FPGA, a history of interconnect

Ivo Bolsens

CTO, Xilinx
The Architectural Shakeout

Mass extinctions in the mid 1990s
Xilinx: 8100, 6200, 4700, Prizm, …
Plessey, Toshiba, Motorola, IBM, …

We were hit by fast-moving CMOS process technology, particularly multiple metal layers.
Trends

![Graph showing trends in LUTs and Wire](image)
Interconnect and ease of use

Longer wire reach gave dramatic improvement for ease of design.

Lower is better.
State-of-the-Art 65nm FPGA

- High-Performance 6-LUT Fabric
- 36Kbit Dual-Port Block RAM / FIFO with ECC
- SelectIO with ChipSync + XCITE DCI
- 550 MHz Clock Management Tile DCM + PLL
- More Configuration Options
- 25x18 Multiplier DSP Slice with Integrated ALU
Predictable Interconnect

Symmetric routing pattern reaches more CLBs with fewer hops

Dramatically increases design performance
Layers of Interconnect

12 layers on chip, 10 layers in package, 10 layers in PCB.

82% of noise is determined by the pin-out and found in package balls and PCB vias.
Sparse Chevron Pin-out Pattern

- Every SelectIO adjacent to return path
- Achieves near-ideal return current loop

IO pins in a regular array of return path pins
Off chip interconnect

- Growing gap between number of logic gates and I/O
- Technology scaling favors logic density

15x drop in I/O-to-logic ratio by 2020

Source: ITRS
I/O to Logic Ratio

Comparison of number of I/O per 1000 logic cell in the largest FPGA in each family

~60x decrease in I/O-logic ratio
Microprocessor’s chip area is dominated by cache memory to overcome the lack of I/O bandwidth
The Move to Serial Connectivity

Arrays of serial Gb transceivers

Source: Intel/Xilinx
Die-Stacking Landscape

(Connection Density, Number of Device Layers)

$>10^6/cm^2$
3-4 device layers

$10^2-10^3/cm^2$
4+ device layers

$10^4-10^6/cm^2$
4+ device layers

1st Si: bottom supporting wafer
2nd Si: thinned to 5.5um
3rd Si: thinned to 5.5um
SiO$_2$

Chip-Stacking
(wire-bonding)

Chip-Stacking
(Through Silicon Vias)

Monolithic
3-D integration

Photo: Amkor

Photo: Tezzaron
Imagine…

Specialized Layers of
DSP fabric,
Memory fabric,
FPGA fabric, …

Optimized for Technology
130nm,
90nm,
32nm, …

With 3D Interconnect

And at its heart…

An FPGA SoC with:
• Embedded Processing
• Embedded DSP
• High-speed Serial Connectivity
• Reprogrammable FPGA Logic Fabric as the Base

Benefits :
• Single package of heterogeneous die
• Multiple configurations of standard products
• Lower cost
• Lower power
• Ultimate customization
• Ultimate flexibility
Wafer Bonding

1) Prep Wafers; Dice Daughter

- Daughter Wafer
- Mother Wafer
- Daughter IC

2) Attach

- Daughter IC
- Mother Wafer

3) Fuse

- Daughter IC
- Mother Wafer

4) Wafers from Foundry

- Daughter Wafer
- Mother Wafer

1) 4 to 15 µm
2) 7 to 50 µm

Source: Cubic Wafer
Inter-Die Connection Performance

- 100X less dynamic power than conventional single-ended I/O
- Link performance ~ 1 Gigabit/sec
2-D FPGA Fabric

Routing Switch

Look Up Table (LUT).

A programmable logic block
3-D FPGA Fabric

- Shorter wire-length and delay
- Higher logic density

Performance Improvement

- Integration of RAM with FPGA by high bandwidth die-stacking
- 2 Terabit/sec/sec bandwidth between FPGA and RAM

<table>
<thead>
<tr>
<th>High Performance Applications</th>
<th>Projected Performance Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Matrix/ Vector Multiply</td>
<td>4X-8X over 2.4GHz Pentium</td>
</tr>
<tr>
<td>Traffic Simulation</td>
<td>170X over 2.2GHz Opteron</td>
</tr>
<tr>
<td>Radiative Heat Transfer</td>
<td>20X over 1.7GHz Pentium</td>
</tr>
<tr>
<td>Molecular Dynamics</td>
<td>~20X over 3 GHz Processor</td>
</tr>
</tbody>
</table>

(Courtesy of Maya Gokhale, LANL)
The FPGA System

- MGTs
- Logic Emulation
- DSP
- Memory
- PowerPC
- I/Os
- Communication Port
- Custom Logic
- DSP Accelerator
- Internal Memory
- μP
- External Memory Port
MicroBlaze’s Flexible Acceleration Interface

- **FSL = Fast Simplex Link**
  - Eliminates bus signaling overhead
    - No address decode
    - No arbitration
    - No acknowledge cycles
- Simple instruction programming
- Flexible number master and slave FSL ports
  - Configurable depth FIFO in FSL
  - Input and output FSL port width is configurable as 8, 16, or 32 bits.
- Dedicated MicroBlaze instruction
  - Get fromInputFSL M, toReg N
  - Put toOutputFSL M, fromReg N
  - Blocking and non-blocking support
Application-Specific Hardware Acceleration

- When the processor core begins to reach software task capacity, then Fabric Acceleration to the rescue
  - Use Fast Simplex Link (FSL) to interface to customer-defined accelerators
  - Enables dramatic improvements in performance
FPGA/processor

- CoreConnect Architecture
  - Processor Local Bus (PLB)
    - Ideal for burst transfers
      - Memory, High Speed Peripherals, Cache Interface
    - 32-bit address, 64-bit data
    - 2.1 GB/s Max BW @ 133 MHz
  - On-Chip Peripheral Bus (OPB)
    - Low speed peripherals
    - 32-bit address, 32-bit data
  - Device Control Register Bus (DCR)
    - For peripheral setup and control

- On Chip Memory Interface (OCM)
  - 4 Processor Cycle Latency
  - Lowest Latency and good data rate
  - Not a bus interface
Accelerate Performance Beyond the Core

- Extends PPC 405 Instruction Set
  - Floating point support
  - User Defined Instructions
- Offloads CPU intensive operations
  - Matrix calculations
    - Video processing
  - Floating point mathematics
    - 3D data processing
- Direct interface to HW accelerators
  - High Bandwidth
  - Low Latency
- Reduce number of bus cycles by factor of 10X
- Increase performance by over 20X
Comparison with Traditional Bus-based APU

**APU**

- **Processor Block**
- **Soft Aux. Processor**

- **Write Instruction and operands**
  - 1 APU cycle

- **Execution**
  - \( N_{EX} \) APU cycle

- **Read Result and Status**
  - 1 APU cycle + 1 CPU cycle

**PLB**

- **Processor Block**
- **Soft Aux. Processor**

- **Write Operand1**
  - 5 PLB cycles + 2 CPU cycles

- **Write Operand2 and Instruction**
  - 5 PLB cycles + 2 CPU cycles

- **Execution**
  - \( N_{EX} \) PLB cycle

- **Read Status**
  - 6 PLB cycles + 3 CPU cycles

- **Read Result**
  - 6 PLB cycles + 3 CPU cycles
**Processor Performance and Fabric Acceleration**

- **2002**: Virtex-II Pro
  - PowerPC 405 at 450 MHz
  - APU

- **2004**: Virtex-4
  - PowerPC
  - APU

- **2006**: Virtex-5
  - PowerPC
  - APU

- **2008**: Next generation PowerPC

- **2010**: "Traditional" Frequency Scaling

---

**Fabric Acceleration**
- PowerPC – APU
- MicroBlaze - FSLs
PowerPC Architecture

Full System Customization & High Performance
Scalability

Scalable in performance, and re-usable solution

Processing rate

FPGA
Flexible system on chip used for tailored system architectures

Fixed processor
Processor / memory bottlenecks worsen

2005
Next fixed architecture

2007
Next fixed architecture

2009

No re-use, architecture dependent
You must customize your program to make use of the busses of the Cell processor.

The FPGA connectivity can be customized to fit your program.
Flexibility

Woodcrest: 24 GFlops @3Ghz
Virtex-5: 60 GFlops (70 GFlops-SP) @350Mhz

Woodcrest = 80 Watt
Virtex-5 = 10 Watt
Internal Memory Bandwidth
Conclusions

• It is all about connectivity
• Important aspects
  – Off-chip interconnect goes serial
    • Latency, power
  – On-chip interconnect has to be
    • Scalable
    • Hierarchical
    • Flexible
    • Easy to use
  – System interconnect heterogeneous and
tailored to application