The Case for Custom Parallel Memories: an Application-Centric Analysis

Giulio Stramondo, Cătălin Bogdan Ciobanu, Ana Lucia Varbanescu, Cees de Laat

The main idea

For many applications, the performance bottleneck is memory throughput. In this work, we propose to use FPGAs for building parallel custom memory systems to accelerate applications.

Application Access Patterns

Metrics

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Sparse 50%</th>
<th>Sparse 33%</th>
<th>Sparse 66%</th>
<th>Irregular</th>
</tr>
</thead>
</table>

Speedup = \#SequentialAccesses / \#ParallelAccesses

Penalty from using a single memory

Gain from using a parallel memory

Efficiency = ElementsReadByApplication / ElementsReadByParallelMemory

Parallel Reconfigurable Memory

We use the PRF [1] as a 2D reconfigurable scratchpad parallel memory.

Take Home Message

Using parallel memories is not trivial. Our approach offers an integrated semi-automatic approach to customize the memory system for your application, and evaluate its potential performance.

More info?