Eliminating Dark Bandwidth

ARM

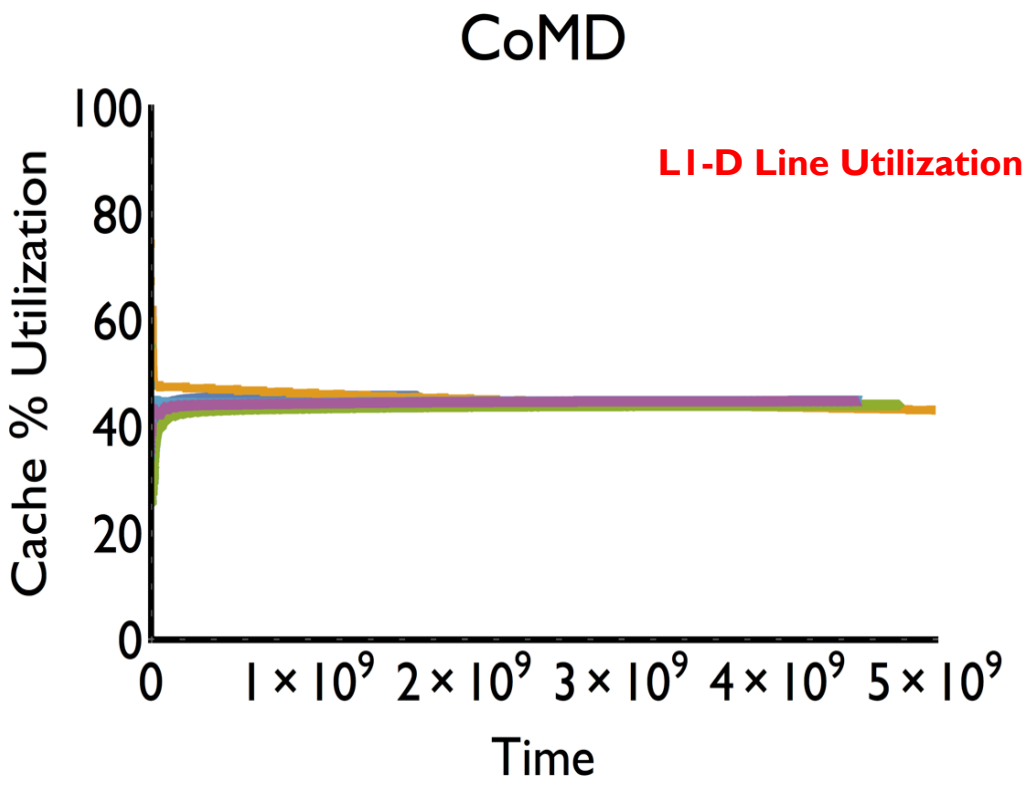
Jonathan Beard (twitter: @jonathan_beard)
Staff Research Engineer
Memory Systems, HPC

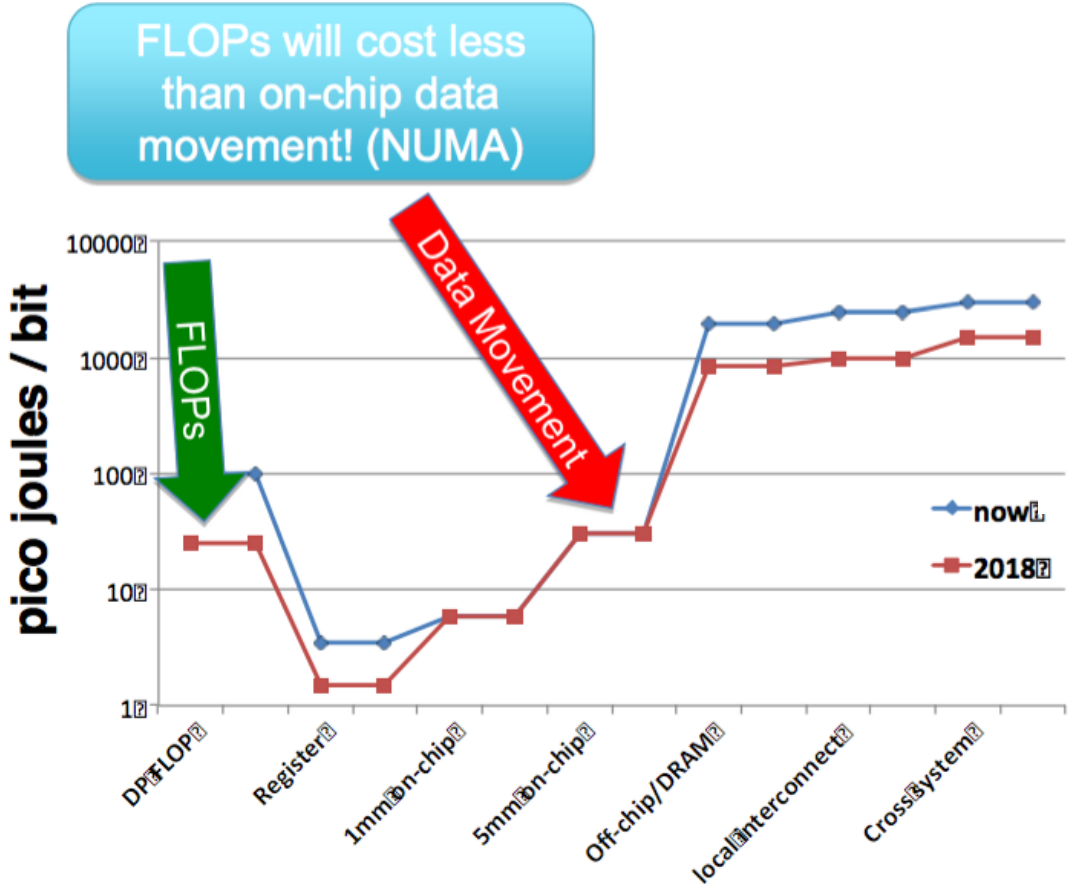
HCPM 2017
22 June 2017

Dark Bandwidth:

Data moved throughout the memory hierarchy that is not used after it is moved.

Data Movement Dominates





Source: Simon, Horst. "Why we need exascale and why we won't get there by 2020." Opt. interconnects conf., Lawrence Berkeley National Laboratory, Santa Fe. 2013.

ARM

Superfluous Data Movement

Utilized Bandwidth

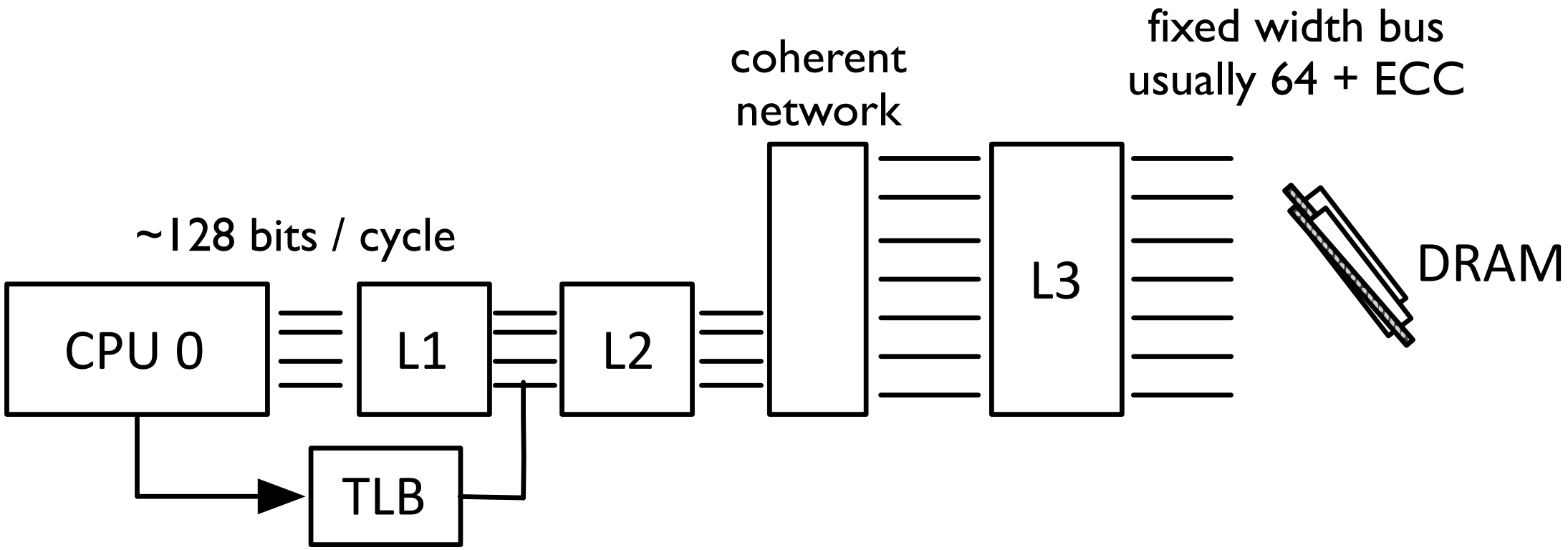
Dark Bandwidth

ARM

By eliminating Dark Bandwidth, we can make computing more efficient, enabling performance gains despite the impending end of Moore's Law.

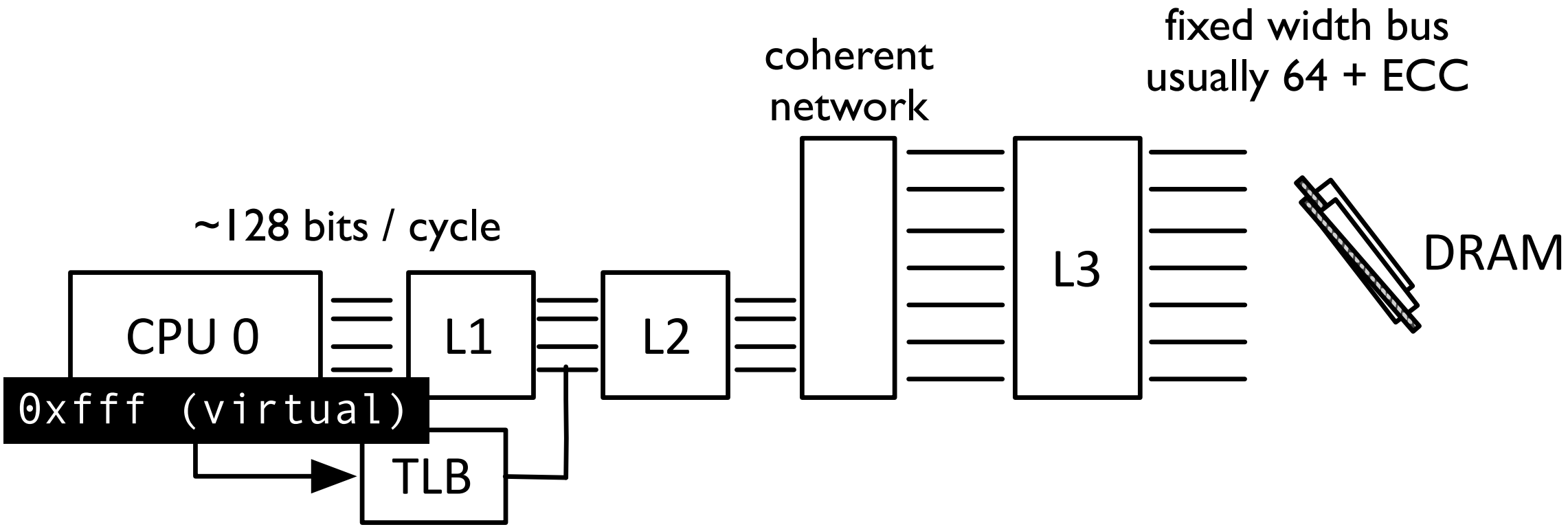
How we get data

```
move    -16(%r1), %r2
```



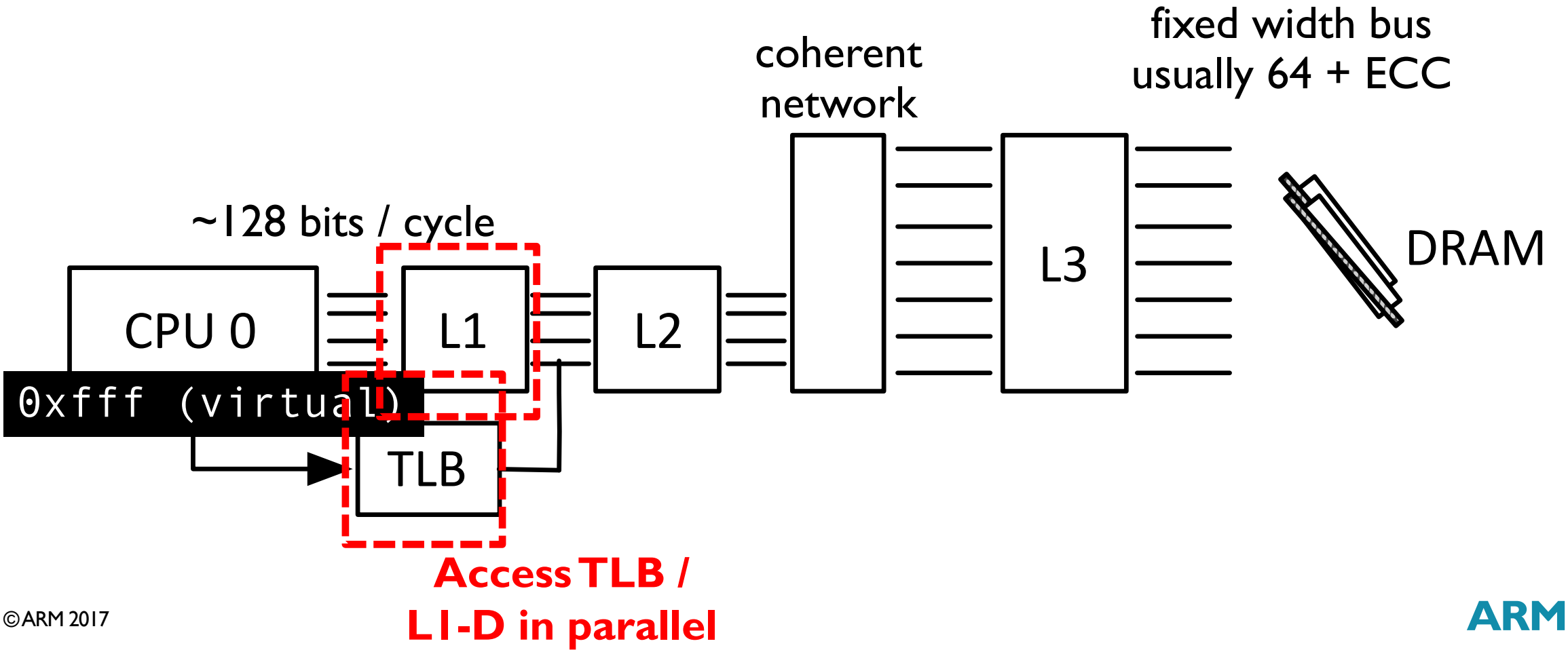
How we get data

```
move    -16(%r1), %r2
```



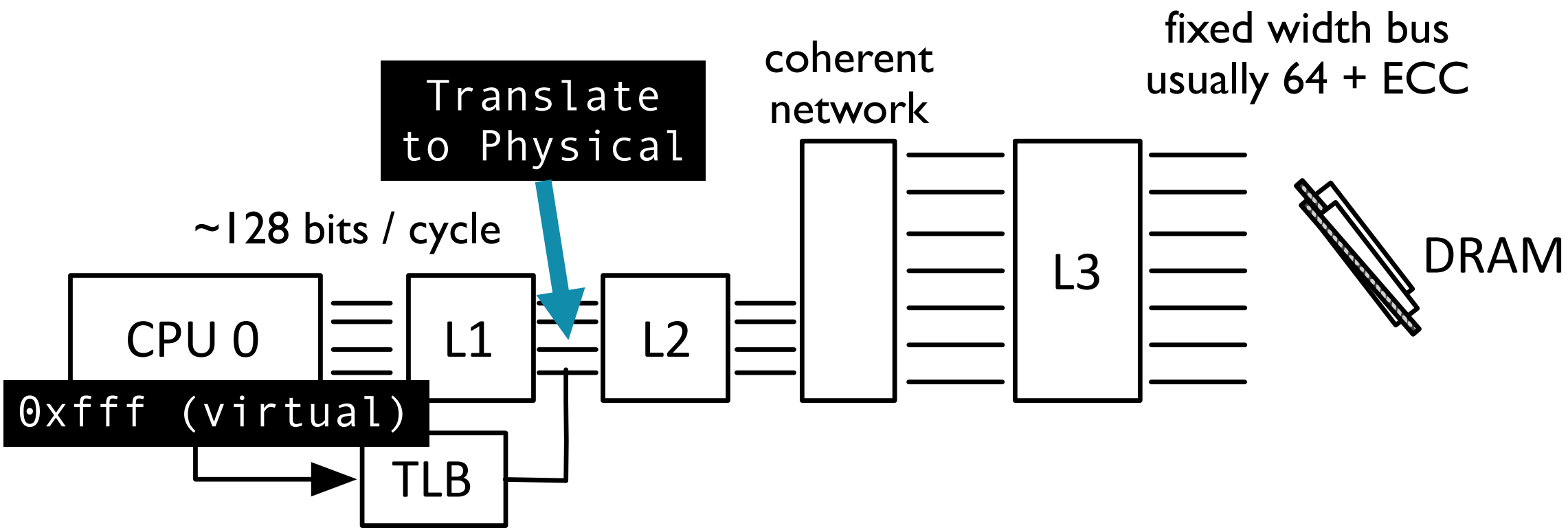
How we get data

```
move    -16(%r1), %r2
```



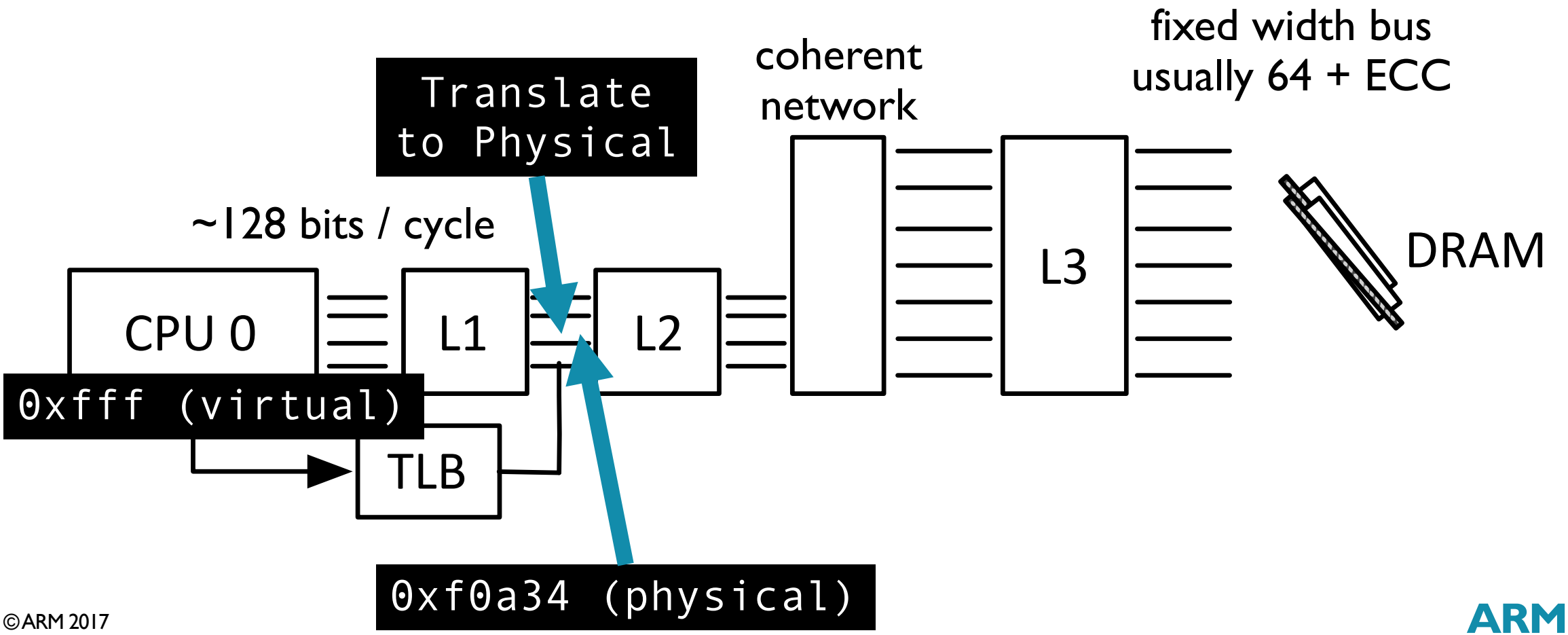
How we get data

```
move    -16(%r1), %r2
```



How we get data

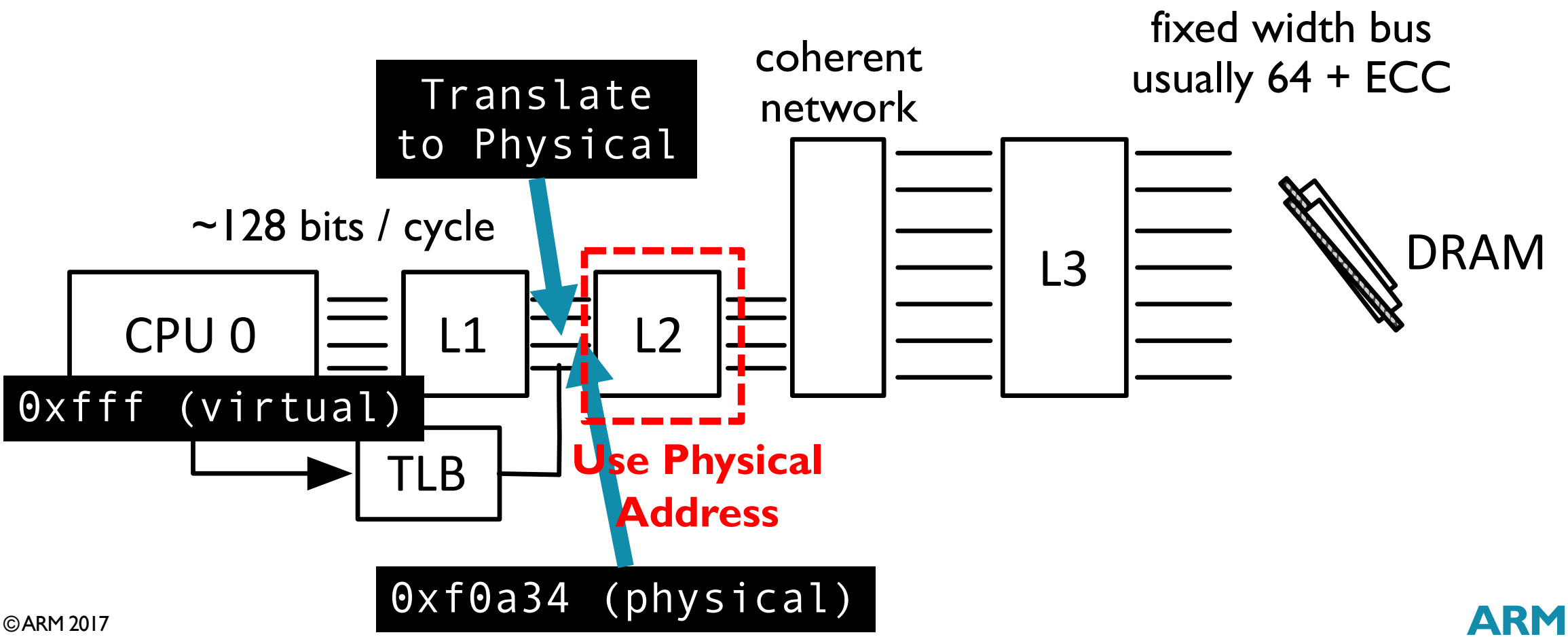
```
move    -16(%r1), %r2
```



ARM

How we get data

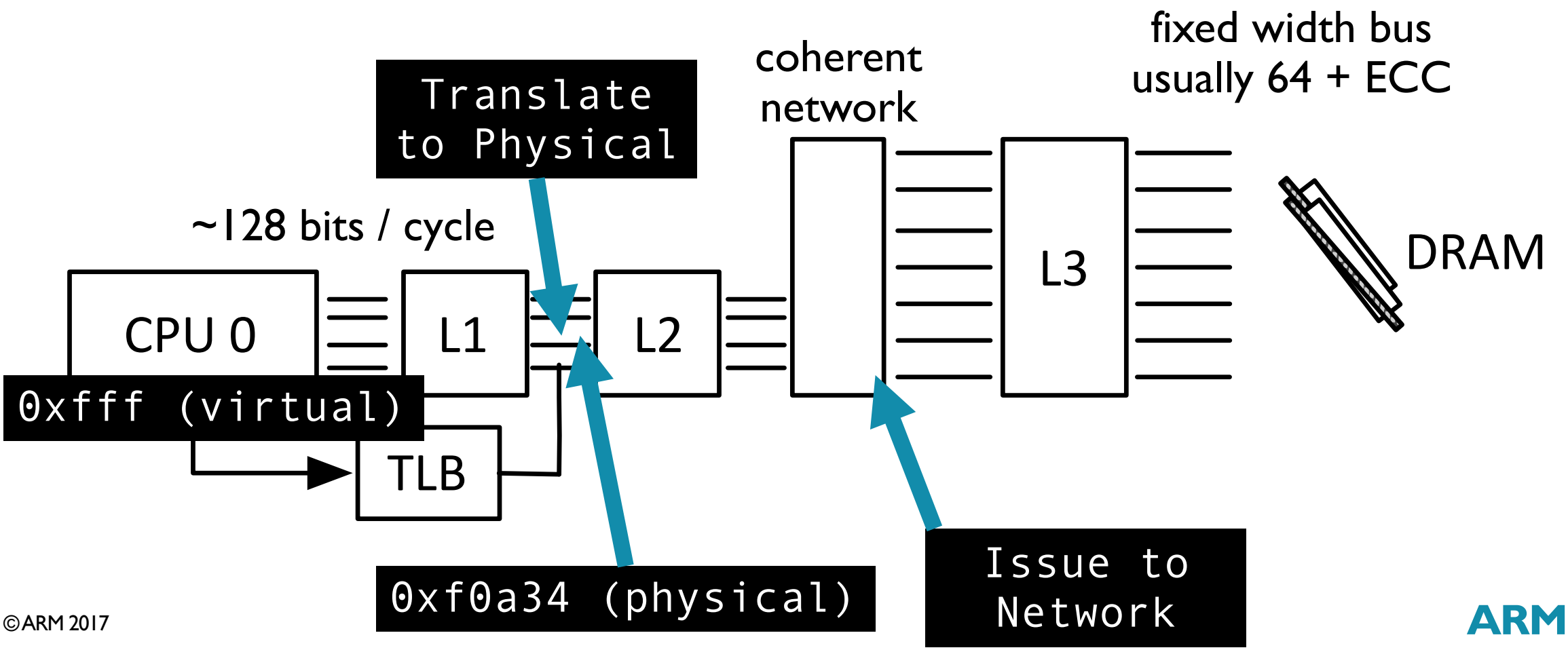
```
move    -16(%r1), %r2
```



ARM

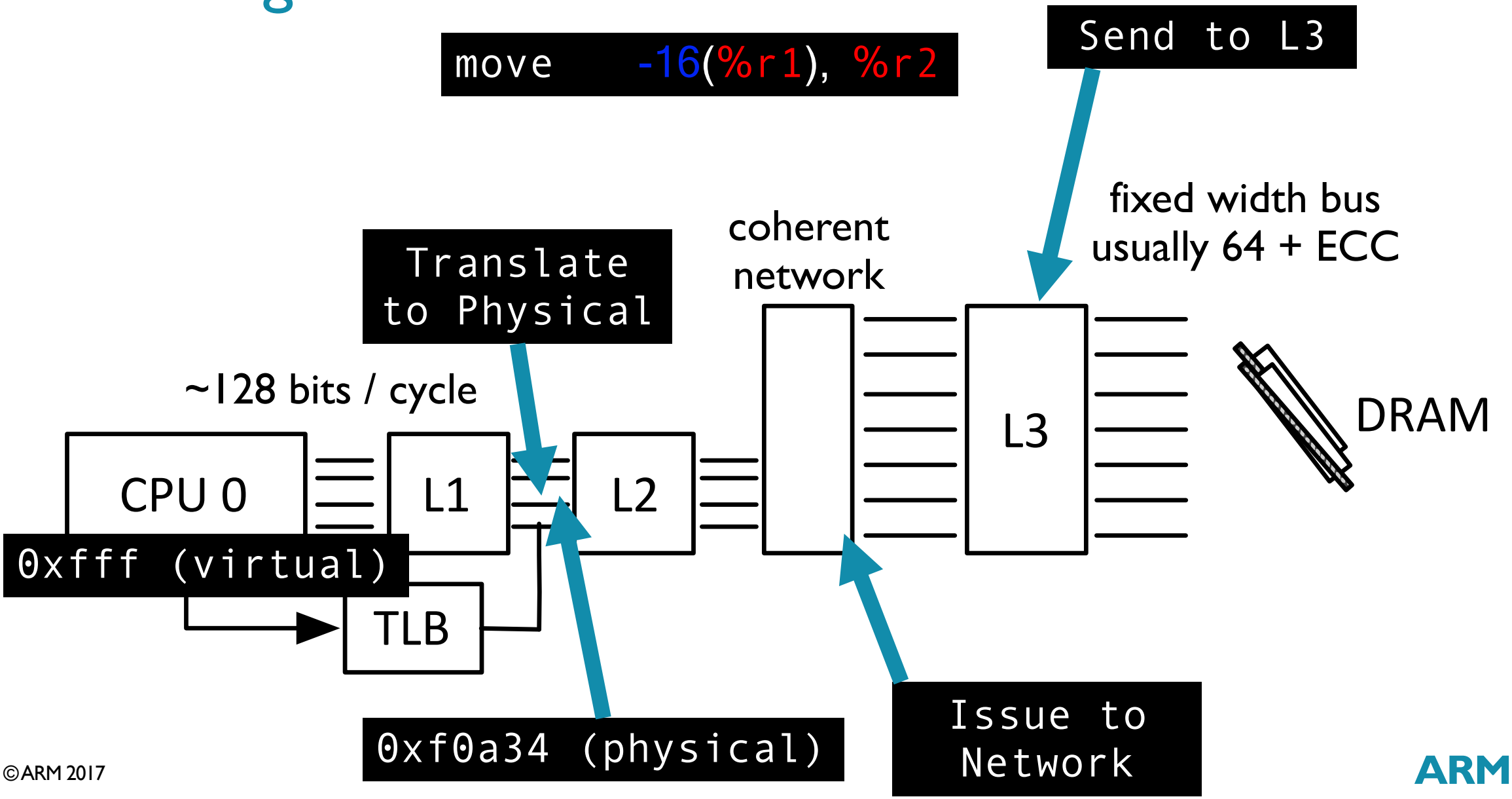
How we get data

```
move    -16(%r1), %r2
```



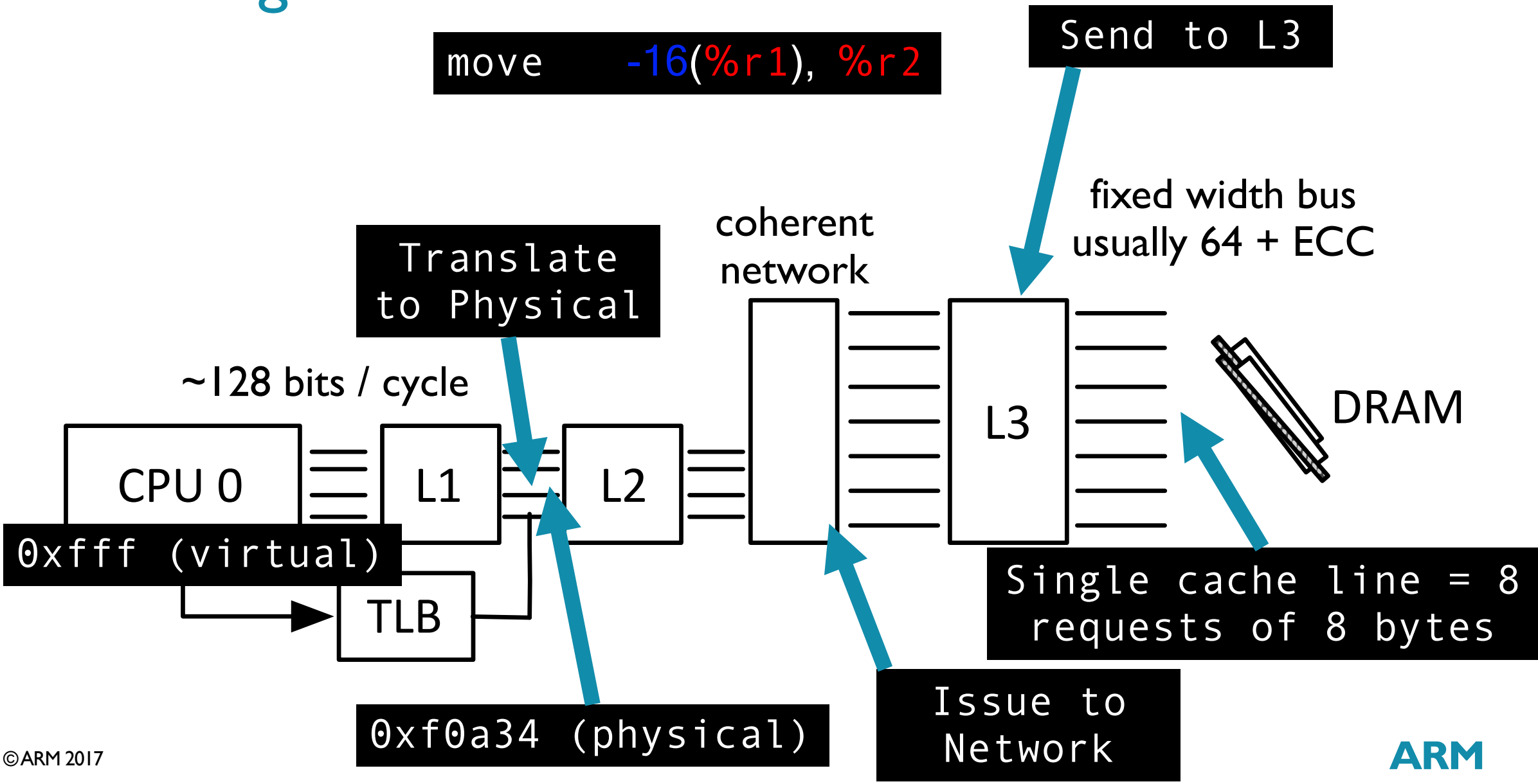
ARM

How we get data



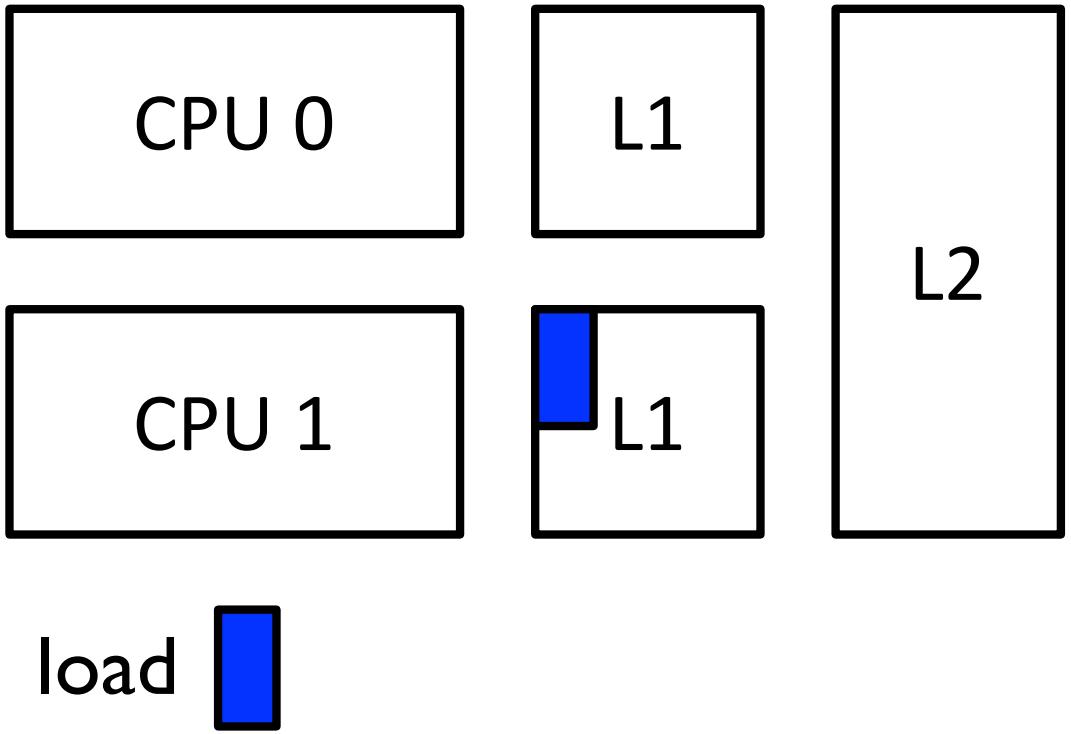
ARM

How we get data



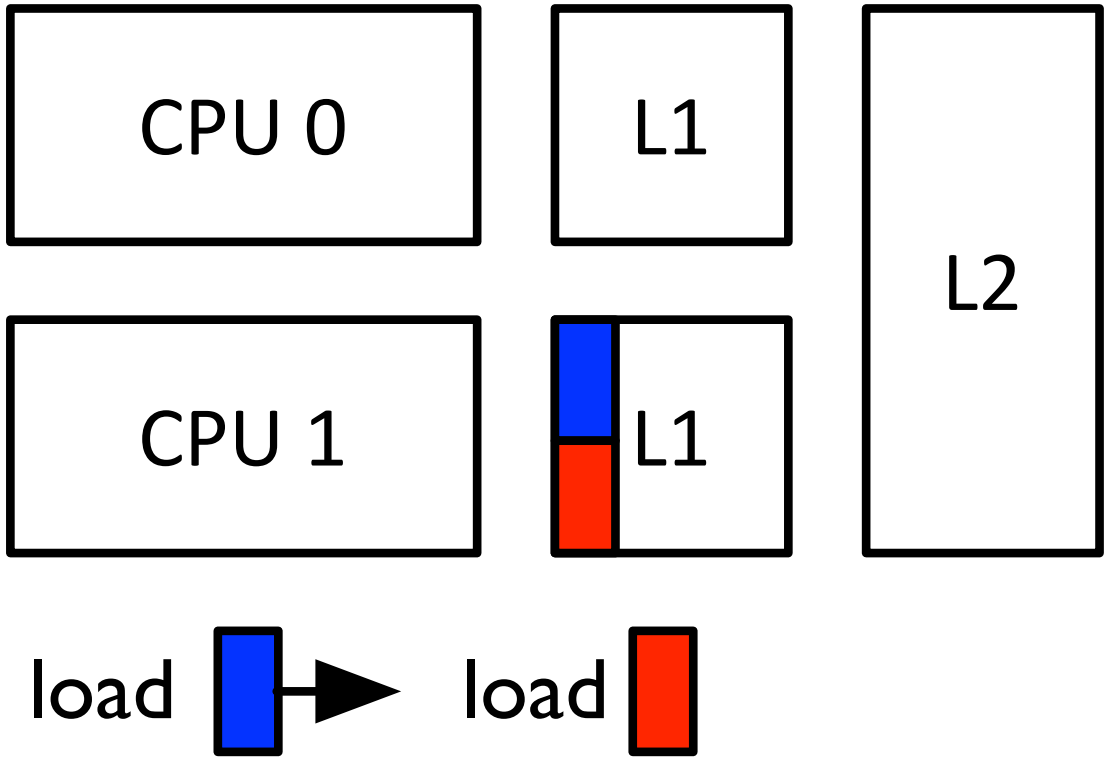
Reuse Distance

- If data within a cache line is used more than once, then the cache line is said to be reused
- We need to answer: how many other cache lines are used in between accesses?



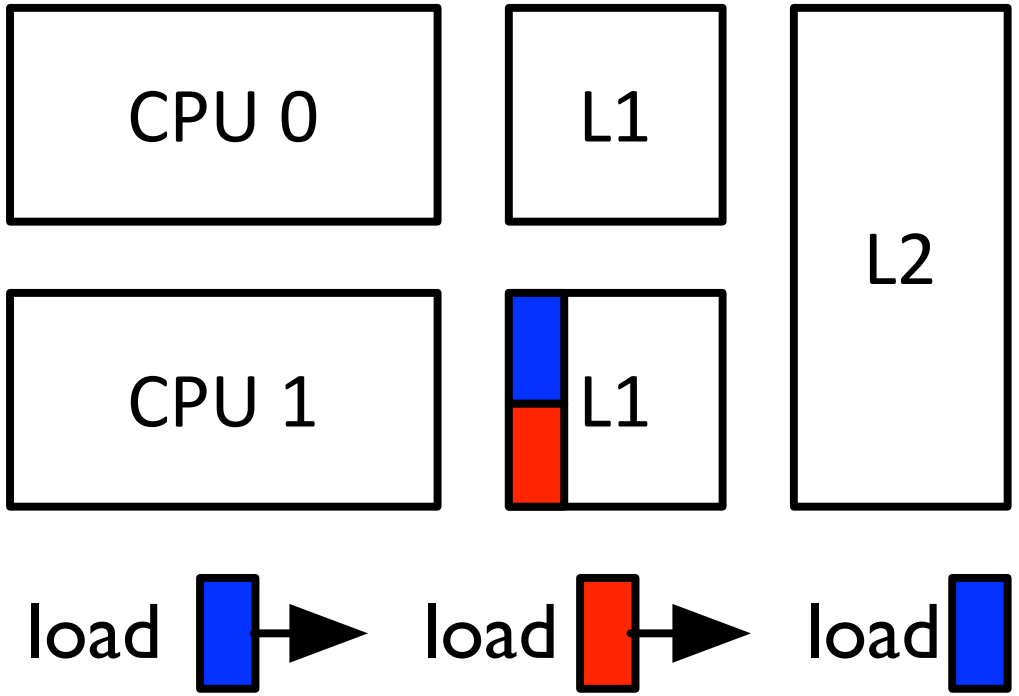
Reuse Distance

- If data within a cache line is used more than once, then the cache line is said to be reused
- We need to answer: how many other cache lines are used in between accesses?



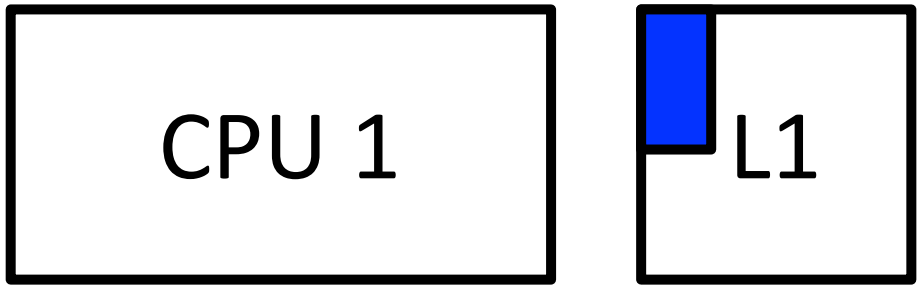
Reuse Distance

- If data within a cache line is used more than once, then the cache line is said to be reused
- We need to answer: how many other cache lines are used in between accesses?
- Blue has a reuse distance of one if the metric is in units of a cache line

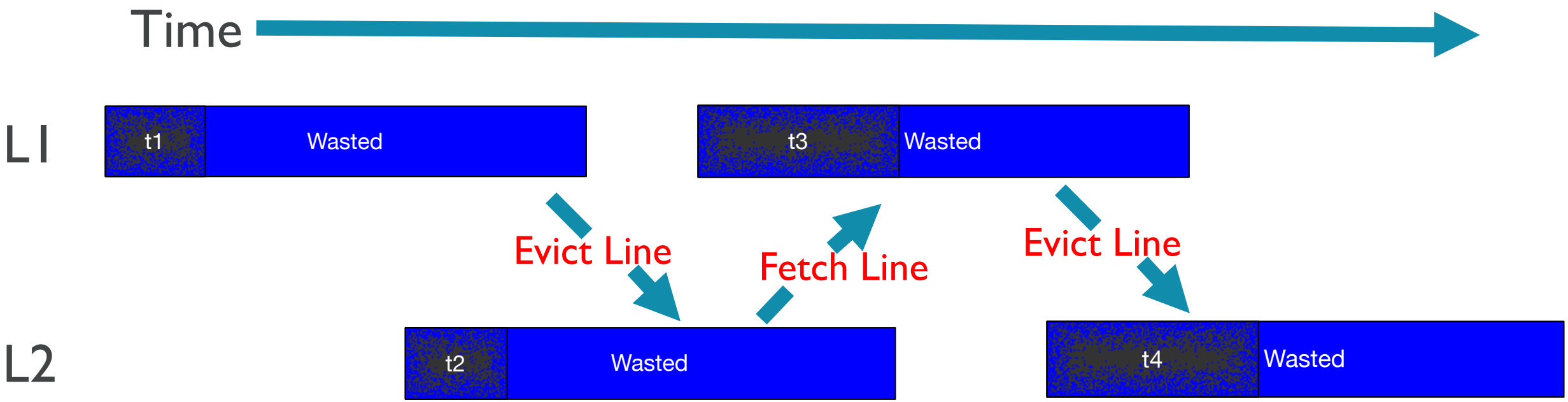


Utilization

- How much of a cache line is used before it is evicted?
- **We moved all the data labeled waste through that entire maze of wires only to evict it again!!**

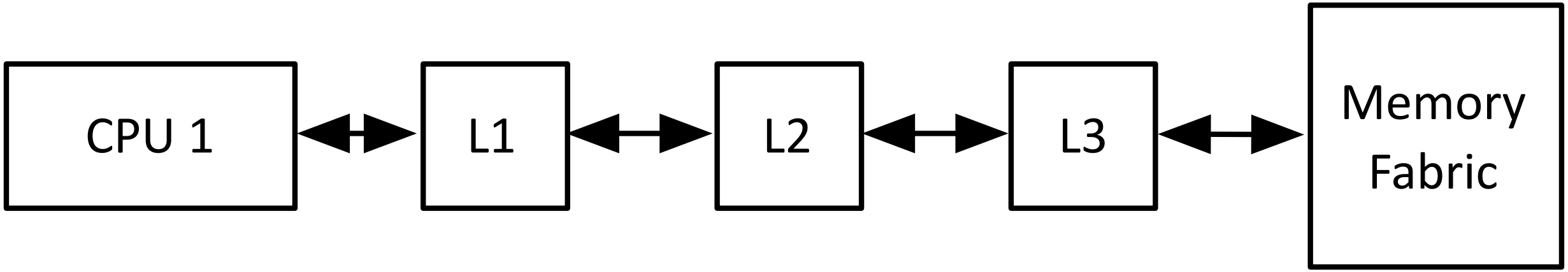


Dance of reuse and utilization



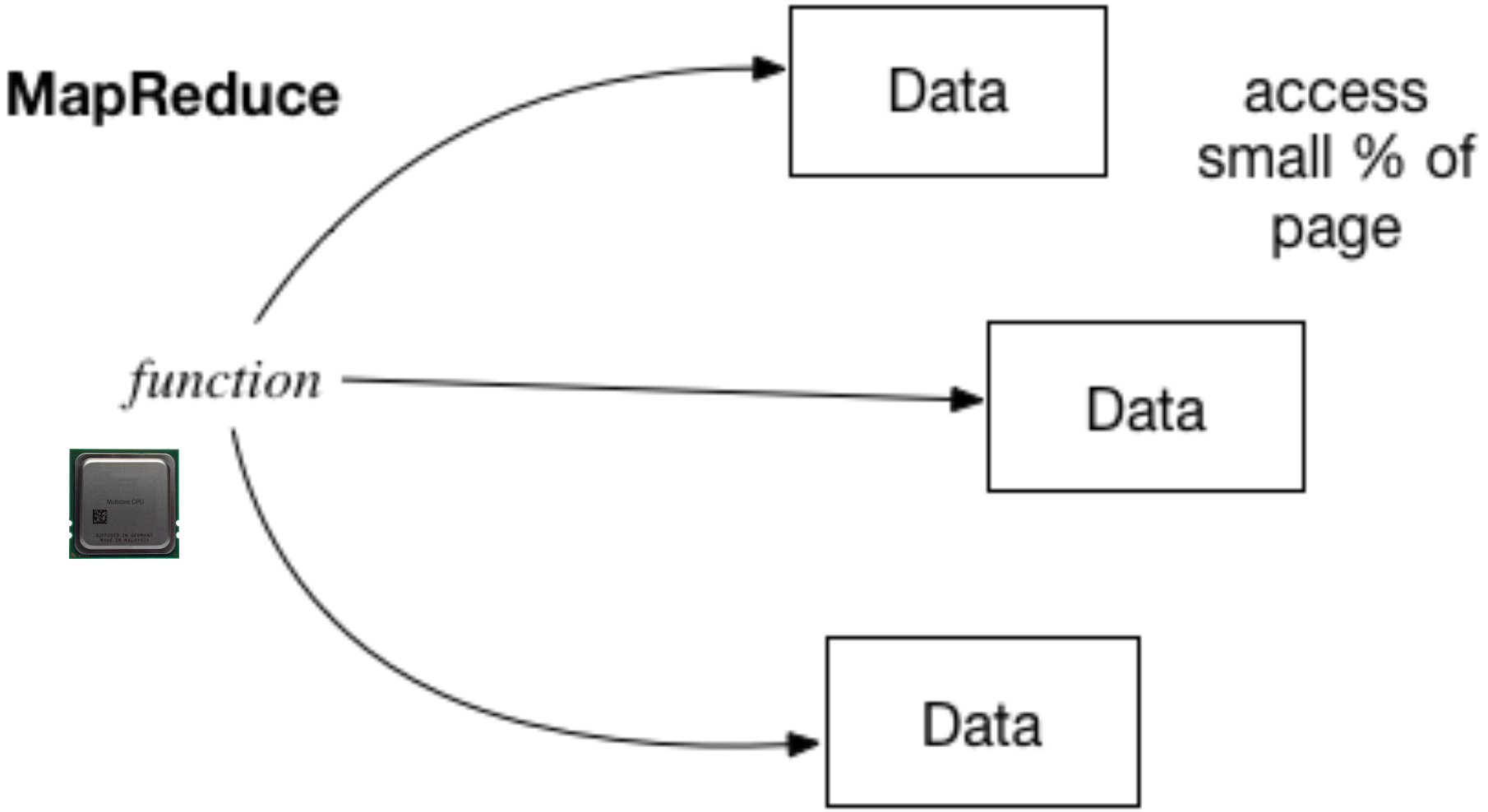
A balanced system

- Dark bandwidth results in:
  - **More energy**
  - **Higher latency**
  - **Less usable bandwidth between levels of cache hierarchy**

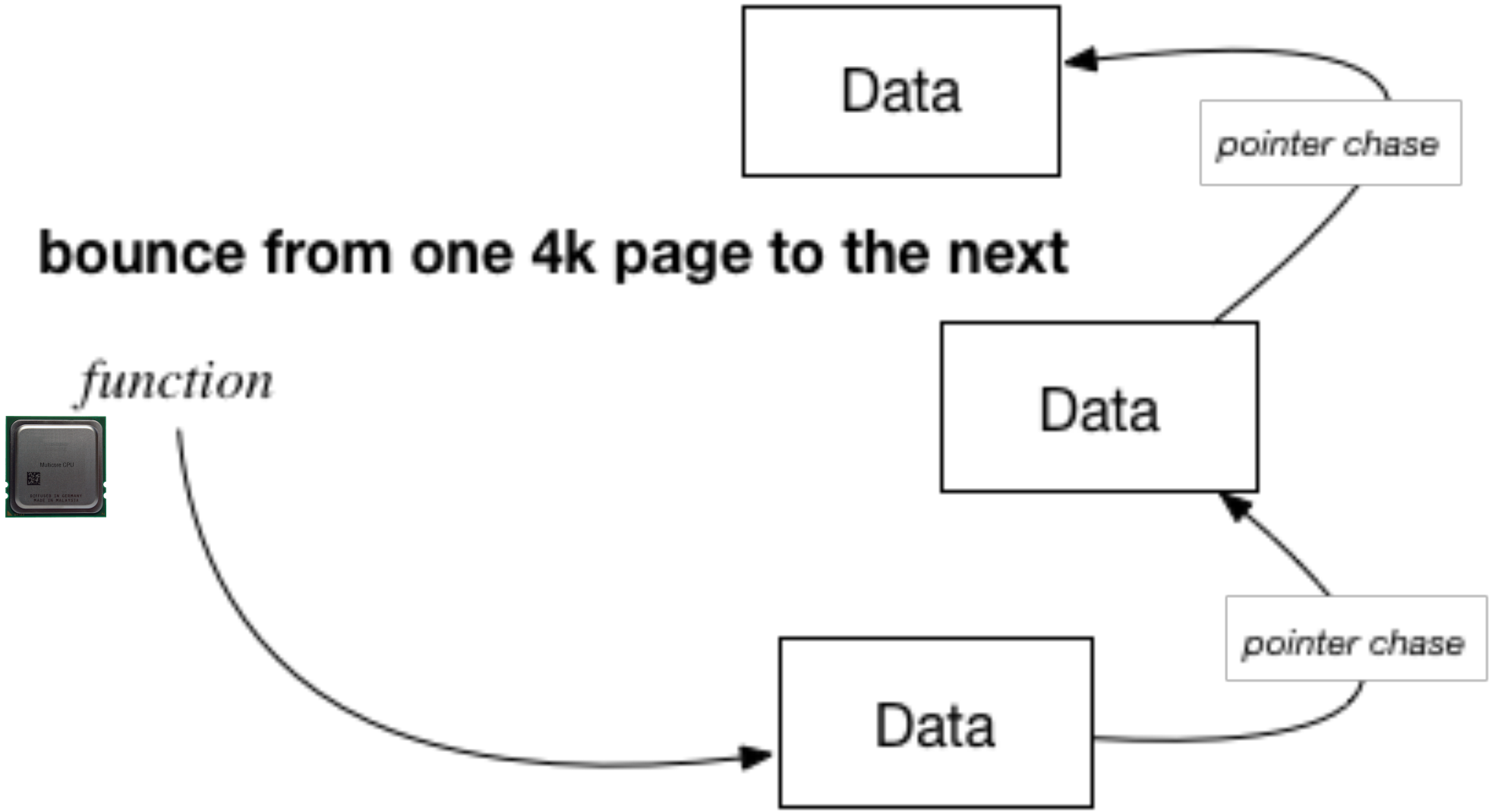


Solutions

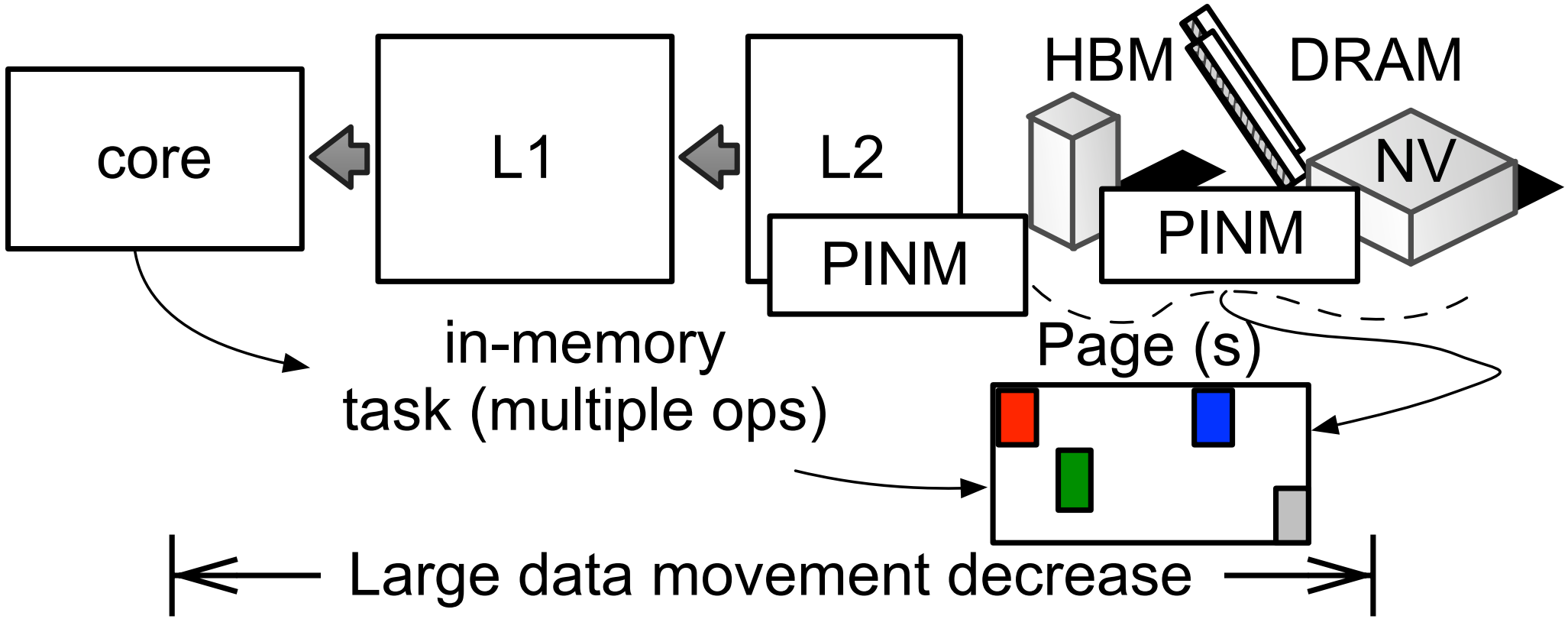
Moving Instructions to Data – Map Reduce



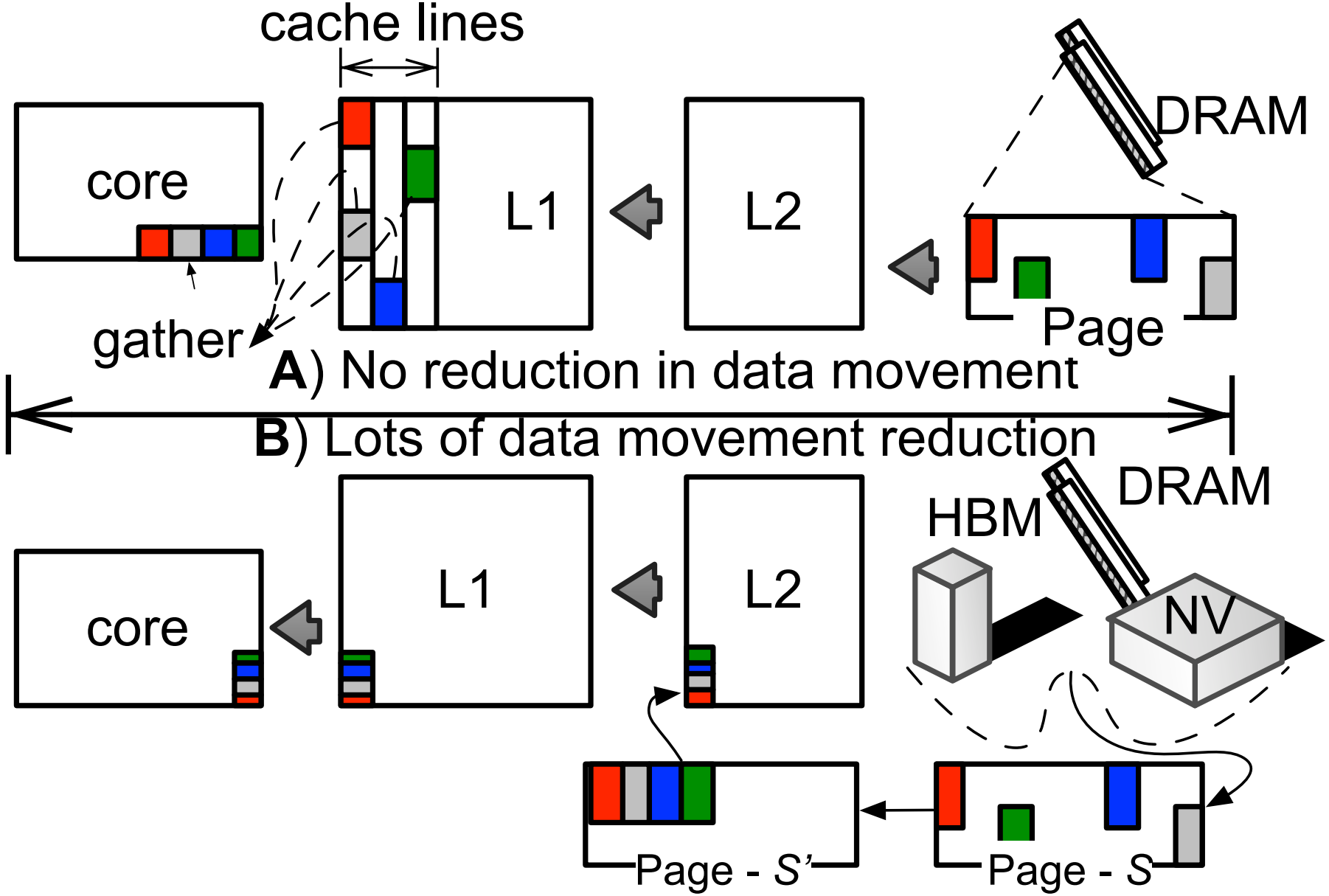
Moving Instructions to Data – Pointer Chasing



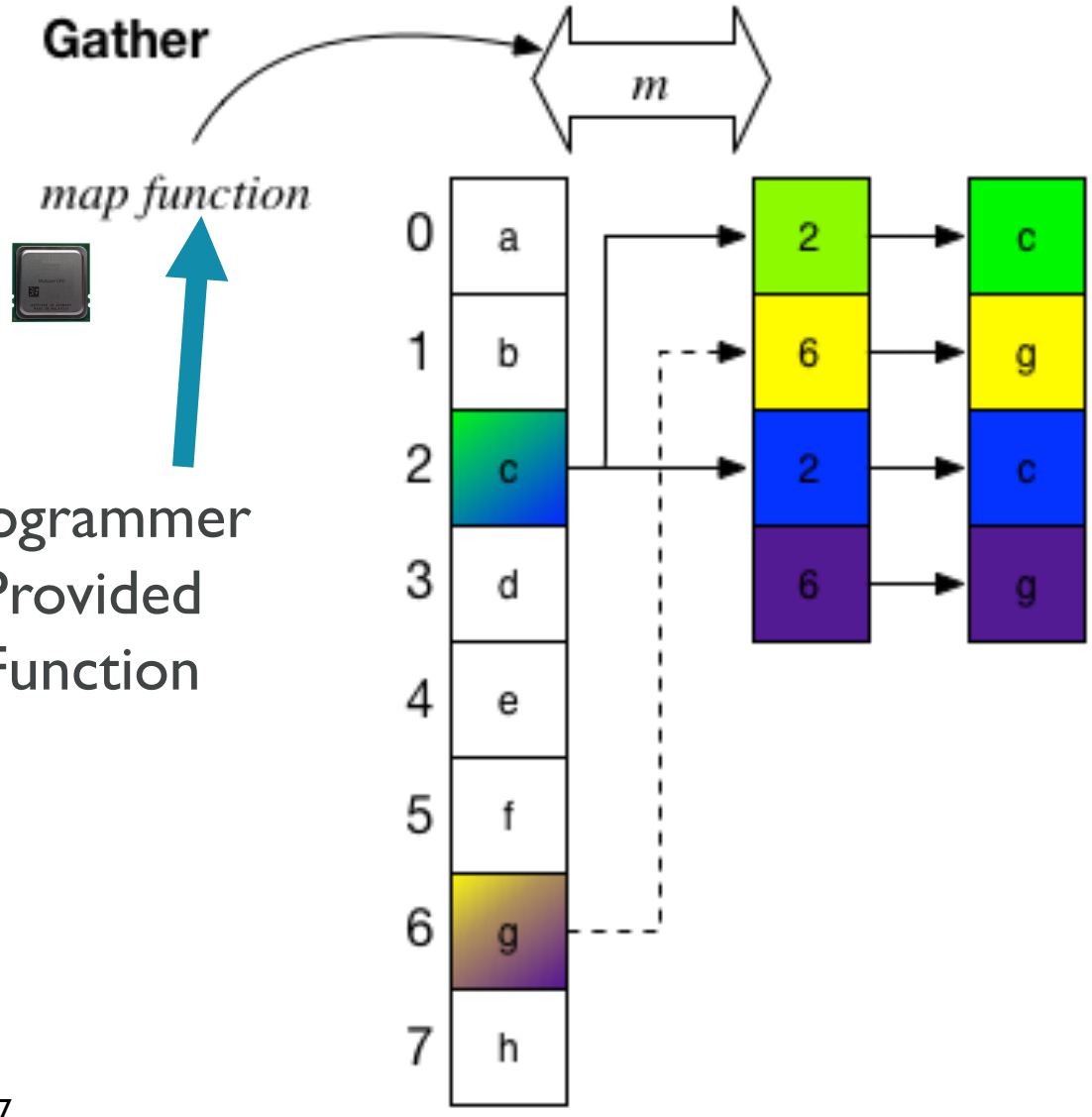
Processing In-/Near-Memory



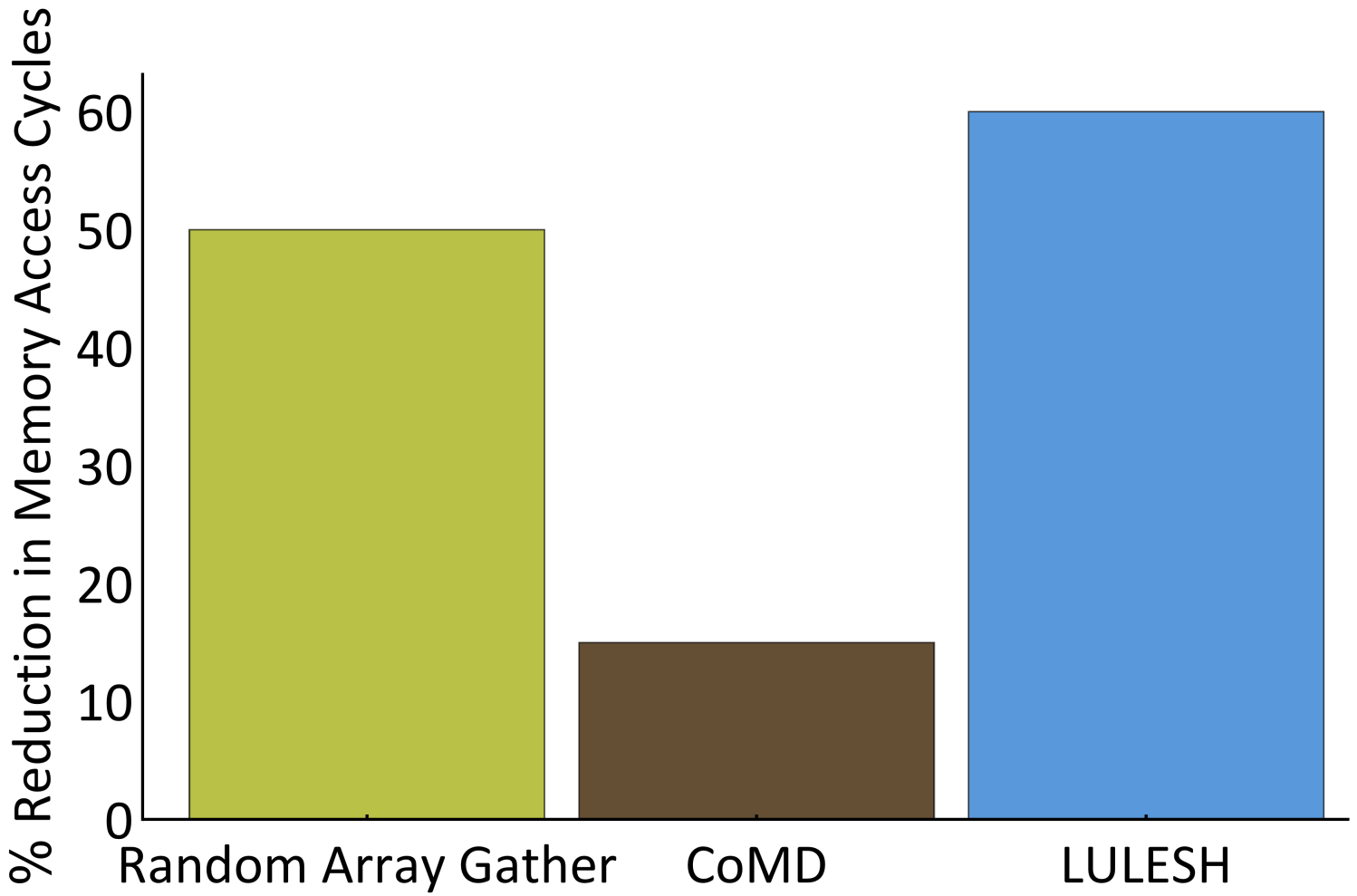
Gather-Scatter



Near-memory Gather-Scatter



Data Reduction Potential (preview of MEMSYS17)

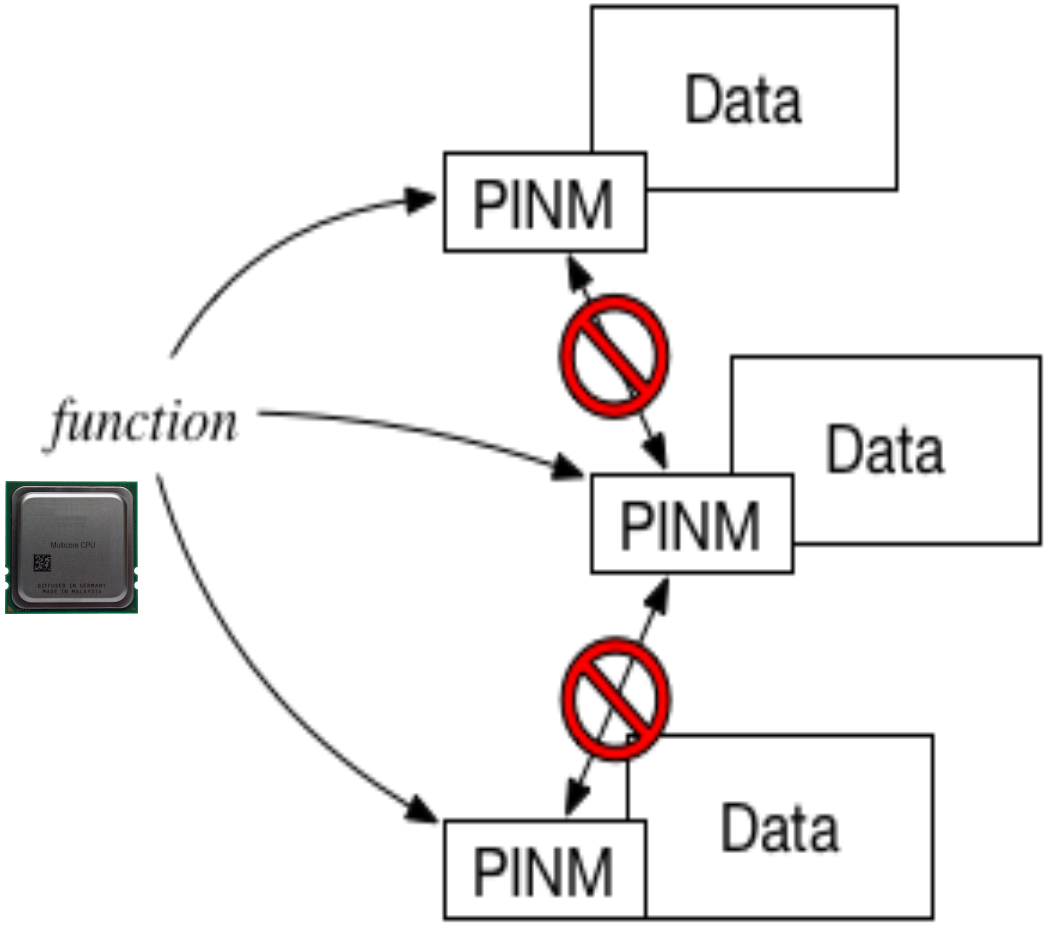


ARM

*Cycles based on simulation model with 4-cycle L1-D latency, 14-cycle L2 latency

“Ideal” PINM Characteristics

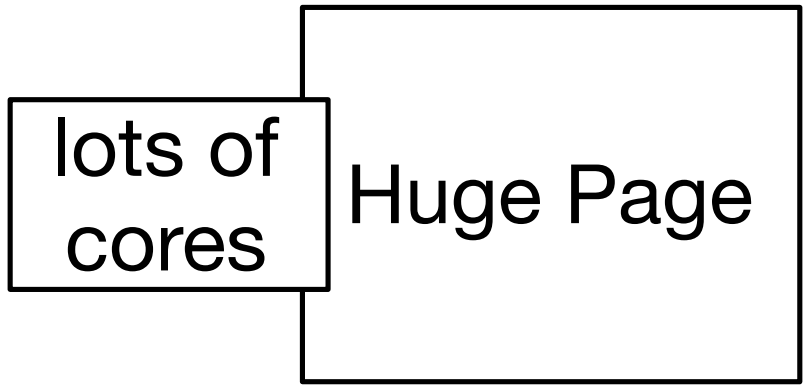
- Many small cores near memory
- Maximize number of independent functions
- Minimize communications between PINM units
- Minimize communication to main cores
- Determine address and bounds of data to be operated on as early as possible to issue to memory



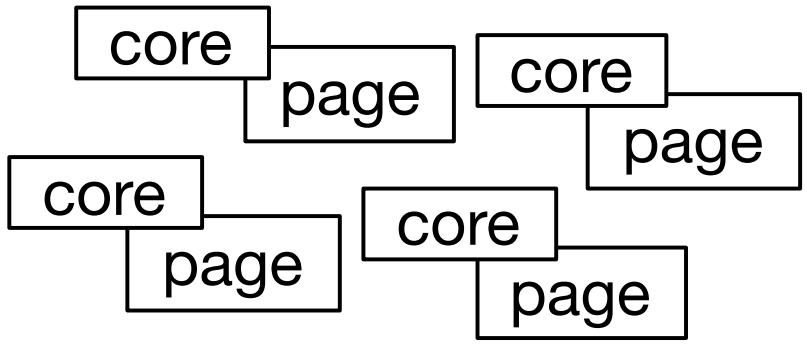
Lost in Translation

Two Options

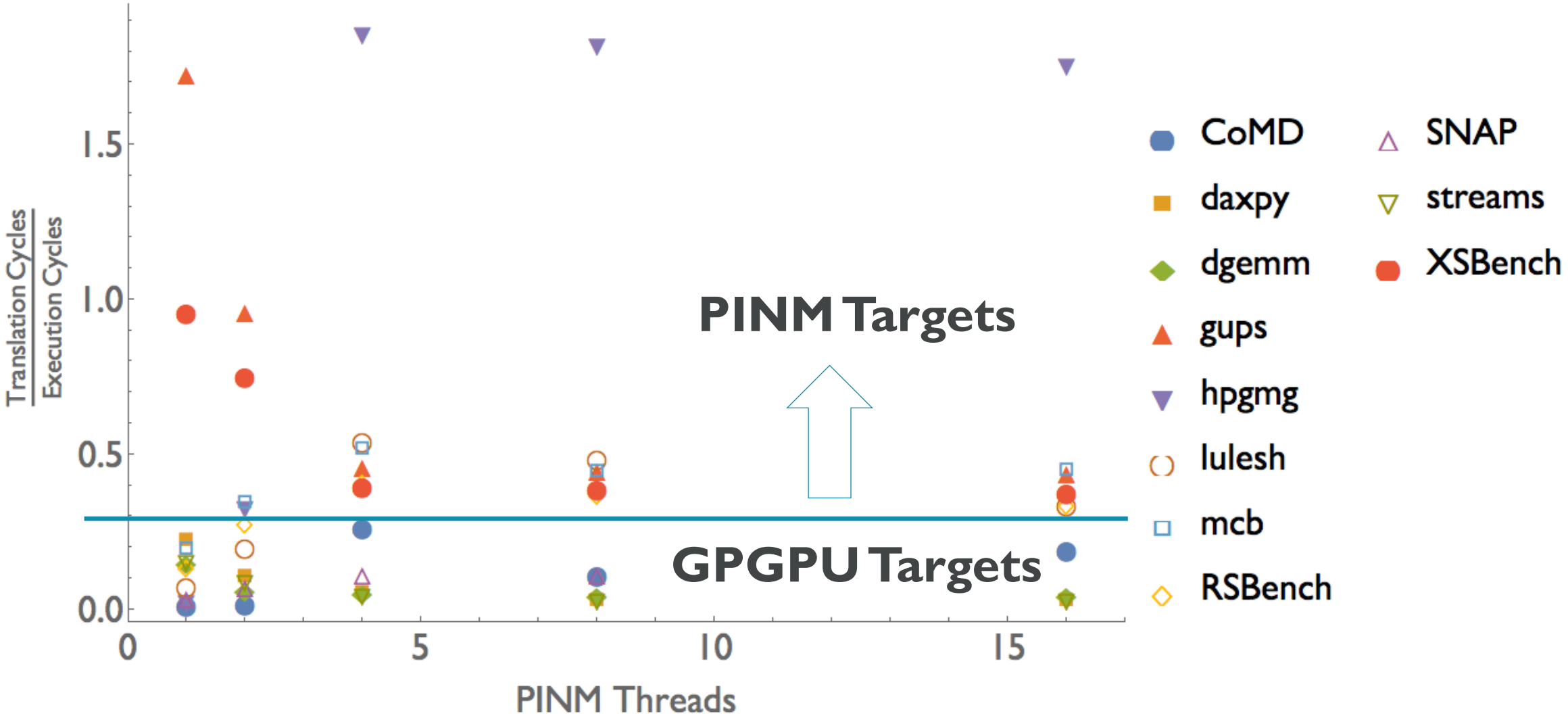
- Contiguous pages / huge pages
  - Well suited for some things, but not for others
  - PINM devices need more hardware for coherence/synchronization
  - Pointer chasing outside page not easy



- Small pages
  - Suited for all applications
  - PINM device simpler
    - page level synchronization
    - limited coherence
  - Pointer chasing outside of page not easy, maybe IOMMU?



How Bad Can It Be?



ARM

A more addressable future for accelerators

Folding (a.k.a. Manual Virtual Memory)

- Virtual memory started b/c of lack of available memory vs. storage
- Programmers wrote code that manually folded/unfolded to storage
- Huge controversy existed b/c all code at the time was written this way – Automatic folding hardware was ~25% slower than manual folding

"In ceasing to expend energy (item 1) **in a process whose main result is to make programs less fit to run on other machine configurations** (item 2), or **to run in company with other programs** (item 3), or to **run with temporarily reduced resources** (item 4), we do more than reduce costs; **we remove self-created obstacles** which today are impeding the development of needed types of systems"

- D. Sayre, IBM Yorktown Research (1969)

1970 - The Circle of Tech - "Self Created Obstacles"

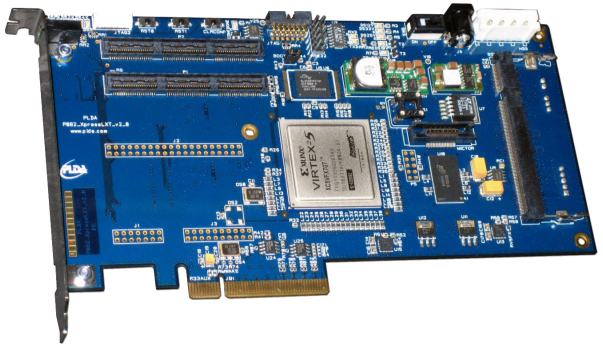
Move In

Move Out

2017 - The Circle of Tech - "Self Created Obstacles"

Move In

Move Out



ARM

Can we live with virtual memory?

- Yes, but for how long?
  - Data structures for specific systems (embed translations)
  - Abstractions to made to extend VM try to make it better, but really make it worse (unified virtual addressing).
  - Bigger hardware page cache's only mean extending the current system for a few years at best.
- Fundamental flaw in page-based virtual memory
  - It's not scalable – single point bottleneck
  - Inherently local (page cache needed)
  - Accelerators attached as an afterthought, the future is heterogeneous.
  - Many solutions to reducing data movement hobbled by translation machinery.

"In ceasing to expend energy (item 1) **in a process whose main result is to make programs less fit to run on other machine configurations** (item 2), or **to run in company with other programs** (item 3), or to **run with temporarily reduced resources** (item 4), we do more than reduce costs; **we remove self-created obstacles** which today are impeding the development of needed types of systems"

- D. Sayre, IBM Yorktown Research (1969)

Conclusions

- Every byte move costs:
  - Energy, Latency, Bandwidth
- A balanced system could consist of:
  - Heavyweight throughput cores
  - DMA / gather/scatter engines near memory
  - Lightweight, simple, near memory processors

Getting there requires changing the way we do virtual memory

Compute Intensity	RUD	CLU	Proposed Ideal Processing Modality
Low	High	Low	PINM
High	High	Low	Data Rearrangement
Low / High	High	High	PINM
Low / High	Low	Low	Data Rearrangement
Low / High	Low	High	Current Processor Pipeline

Twitter: @jonathan_beard
Thanks for listening!!