Heterogeneous Shared-Memory Multicore Processors

H. Peter Hofstee
IBM Austin / TU Delft
Transformation Hierarchy

Problem

Algorithm

Program

ISA (Instruction Set Arch)

Microarchitecture

Circuits

Electrons
Playing at the program/ISA boundary ....

E però sappia ciascuno che nulla cosa per legame musaico armonizzata si può de la sua loquela in altra transmutare, senza rompere tutta sua dolcezza e armonia.

Dante (Convivio)

And yet each of them knows that nothing by a harmonized mosaic can be told of his talk in another transmutation without breaking all his sweetness and harmony.

... and Google Translate
Playing at the program/ISA boundary ....

E però sappia ciascuno che nulla cosa per legame musaico armonizzata si può de la sua loquela in altra transmutare, senza rompere tutta sua dolcezza e armonia.

Dante (Convivio)

And so everyone should know that nothing harmonized according to the Muse’s rules can be translated from its native speech into another without breaking all its sweetness and harmony.
Overview

- It is (indeed) all about the memory
- Programming Heterogeneous Multicore
- Apache Arrow and Fletcher
CPU

A(another) story for a 10-year old ...
COHERENT BUS (+RAG)

Memory

LS Alias

LS Alias

Syn. Proc. ISA

MMU/DMA +RMT

Local Store Memory

Syn. Proc. ISA

MMU/DMA +RMT

Local Store Memory

Power ISA +RMT

Power ISA +RMT

MMU/BIU +RMT

MMU/BIU +RMT

IO transl.
Memory Managing Processor vs. Traditional General Purpose Processor

Cell

BE

AMD

IBM

Intel
What do you most associate Cell with?
Cell

PlayStation 3?
Cell

Roadrunner Supercomputer?
Cell

Difficult to program?
Programming Cell

• Cell memory model

• A near miss?

• Tasks

• CellSuperScalar

• OpenCL

• OpenMP
Is shared coherent memory enough? ...
Apache Arrow

**Schema**

```
X {
  A: Float (nullable)
  B: List<Char>
  C: Struct{
    E: Int16
    F: Double
  }
}
```

**Arrow terminology:**

- **Schema:**
  Description of data types in a `RecordBatch`
- **RecordBatch:**
  Tabular structure containing `arrays`
- **Arrays:**
  Combination of `buffers`, can be nested
- **Buffers:**
  Contiguous C-like arrays

**Arrow in-memory example:**

```
<table>
<thead>
<tr>
<th>Index</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.33f</td>
<td>beer</td>
<td>{1, 3.14}</td>
</tr>
<tr>
<td>1</td>
<td>7.01f</td>
<td>is</td>
<td>{5, 1.41}</td>
</tr>
<tr>
<td>2</td>
<td>∅</td>
<td>nice</td>
<td>{3, 1.61}</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Index</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>b</td>
</tr>
<tr>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>2</td>
<td>r</td>
</tr>
<tr>
<td>3</td>
<td>i</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Index</th>
<th>Offset</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>b</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>e</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>r</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>i</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>s</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>i</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>c</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>e</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Index</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Index</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.14</td>
</tr>
<tr>
<td>1</td>
<td>1.41</td>
</tr>
<tr>
<td>2</td>
<td>1.61</td>
</tr>
</tbody>
</table>
```
Old Way

Application
- Apache Spark
  - JVM
  - Python library
  - Native library

Serialize / Deserialize

Network
Disk
Accelerator

Apache Arrow & Fletcher

Application
- Apache Spark
  - JVM
- Python tool
- Native library
- FPGA Accelerator

Apache Arrow libraries

Shared data set / memory in Arrow format

Storage
Network

J. Peltenburg, e.a., TU Delft (OpenPOWER Summit USA 2018)
Fletcher in Action

R=16 different regular expressions per unit

AWS EC2 F1:
• Virtex Ultrascale+
• N=16 regex units
• 256 regexes being matched in parallel

POWER8 CAPI (Supervessel, & soon at Nimbix):
• AlphaData KU3 (Kintex Ultrascale)
• N=8 regex units
• 128 regex being matched in parallel

Johan Peltenburg e.a., TU Delft
Sharing Memory
Dimitris Syrivelis, IBM Research - Ireland

OpenCAPI Extended Main System Memory over OpenCAPI
OpenPOWER Summit Europe
NVIDIA RAPIDS

- A set of libraries that operate on Apache Arrow-based data in GPU memory
Conclusions

- The curve is flattening … we need some (new) ideas
- Preserve more of the programmer’s information
  - Tasks
  - Defined layout for data types (across languages)
- Ideas have made their way into mainstream
  - OpenMP
  - Apache Arrow