Smallest RISC-V Device for Next-Generation Edge Computing

Seiji Munetoh¹, Chitra K Subramanian², Arun Paidimarri², Yasuteru Kohda¹
IBM Research – Tokyo¹ & T.J. Watson Research Center²
• A simple microprocessor core uses 100K-1M transistors, and can fit in an area as small as 100X100 um^2 using advanced technology nodes.
• Running at 10 MHz, such a microprocessor will consume 1-10 mW (and much less, if it runs slower).
• The creation of **compute elements that are ultra-compact and low cost** will enable a dramatic expansion of applications in areas from security to IoT to health care and beyond.
Our 1\textsuperscript{st} target application – Authentication

- Hash based authentication
  - HMAC-SHA256 and variants

- Host/device communication – Optical
  - With micro-LED and micro-PV/PD cells
  - Protocol, UART, HDLC frame, custom payloads

- Bootloader ROM (synthesized, embedded in the proc. chip)
  - Basic device authentication
  - Upload new application to the device

- Storage Memory (as external chip)
  - Use SRAM to emulate NV memory chip
Our 1st gen. processor and 2.5D integrated device

ASIC: 300um x 250um, GF14LPP
SoC: Based on PULPino (RV32IMC)
Memory: 2KB data SRAM
+ Authentication engine
+ Analog custom circuits (LDO, Clock...)

Si interposer < 1mm², 20μm bump
+ Processor
+ Memory (32KB SRAM)
+ Optical I/O: PD, MicroLED
+ Power: PV cells (1V,3V)
Original PULPino Architecture (32KB SRAM x2)

- 32KB Inst SRAM
- 32KB Data SRAM
- Boot ROM (xKB)
- SPI-Flash (xKB)
Architecture evaluation using COTS FPGA boards

- Reduce Memory (SRAM/Flash) size
- Reduce # of I/O pins
- Confirm Performance
- Develop BootROM w/ Uploader
- Emulate & Test ASIC design
  - Custom analog circuits
    (LDO, Clock OSC, PD, LED)
Reduce Memory footprint:
Original PULPino: 32KB(I)+32KB(D)+Ext Flash
Reduced memory: 2KB(D) + 32KB Ext SRAM

- PULPino constraint: SRAM size > App size
- Remove Inst. SRAM from processor die, and use external memory to store and exec the code (XIP)
  - SRAM area (in 14nm):
    \[ 60 \text{Kmm}^2/64\text{KB} \Rightarrow 4\text{Kmm}^2/2\text{KB}, \quad (1/14) \]
- Support XIP
- Expand bus widths from 4 to 8,16,32

Evaluate the performance of 4,8,16,32 bit data bus widths between Proc. and ext. SRAM
Reduced memory footprint: Application Execution Performance Trade Off

- Clock/Instruction
  - 1 (original), 76 (4), 40 (8), 24 (16), 16 (32)

- Application execution time from ext. SRAM
  - 8 ~ 36 times slow

- Choose the 8-bit bus configuration,
  - IO footprint fits within SRAM memory chip size
  - Adequate for most applications

But, a hash calculation by SW (XIP on ext. SRAM) is too slow since it requires greater computational power
With or without of Authentication Engine HW

- **Authentication engine**
  - SHA256, HMAC, Benes-network

- **SHA256 performance**
  - Original: 7760 clk/blk
  - 8-bit bus: 242966 clk/blk
  - 8-bit bus + HW assist: 12474 clk/blk
  - Original + HW assist: 350 clk/blk

- **Area (FPGA)**
  - LUT 18K-> 21K, + 33%
  - FF 15K->17K, + 21%

**Hash Algorithm performance is adequately restored with addition of the authentication engine HW. And the BootROM can support Hash calc. w/ Auth. engine**
Our Modified Processor Architecture
(2KB SRAM + external 32KB MPI-SRAM)

- 2KB Data SRAM
- Boot ROM (xKB)
- 32KB MPI-SRAM (XIP)
- Analog
ASIC Implementation: Processor, MPI-SRAM, Debug Chip

- Global foundries 14LPP
# Testing the Debug Chip

<table>
<thead>
<tr>
<th>Singulation</th>
<th>Packaging</th>
<th>Testing the operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dicing</td>
<td>QFP64, wire bonding</td>
<td>On testbed PCB</td>
</tr>
</tbody>
</table>

**Boot – OK**  
**Application upload & run - OK**
## Testing the 2.5D integrated Device

<table>
<thead>
<tr>
<th>Singulation</th>
<th>Packaging</th>
<th>Testing the operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etching</td>
<td>On Si interposer</td>
<td>On probe station</td>
</tr>
</tbody>
</table>

![Images showing different stages of the process]

**Boot – OK**
# Summary of processor specs

<table>
<thead>
<tr>
<th></th>
<th>PULPino</th>
<th>Our processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASIC node</td>
<td>65nm</td>
<td>14nm</td>
</tr>
<tr>
<td>ASIC size</td>
<td>1mm²</td>
<td>0.076mm²</td>
</tr>
<tr>
<td>Memory</td>
<td>I-SRAM 32KB</td>
<td>I-SRAM N/A</td>
</tr>
<tr>
<td></td>
<td>D-SRAM 32KB</td>
<td>D-SRAM 2KB</td>
</tr>
<tr>
<td>Ext. Memory</td>
<td>128MB Flash</td>
<td>MPI-SRAM 32KB</td>
</tr>
<tr>
<td>Clock</td>
<td>XXMhz</td>
<td>1-100MHz</td>
</tr>
<tr>
<td>I/O</td>
<td>UART, I2C.SPI</td>
<td>UART</td>
</tr>
<tr>
<td>Debug</td>
<td>SPI slave</td>
<td>(SPI slave)</td>
</tr>
<tr>
<td>Analog</td>
<td>-</td>
<td>LDO, Clock/Reset, LED driver, PD input</td>
</tr>
</tbody>
</table>
Conclusions and future plans

- Our 1\textsuperscript{st} generation device is under full evaluation
  - Preliminary tests showed functionality
  - Target application: authentication

- Our 2\textsuperscript{nd} generation device was taped out Feb. 2018
  - New SoC design with I-cache, radio interface, sensors
  - Target applications: Blockchain and IoT application

- Our 3\textsuperscript{rd} generation device is under consideration

- It's all RISC-V
Thank you