ESE680-002 (ESE534): Computer Organization

Day 10: February 12, 2007
Empirical Comparisons

Last Time
• Instruction Space Modeling

Today
• Empirical Data
  – Custom
    • Gate Array
    • Std. Cell
    • Full
  – FPGAs
  – Processors
  – Tasks

Empirical Comparisons

Empirical
• Ground modeling in some concretes
• Start sorting out
  – custom vs. configurable
  – spatial configurable vs. temporal

Full Custom
• Get to define all layers
• Use any geometry you like
• Only rules are process design rules
• ESE570
Standard Cell Area

- \texttt{inv}
- \texttt{nand3}
- \texttt{AOI4}
- \texttt{nor3}
- \texttt{inv}

All cells uniform height

Width of channel determined by routing

Cell area

Identify the full custom and standard cell regions on 386DX die

http://microscope.fsu.edu/chipshots/intel/386dxlarge.html

MPGA

- Metal Programmable Gate Array
  - Resurrected as “Structured ASICs”
  - …and already dead?...
    - Structured ASICs: what happened at LSI Logic?
      - EDN, Oct. 2006
    - Still headed to $1B business?
      - http://www.elecdesign.com/Articles/Index.cfm?AD=1&AD=1&ArticleID=14442
  - Gates pre-placed (poly, diffusion)
  - Only get to define metal connections
    - Cheap – only have to pay for metal mask(s)

MPGA/SA vs. Custom?

- AMI CICC’83
  - MPG 1.0
  - Std-Cell 0.7
  - Custom 0.5
- AMI CICC’04
  - Custom 0.6 (DSP)
  - Custom 0.8 (DPath)
- Toshiba DSP
  - Custom 0.3
- Mosaid RAM
  - Custom 0.2
- GE CICC’86
  - MPG 1.0
  - Std-Cell 0.4–0.7
  - FF/counter 0.7
  - FullAdder 0.4
  - RAM 0.2

Metal Programmable Gate Arrays

- Modern -- “Sea of Gates”
- Yield 35–70%
- Maybe $5/\text{gate}$?
  - (Quite a bit of variance)
FPGA Table

<table>
<thead>
<tr>
<th>Year</th>
<th>Design</th>
<th>Organization</th>
<th>Max</th>
<th>( \lambda )</th>
<th>( \lambda^2 ) area</th>
<th>cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>Xilinx 2K</td>
<td>CLB (4-LUT)</td>
<td>100</td>
<td>1.9</td>
<td>800K</td>
<td>20 ns</td>
</tr>
<tr>
<td>1988</td>
<td>Xilinx 3K</td>
<td>CLB (2-4-LUT)</td>
<td>320</td>
<td>0.6</td>
<td>1.2M</td>
<td>13 ns</td>
</tr>
<tr>
<td>1992</td>
<td>Xilinx 4K</td>
<td>CLB (2-4-LUT)</td>
<td>1024</td>
<td>0.8</td>
<td>1.25M</td>
<td>7 ns</td>
</tr>
<tr>
<td>1995</td>
<td>Xilinx 5K</td>
<td>CLB (4-LUTS)</td>
<td>484</td>
<td>0.3</td>
<td>2.26M</td>
<td>6 ns</td>
</tr>
<tr>
<td>1995</td>
<td>Altera 5K</td>
<td>LE (4-LUT)</td>
<td>1256</td>
<td>0.3</td>
<td>2.25K</td>
<td>7.5 ns</td>
</tr>
<tr>
<td>1999</td>
<td>Orca 2K</td>
<td>FLC (4-LUT)</td>
<td>600</td>
<td>0.3</td>
<td>4.3M</td>
<td>7 ns</td>
</tr>
<tr>
<td>1998</td>
<td>HSRA</td>
<td>BLB (5-LUT/2 = 4-LUT?)</td>
<td>0.2</td>
<td>2M</td>
<td>4 ns</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>4-LUT</td>
<td>2K</td>
<td>-</td>
<td>800K</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>4-LUT</td>
<td>16K</td>
<td>-</td>
<td>1M</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

(semi) Modern FPGAs

- APEX 20K1500E
  - 52K LEs
  - 0.18\( \mu \)m
  - 24mm \times 22mm
  - 1.25M\( \lambda^2 \)/LE
- XC2V1000
  - 10.44mm \times 9.90mm
  - 0.15\( \mu \)m
  - 11,520 4-LUTs
  - 1.5M\( \lambda^2 \)/4-LUT

[Both also have RAM in cited area]

Conventional FPGA Tile

K-LUT (typical \( k=4 \))

w/ optional output Flip-Flop

Interconnect

How many gates?

```
00 x_0
01 x_1
10 x_2
11 x_3
```

“gates” in 2-LUT

```
\[ 0 \rightarrow 0.5 \rightarrow 1 \rightarrow 1.5 \]
```

```
\[ \text{LUT} \]
```

Toronto FPGA Model
Now how many?

![Diagram of XOR gate]

Which gives:

- More usable gates?
- More gates/unit area?

Gates Required?

![Diagram of FF-FF with input signal]

Depth=3, Depth=2048?

Gate metric for FPGAs?

- Day8: several components for computations
  - compute element
  - interconnect:
    - space
    - time
  - instructions
- Not all applications need in same balance
- Assigning a single "capacity" number to device is an oversimplification

MPGA vs. FPGA

- Xilinx XC4K
- 5K/2/gate
- 35-70% usable (50%)
- 7-17K/2/gate net
- Ratio: 2–10 (5)

Adding ~2x Custom/MPGA,
Custom/FPGA ~10x

![Table of FPGA Comparisons]

- 90nm
- FPGA: Stratix II
- STMicro CMOS090
  - Standard Cell
    - Full custom layout
    - …but by tool

Adding ~2x Custom/MPGA,
Custom/FPGA ~10x

[Kuon/Rose TRCADv26n2p203–215 2007]
MPGA vs. FPGA

- MPGA (SOG GA)
  \[ \lambda = 0.6 \mu \]
  \[ \tau_{gd} \approx 1 \text{ ns} \]
- Xilinx XC4K
  \[ \lambda = 0.6 \mu \]
  1-7 gates in 7ns
  2-3 gates typical
- Ratio: 1-7 (2.5)
  - Altera claiming 2x
  - LSI claiming 3x
  - For their SA [2007]

Processors vs. FPGAs

- 90nm
- FPGA: Stratix II
- STMicro CMOS090

Component Example

- Single die in 0.35μm
  XC4085XL-09 3,136 CLBs 4.6ns
  682 Bit Ops/ns
  Alpha 1996 2×64b ALUs 2.3ns
  55.7 Bit Ops/ns

[1 “bit op” = 2 gate evaluations]
Raw Density Summary

- **Area**
  - MPGA 2-3x Custom
  - FPGA 5x MPGA
    - FPGA:std-cell custom ~ 15-30x
- **Area-Time**
  - Gate Array 6-10x Custom
  - FPGA 15-20x Gate Array
    - FPGA:std-cell custom ~ 100x
  - Processor 10x FPGA

Raw Density Caveats

- Processor/FPGA may solve more specialized problem
- Problems have different resource balance requirements
- …can lead to low yield of raw density

Task Comparisons

Broadening Picture

- Compare larger computations
- For comparison
  - throughput density metric: results/area-time
  - normalize out area-time point selection
  - high throughput density
    - most in fixed area
    - least area to satisfy fixed throughput target

Multiply

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Feature Size (u)</th>
<th>Area and Time</th>
<th>16 x 16</th>
<th>8 x 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom 16 x 16 Custom 6 x 6</td>
<td>0.50um, 0.40um</td>
<td>2.6M^2, 40ns</td>
<td>9.6</td>
<td>9.6</td>
</tr>
<tr>
<td>Gate-Array 16 x 16</td>
<td>0.50um</td>
<td>3.3M^2, 4.3ns</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>FPGA (Xilinx)</td>
<td>0.50um</td>
<td>1.2M^2/2.3ns</td>
<td>0.09</td>
<td>0.30</td>
</tr>
<tr>
<td>1st DSP</td>
<td>0.50um</td>
<td>3.5M^2, 20ns</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>RISC (16-bit multiply)</td>
<td>0.50um</td>
<td>12M^2/60 ns/cycle</td>
<td>0.002</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Example: FIR Filtering

Application metric:
TAPs = filter taps multiply accumulate

\[ Y_i = w_1 x_i + w_2 x_{i+1} + ... \]

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Feature Size (u)</th>
<th>( f / 1^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>32b RISC</td>
<td>0.75um</td>
<td>0.020</td>
</tr>
<tr>
<td>16b DSP</td>
<td>0.625um</td>
<td>0.07</td>
</tr>
<tr>
<td>32b RISC/USM</td>
<td>0.25um</td>
<td>0.021</td>
</tr>
<tr>
<td>64b RISC</td>
<td>0.125um</td>
<td>0.064</td>
</tr>
<tr>
<td>FPGA (Xilinx)</td>
<td>0.60um</td>
<td>1.9</td>
</tr>
<tr>
<td>Altera 8K</td>
<td>0.30um</td>
<td>3.6</td>
</tr>
<tr>
<td>Full Custom</td>
<td>0.75um</td>
<td>3.6</td>
</tr>
<tr>
<td>(fixed coefficient)</td>
<td>0.60um</td>
<td>4.8</td>
</tr>
<tr>
<td>(n.b. 16b samples)</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
IIR/Biquad

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Feature Size (µm)</th>
<th>Area and Time</th>
<th>16b TAPs</th>
<th>10b TAPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>16b DSP</td>
<td>0.60x0.60</td>
<td>200MHz, 500ns/biquad</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>FPGA</td>
<td>0.60x0.60</td>
<td>60 CLBs, 320ns/biquad</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>(XC4K)</td>
<td>0.90x0.90</td>
<td>43 CLBs, 200ns/biquad</td>
<td>0.093</td>
<td>5.0</td>
</tr>
<tr>
<td>Full Custom</td>
<td>0.90x0.90</td>
<td>68MHz, 11.8ns/4 biquads</td>
<td>0.044</td>
<td></td>
</tr>
</tbody>
</table>

Simplest IIR: \( Y_i = A \times X_i + B \times Y_{i-1} \)

DES Keysearch

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Feature Size (µm)</th>
<th>Area</th>
<th>Keys/Second</th>
<th>Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>DES IC</td>
<td>1.5x1.5</td>
<td>11.1MHz</td>
<td>310K</td>
<td>0.028</td>
</tr>
<tr>
<td>FPGA (Altera 8K)</td>
<td>0.30x0.30</td>
<td>811.8MHz (93OMHz)</td>
<td>800K</td>
<td>0.00006</td>
</tr>
<tr>
<td>RISC</td>
<td>0.30x0.30</td>
<td>1.8GHz</td>
<td>1.1K</td>
<td>0.000023</td>
</tr>
</tbody>
</table>

<http://www.cs.berkeley.edu/~iang/isaac/hardware/>

DNA Sequence Match

- **Problem**: "cost" of transform \( S_1 \rightarrow S_2 \)
- **Given**: cost of insertion, deletion, substitution
- **Relevance**: similarity of DNA sequences
  - evolutionary similarity
  - structure predict function
- **Typically**: new sequence compared to large database

DNA Sequence Match

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Feature Size (µm)</th>
<th>Area</th>
<th>Cell Updates per Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom FPGA</td>
<td>2.0x2.0</td>
<td>270MHz</td>
<td>500M       1.9</td>
</tr>
<tr>
<td>(SPLASH 2)</td>
<td>0.60x0.60</td>
<td>43G</td>
<td>3,000M 0.070</td>
</tr>
<tr>
<td>(SPLASH)</td>
<td>0.60x0.60</td>
<td>33G</td>
<td>370M 0.012</td>
</tr>
<tr>
<td>RISC</td>
<td>0.60x0.60</td>
<td>273M</td>
<td>873M 0.0032</td>
</tr>
<tr>
<td>(SparcStation 10)</td>
<td>0.40x0.40</td>
<td>1.6G</td>
<td>1.2M 0.00075</td>
</tr>
</tbody>
</table>

N.B. includes memory area for SPLASH

Degrade from Peak

- Long path length \( \rightarrow \) not run at cycle
- Limited throughput requirement
  - bottlenecks elsewhere limit throughput req.
- Insufficient interconnect
- Insufficient retiming resources (bandwidth)
Degrade from Peak: Processors

- Ops w/ no gate evaluations (interconnect)
- Ops use limited word width
- Stalls waiting for retimed data

\[ E(\text{Functional Density}) = \frac{\text{Gate Evaluations} \times \text{Datapath Bits} \times \text{pins} \times \text{Issue Slots} \times \text{Clock Cycle} \times \text{area}}{\text{clock}} \]

Degrade from Peak: Custom/MPGA

- Solve more general problem than required
  - (more gates than really need)
- Long path length
- Limited throughput requirement
- Not needed or applicable to a problem

Degrade Notes

- We’ll cover these issues in more detail as we get into them later in the course

Big Ideas
[MSB Ideas]

- Raw densities:
  - custom:ga:fpga:processor
  - 1:5:100:1000
  - close gap with specialization

Admin

- No class next Monday (2/19)
- No office hours Tuesday 2/20