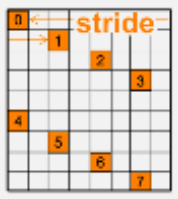
AXI-PACK: Near-Memory Bus Packing for Bandwidth-Efficient Irregular Workloads

Chi Zhang[1], Paul Scheffler[1], Thomas Benz[1], Matteo Perotti[1], Luca Benini[1,2]
[1]Integrated Systems Laboratory, ETH Zurich; [2]DEI, University of Bologna

1 Challenge: irregular memory accesses

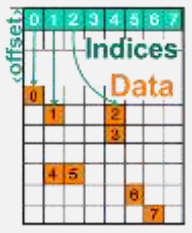
- **Applications**
  - Graph analytics
  - Fluid dynamics
  - Recommender systems

```
For (i=0,i<8,i++)
  var+=Data[ Stride*i ]
```



- **Challenge to processors**
  - Poor bus utilization
  - Cache trashing
  - Long latencies

```
For (i=0,i<8,i++)
  var+=Data[ Index[ i ]]
```



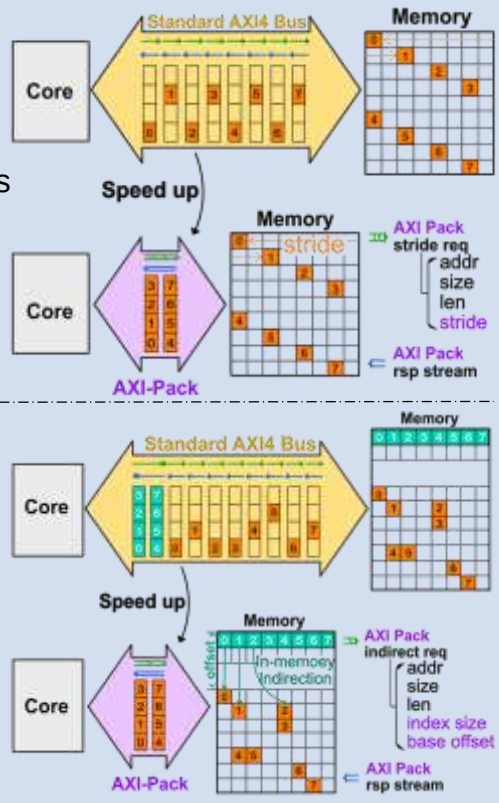
2 State-of-art solutions

- **Core-side stream ISA extensions[1]**
  - \+ Decouple computing and memory access: hide latency
  - \- Inherent inefficiency of narrow bus accesses
  - \- High index-fetching overhead and bus traffic
- **Memory-side extensions[2]**
  - \+ Prefetch and pack irregular elements at memory-side
  - \- Not well co-integrated with Core-side
  - \- Non-standard solutions

gap

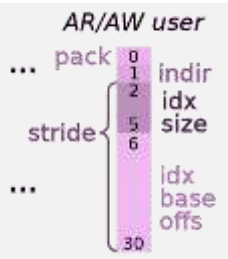
3 Proposal: AXI-Pack on-chip protocol

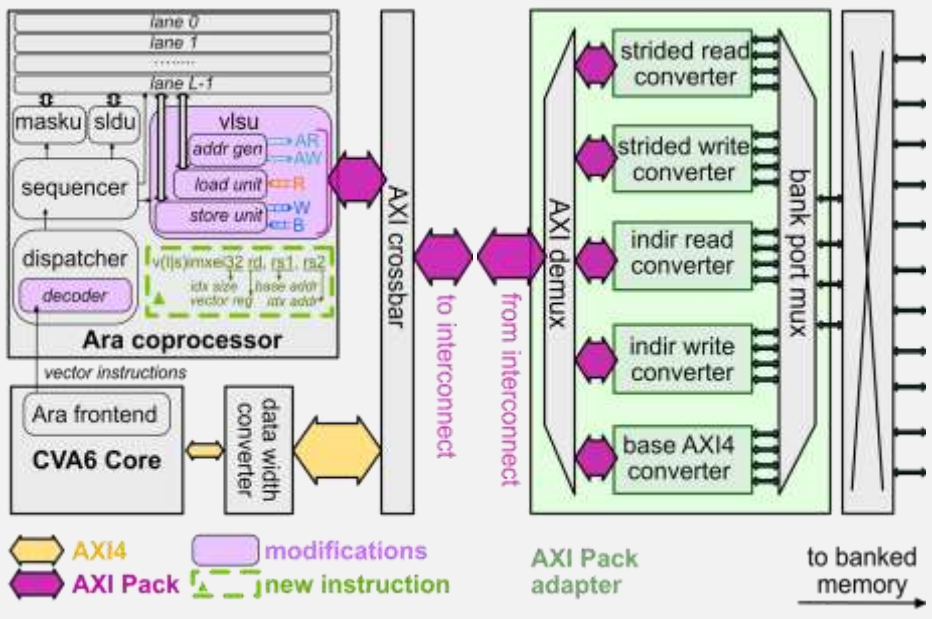
- **Extends Advance eXtensible Interface4 (AXI4)[3] on-chip protocol**
  - Core-side issues
    Pattern-aware requests
  - Memory-side responses
    Densely packed stream
- **Features**
  - End-to-end
    irregular streaming
  - Process-In-Memory
  - Bus-packing
  - **Standard based**
  - **Backward compatible**
  - **Transparent**
  - **Scalable**



4 Design work

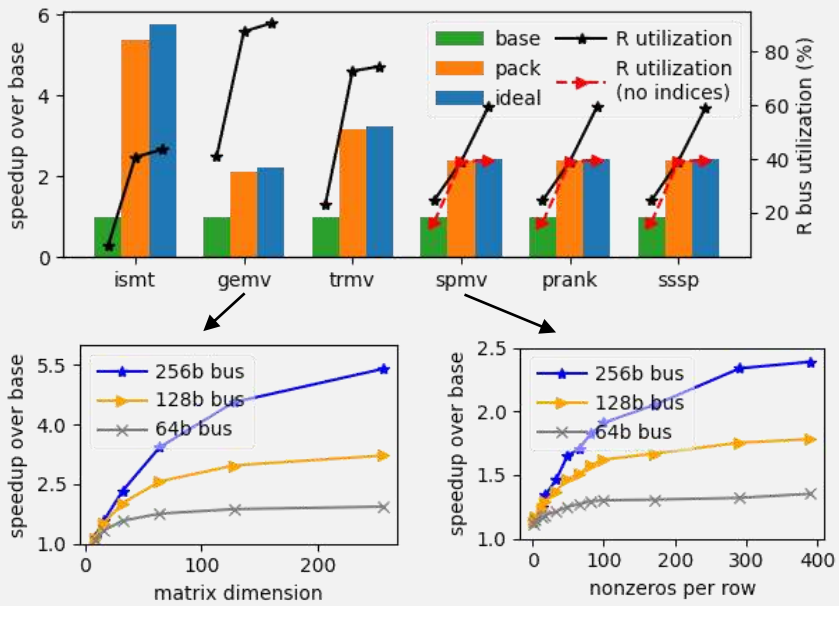
- **Define AXI4 user extension (7% more bits)**
- **Extend a RISC-V vector processor**
- **Design an AXI-Pack adapter for banked Memory**



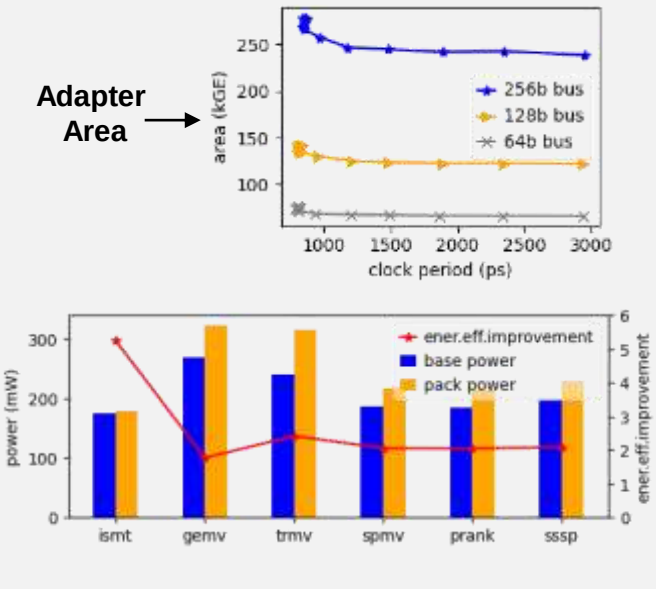


5 Results

- **Speedup irregular workloads**
  - 5.4x (stride)
  - 2.4x (indirect)
- **light-weight and scalable**
  - 6.2% extension area
- **Improve energy efficiency**
  - 5.3x (stride)
  - 2.1x (indirect)



Adapter Area



References

1. Domingos, Joao Mario, et al. "Unlimited vector extension with data streaming support." 2021 ISCA
2. Tanabe, Noboru, et al. "A memory accelerator with gather functions for bandwidth-bound irregular applications." Proceedings of the 1st Workshop on Irregular Applications: Architectures and Algorithms. 2011.
3. Arm, “AMBA AXI and ACE Protocol Specification,” https://developer.arm.com/documentation/ihi0022/hc.