Specialization for Data Analytics

Martha Kim
Columbia University
Why Specialize?

- General-Purpose Processor
  - 40-500x
- Application-Specific Processor (ASIP)
  - 10-350x
- Field-Programmable Gate Array (FPGA)
  - 10-40x
- Standard Cell ASIC
  - 3-10x
- Full Custom ASIC
  - 6-8x
Why **Not** Specialize?

<table>
<thead>
<tr>
<th>Homogeneous (design, maintenance)</th>
<th>Heterogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform software interface</td>
<td>Complex, arcane</td>
</tr>
<tr>
<td>Arbitrary software</td>
<td>Fixed function(s)</td>
</tr>
</tbody>
</table>
In-connection

Memory

Q100
Data Processing Unit

Data Processing Unit 2.0

System Integration Challenges

CPU
Accelerator
Data-Centric Specialization

8KB data cache read: 131pJ

8KB data cache read: 131pJ

32 bit addition: 10pJ
8KB data cache write: 131pJ

40-word regfile: 17pJ

Data supply accounts for 83%-97% of total energy!
Q100 Database Processing Unit

- Targets analytical queries (not updates or txns)
- Hardware implementations of relational operators
- Processes streams representing columns or tables
- Spatial and temporal ISA
SELECT s_season, SUM(s_qty) as sum_qty
FROM sales
WHERE s_shipdate >= '2013-01-01'
GROUP BY s_season
ORDER BY s_season
SELECT s_season, SUM(s_qty) as sum_qty
FROM sales
WHERE s_shipdate >= '2013-01-01'
GROUP BY s_season
ORDER BY s_season
Key Sources of Efficiency

- Specialized interface to memory (think stream via DMA)
- Division of concerns: compute tiles never see an address
- Read datum once, multiple operations
- Pipeline records, parallel streams
Example Tile: BoolGen

BOOLGEN

IN 0

IN 1

CMP

EN

PRED

COLUMN FILTER
Range Partitioning Operation

Keys:
- 36
- 4
- 21
- 16
- 30
- 10
- 34
- 29
- 30
- 9
- 33
- 15
- 10
- 39
- 22
- 30
- 8
- 38

Splitters:
- 4
- 9
- 8
- 10
- 16
- 15
- 21
- 29
- 22
- 30
- 30
- 30
- 36
- 34
- 33
- 39
- 38
Hardware Accelerated Range Partitioner (HARP)
HARP Tile Internals
Tile Characterization

- AGG
- ALU
- BOOLGEN
- COLFILTER
- JOIN
- PART
- SORT
- APPEND
- COLSELECT
- CONCAT
- STITCH

Critical Path (ns)

Area (mm²)

Power (mW)

Area and power dominated by two tiles

Max Frequency 315 MHz
TPC-H Runtime (milliseconds) vs. Power (Watts)

- **Low Power**:
  - 1 ALU
  - 1 Partitioner
  - 1 Sorter
  - Points are scattered towards the lower left corner.

- **Pareto**:
  - 4 ALUs
  - 2 Partitioners
  - 1 Sorter
  - Points are located near the Pareto front.

- **High Perf**:
  - 5 ALUs
  - 3 Partitioners
  - 6 Sorters
  - Points are scattered towards the upper right corner.

Legend:
- Blue arrow: Low Power
- Yellow arrow: Pareto
- Red arrow: High Perf
Inter-connect

Memory

Q100
Data Processing Unit

Data Processing Unit
2.0

System Integration Challenges

CPU
Accelerator
Is a heterogeneous design best?

- Tiles are unitaskers
- Yet, they have a lot in common
- Irregular connectivity, tricky NoC
- Ad-hoc support for non-streaming operators
Recall the Partitioner
Just a Small Step to Merge...
Or to Nested Block Loop Join.
Programmable, Homogeneous DPU

- Homogeneous resources
- So, better usage
- Regular, point-to-point NoC
- Unified buffering and routing
Interconnect

Memory

Data Processing Unit 2.0

Q100 Data Processing Unit

System Integration Challenges
Hardware and Software Integration

It is all one problem!
Exploit established software interfaces

SQL

DBMS

CPU

Accelerator

SQL [query, schema]

Data Independent Translation (Compilation)

Data

MetaLang [soft plan]

Data Dependent Translation (JIT)

Bitfile [partial plan]

Q100

Q100

Q100
Unified Address Space

- Globally visible pointers, so no data marshaling
- No pinning pages for accelerator
- Efficient virtual to physical address translation
Thank You

http://arcade.cs.columbia.edu
Save the planet and return your name badge before you leave (on Tuesday)

Microsoft Privacy Policy statement applies to all information collected. Read at research.microsoft.com