Ascenium: A Continuously Reconfigurable Architecture

Robert Mykland
Founder/CTO
robert@ascenium.com
August, 2005
Ascenium: A Continuously Reconfigurable Processor

- Continuously reconfigurable approach provides:
  - The computational efficiency of direct logic implementations in ASICs
  - All the flexibility of microprocessors (pure software)
- Unique architecture enables the array to be effectively targeted by an ANSI Standard C compiler
  - CPU logic is continuously redefined to match the compiler’s needs for each instruction (≈20 cycles)
- Massive parallelism even in programs with little or no inherent instruction-level parallelism
  - Dozens of lines of sequential C code are executed every instruction
Ascenium Block Diagram

- No instruction set
- RLU is combinatorial logic that settles
- Memory is accessed in statically scheduled block operations
- Two instruction streams
  1) Memory
  2) Configuration
- Whole loops are often executed as one RLU instruction

---

Reconfigurable Logic Unit (RLU)

Memory

Ascenium
Ascenium Basic Operation

Conventional

- ALU
- Register File
- Memory

Clock = 12

Ascenium

- Reconfigurable Logic Unit (RLU)
- Memory

Clock = 165

RISC Instructions

Clock = 24
A128X16 Rough Layout

- 130 nm process
- 128 16-bit logic elements
- 64 MACs/clock cycle
- 200 MHz DDR II = 3.2 GB/s external memory bandwidth
- 256KB onchip memory
- 19.2 GB/s peak onchip memory bandwidth
- 500 mW @ 200 MHz
- 52 mm²
Compiler Block Diagram

- GCC Based Compiler (open source)
- Library (open source)
- Bytecode Linker (open source)
- Proprietary Ascenium Code Generator (develop)
- Source code (C, C++, Java)
- Bytecode files
Code Generator Diagram

LLVM PARSER → GENERIC INT. REP. (GIR) → GLOBAL THREADING → GIR OPTS.

ASCENIUM INT. REP. (AIR) → FUNC. ASSIGNMENT & SCORING → ARRAY INST. PACKING & COMP. → DATA DIRECTIVES

FETCH ORDER ASSY. → AIR OPTS. → OUTPUT

● indicates where optimizations occur
Generic Intermediate Representation (GIR)
FFT Inner Loop C Code

i1 = i0 + n2;
i2 = i1 + n2;
i3 = i2 + n2;
r1 = x[2 * i0] + x[2 * i2];
r2 = x[2 * i0] - x[2 * i2];
t = x[2 * i1] + x[2 * i3];
x[2 * i0] = r1 + t;
r1 = r1 - t;
s1 = x[2 * i0 + 1] + x[2 * i2 + 1];
s2 = x[2 * i0 + 1] - x[2 * i2 + 1];
t = x[2 * i1 + 1] + x[2 * i3 + 1];
x[2 * i0 + 1] = s1 + t;
s1 = s1 - t;
x[2 * i2] = (r1 * co2 + s1 * si2) >>15;
x[2 * i2 + 1] = (s1 * co2-r1*si2)>>15;
t = x[2 * i1 + 1] - x[2 * i3 + 1];
r1 = r2 + t;
r2 = r2 - t;
t = x[2 * i1] - x[2 * i3];
s1 = s2 - t;
s2 = s2 + t;
x[2 * i1] = (r1 * co1 + s1 * si1)>>15;
x[2 * i1 + 1] = (s1 * co1-r1 * si1)>>15;
x[2 * i3] = (r2 * co3 + s2 * si3)>>15;
x[2 * i3 + 1] = (s2 * co3-r2 * si3)>>15;
FFT Loop Array Instruction

<table>
<thead>
<tr>
<th>MM</th>
<th>A</th>
<th>D</th>
<th>C</th>
<th>F</th>
<th>I</th>
<th>J</th>
<th>GG</th>
<th>HH</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>K</td>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>F</td>
<td>F</td>
<td>J</td>
<td>J</td>
<td>P</td>
<td>Q</td>
<td>T</td>
<td>U</td>
</tr>
<tr>
<td>Z</td>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>S</td>
<td>V</td>
<td>W</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>AA</td>
<td>AA</td>
<td>BB</td>
<td>BB</td>
<td>CC</td>
<td>DD</td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table represents a grid with specific instructions or values, but the exact meaning is not clear from the image alone.
FFT Loop Data Directives

1. Fetch the array instruction and data directives
2. Load the array instruction
3. Read \( x[t] \) in four 64-bit wide banks 32 times
4. Write \( x[t] \) in one 256-bit wide bank 32 times, changing banks every 8 operations
5. Repeat steps 3 and 4 three more times
6. At the same time as 3 – 5, read in \( w \) coefficients
7. Pass 1: \( w \) every clock; 2: \( w \) every 4 clocks, then every 16, then every 64
8. Stop
## DSP Benchmark Results
(Simulated, Hand-Coded)

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>A128X16 core, 130nm, 250MHz, 250mW</th>
<th>TMS320C64x core, 130nm, 300MHz*, 250mW</th>
</tr>
</thead>
<tbody>
<tr>
<td>256-Point Radix-4 FFT</td>
<td>512 ns</td>
<td>8,920 ns</td>
</tr>
<tr>
<td>Real Block FIR Filter</td>
<td>512 ns</td>
<td>6,827 ns</td>
</tr>
<tr>
<td>Complex Block FIR Filter</td>
<td>512 ns</td>
<td>6,827 ns</td>
</tr>
<tr>
<td>DCT x 24</td>
<td>512 ns</td>
<td>6,080 ns</td>
</tr>
<tr>
<td>IIR Filter NR = 2,048</td>
<td>292 ns</td>
<td>6,827 ns</td>
</tr>
</tbody>
</table>

* The 130nm C64x core will operate up to 720 MHz, but burns more power at higher frequencies.
Ascenium Status

- Patents filed
- Chip architecture complete, polishing ongoing
- Prototype compiler complete & demo-able
- Development plan in place
- Raising seed round
Summary

• Great performance and performance/mW:
  >10x programmable DSPs
  >4x performance/mW of FPGAs

• Great time-to-market:
  -- Normal GPP software development cycle

• Great ease-of-use:
  -- Programmers writing standard, non-optimized code in C/C++ get all the performance