<td><strong>Granularity</strong></td>
<td>On-Demand Unit</td>
<td>Cluster / SoC / Memories</td>
<td>Cluster / IPUs / FPUs</td>
<td>Tile (Core)</td>
<td>FPU Lane</td>
</tr>
<tr>
<td><strong>Best INT Performance</strong></td>
<td>1.5 GOPS</td>
<td>15.6 GOPS (8-bit)</td>
<td>6.8 GOPs (32-bit)</td>
<td>Not Available</td>
<td>Not Available</td>
</tr>
<tr>
<td><strong>Best FP Performance</strong></td>
<td>-</td>
<td>2 GFLOPS (FP32)</td>
<td>13.6 GFLOPS (FP32)</td>
<td>Not Available</td>
<td>3.4 TFLOPS (FP64)</td>
</tr>
<tr>
<td><strong>Reaction Time</strong></td>
<td>207ns</td>
<td>not available</td>
<td>10ns</td>
<td>&lt;2us</td>
<td>few ms</td>
</tr>
<tr>
<td><strong>Energy Efficiency</strong></td>
<td>230 GOPS/W @ 110 MOPS (int8)</td>
<td>79 GFLOPS/W @ 1 GFLOPS</td>
<td>118 GFLOPS/W @ 7.2 GFLOPS</td>
<td>19.6 GFLOPS/W</td>
<td>16.9 GFLOPS/W (FP64)</td>
</tr>
</tbody>
</table>
Conclusion

- **Thestral**: 10-core chip with 20 power domains
  - RISC-V manycore chip
  - Custom, light-weight ISA extensions
  - Agile, sub-10 ns, aggressive, fine-grain power management

- Wide range of applications
  - **IoT**: microcontroller with GP compute acceleration
  - **Datacenter**: compute cluster with decentralized, fine-grain power control

- Datacenter: up to 41.4% power reduction in a 1024-core chiplet

- IoT: Up to 65% power reduction near-threshold, high environmental temperature

- High energy efficiency of 118 GFLOPS/W @ 7.2 GFLOPS