

Lightweight Capability Compression for Bandwidth-Amplified CHERI Workloads

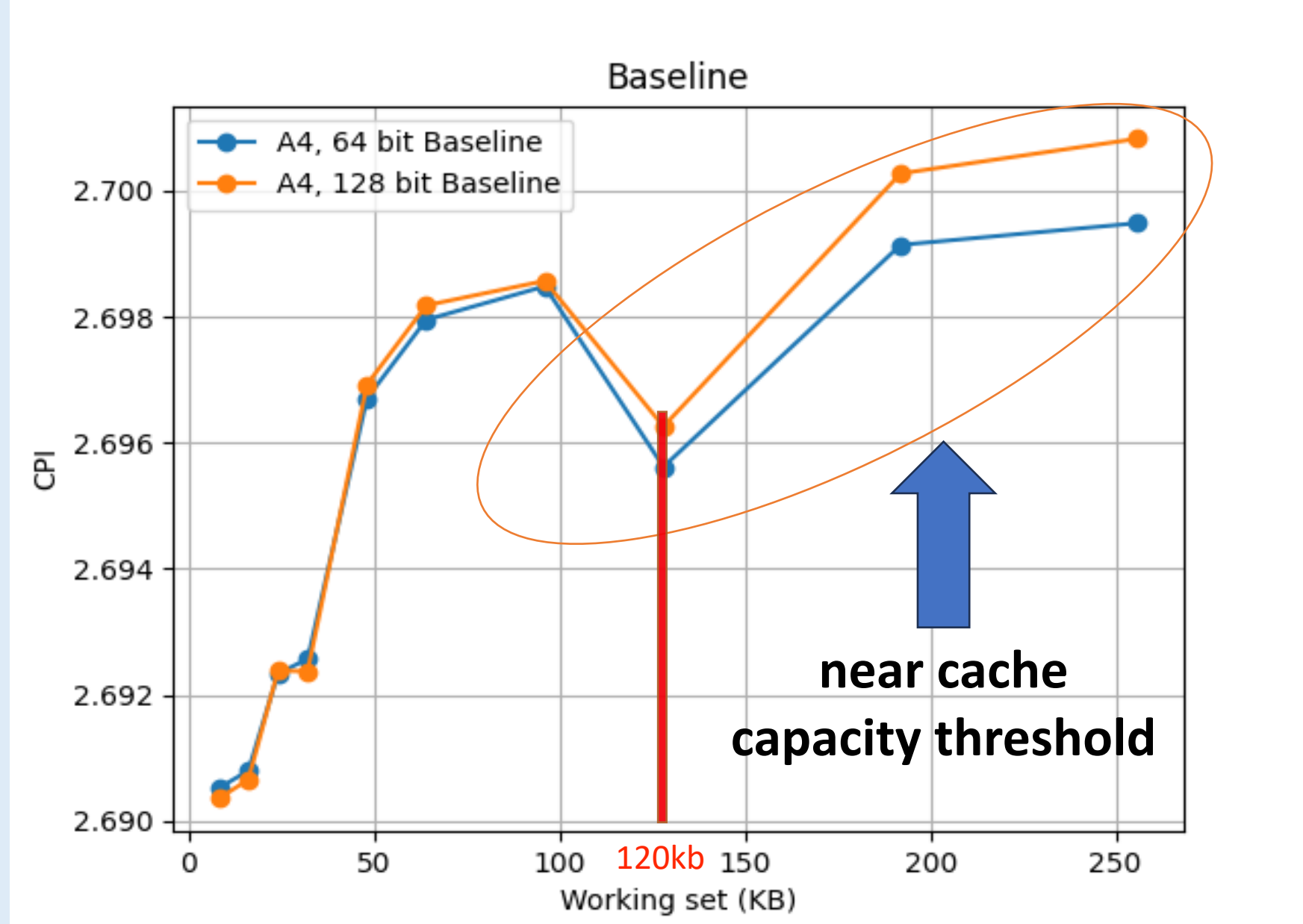
A Case Study on Capability Width and Memory Hierarchy Amplification

Jici Li, University of Edinburgh

Motivation & Setup:

- CHERI uses 128-bit capability
 - Wider representation increases memory traffic
 - Overheads is often assumed uniform
- Question:** *Where is this overhead actually amplified?*

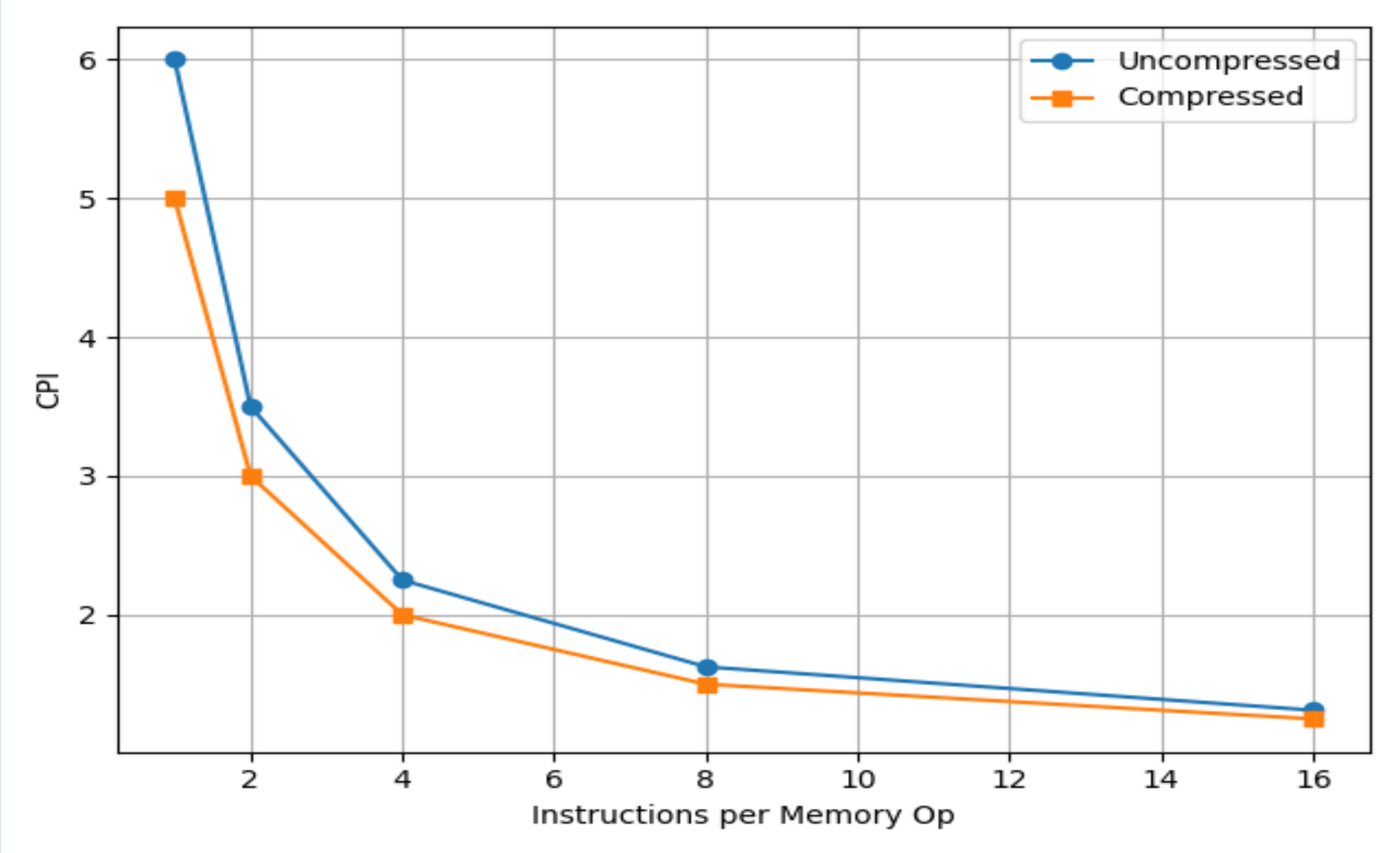
Key Observation:



Observation: CHERI overhead is **not** intrinsic latency
It is **bandwidth-amplified** beyond cache thresholds

- CPI nearly identical in cache-resident regime
- Divergence emerges beyond cache capacity

Results:

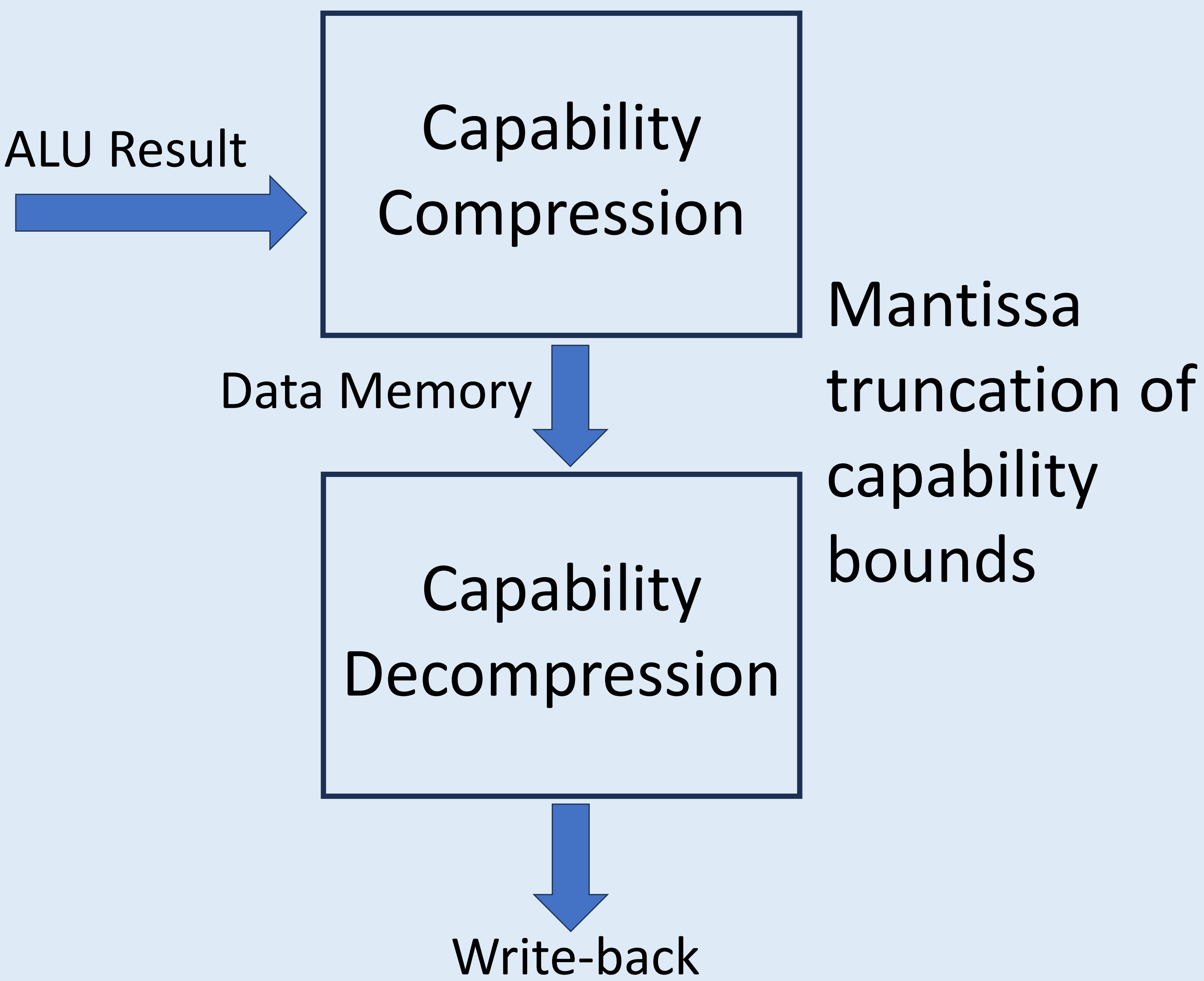


- Identical hardware footprint: 321 cells
- Memory-bound workloads:
CPI ↓ 9.54%

Design:

Target: **Memory traffic reduction**

- Write-back stage
- No pipeline restructuring
- No decompression latency
- No increase in logic cells



Cost & Trade-off:

| | Compressed | Uncompressed |
|-------------|------------|--------------|
| SB_LUT4 | 134 | 134 |
| SB_DFFESR | 2 | 2 |
| SB_CARRY | 30 | 30 |
| SB_CARRY | 59 | 59 |
| Total cells | 321 | 321 |

- Bounded precision loss ($\leq 2^k$ bytes)
- Upper bounds preserved
- Safety semantics unchanged

CHERI performance degradation is bandwidth-driven not pipeline-bound