RowClone: In-DRAM Copy and Initialization of Bulk Data

V. Seshadri, Y. Kim, C. Fallin, D. Lee, G. Pekhimenko, R. Ausavarungnirun, Y. Luo, O. Mutlu, P. B. Gibbons*, M. A. Kozuch*, T. C. Mowry

Carnegie Mellon University ** Intel Labs Pittsburgh*

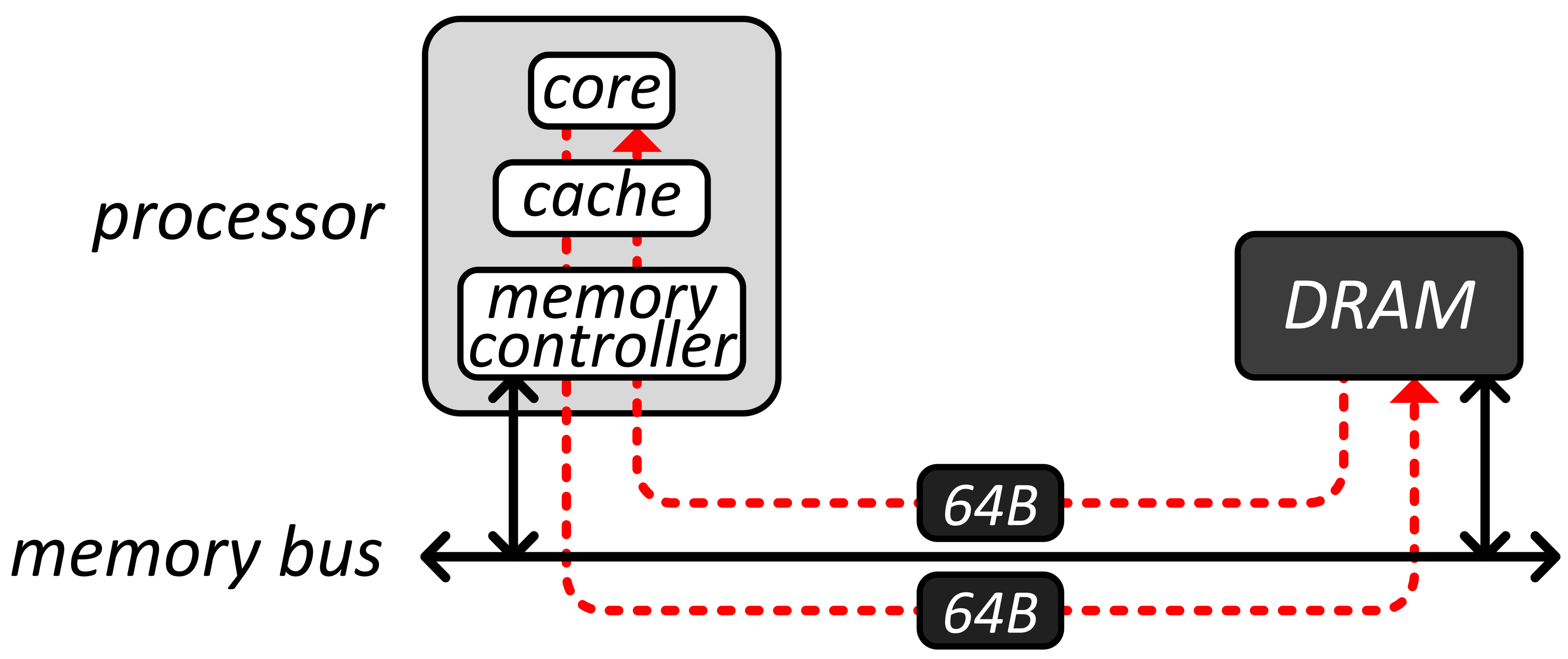
1. Introduction: Copying & Initialization

- Many applications copy or initialize large amounts of memory data
  - Example: `memcpy, memset`
- Copying/initialization *do not* require computation in order to derive their data values
  - **Copying**: simply move data that already exists
  - **Initialization**: simply reset data to constants
- Therefore, the main memory subsystem can potentially copy/initialize data *all by itself* ...
  - ... without involving the processor

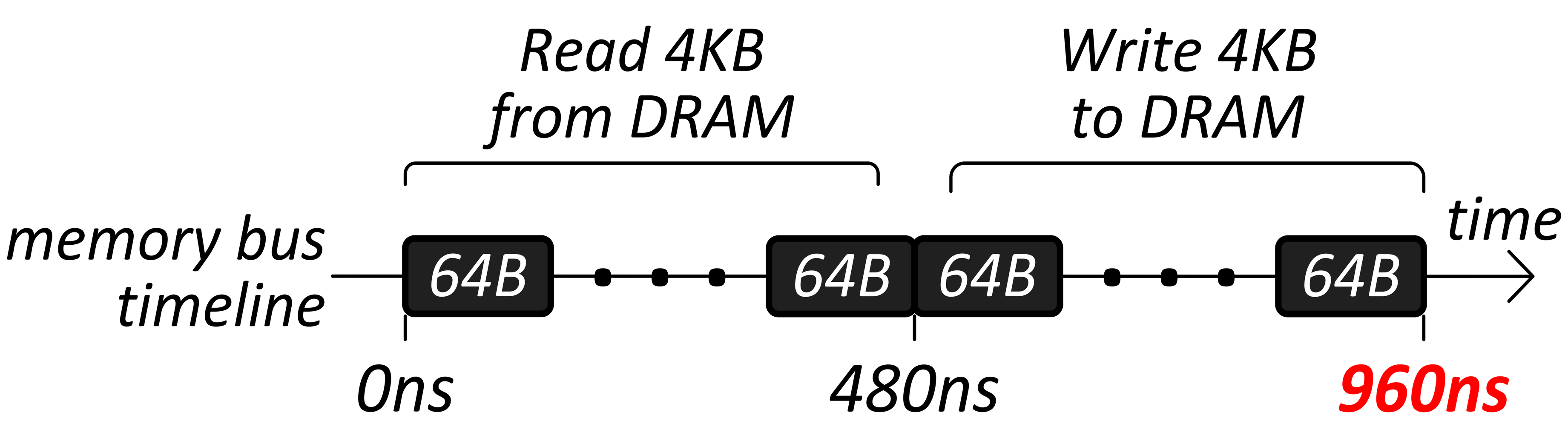
2. Problems in Existing Systems

When copying a page (4KB), for example, existing systems suffer from two major problems

- ***Problem 1.*** *Unnecessary data-transfers between the processor and main memory*



- ***Problem 2.*** *Serialized data-transfers – a page is copied one small cache-line (64B) at a time*

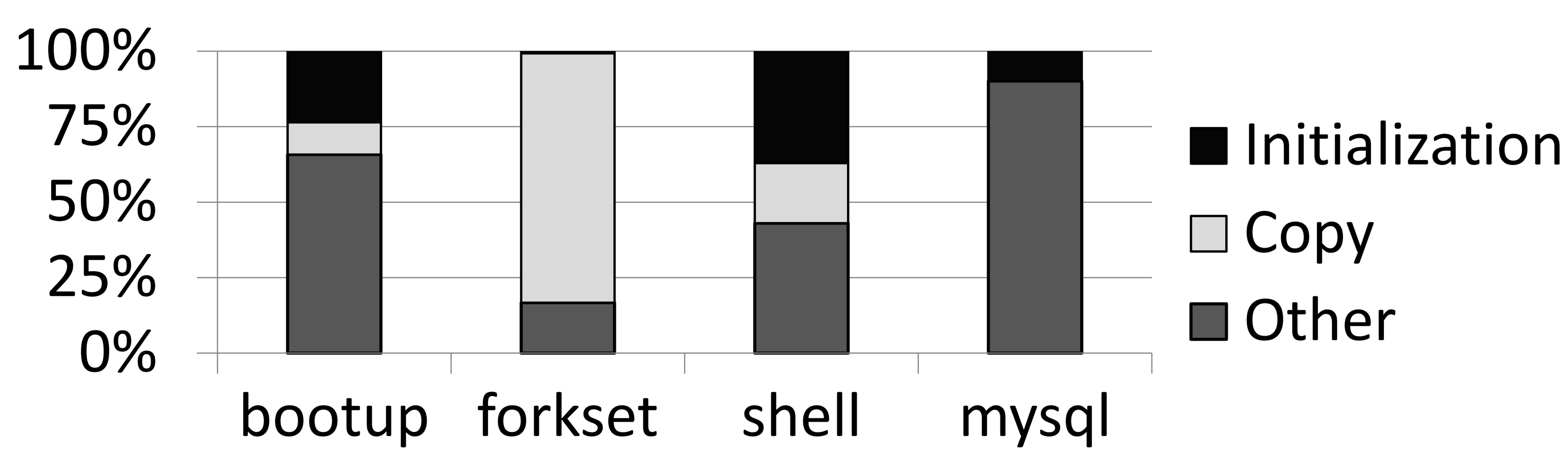


As a result, existing systems suffer from:

1. Large memory latency
2. Wasted memory bandwidth
3. Wasted memory power

3. Cases for Copying & Initialization

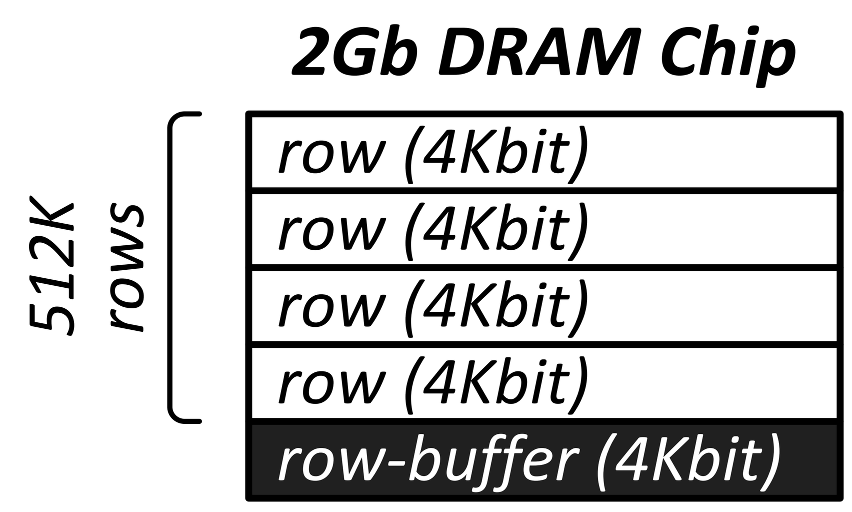
- Prevalence of copying/initialization



- Usages of copying/initialization
  - System/application startup
  - Secure memory deallocation
  - Process cloning & copy-on-write
  - Memory checkpointing
  - Graphics processing

4. Key Observation

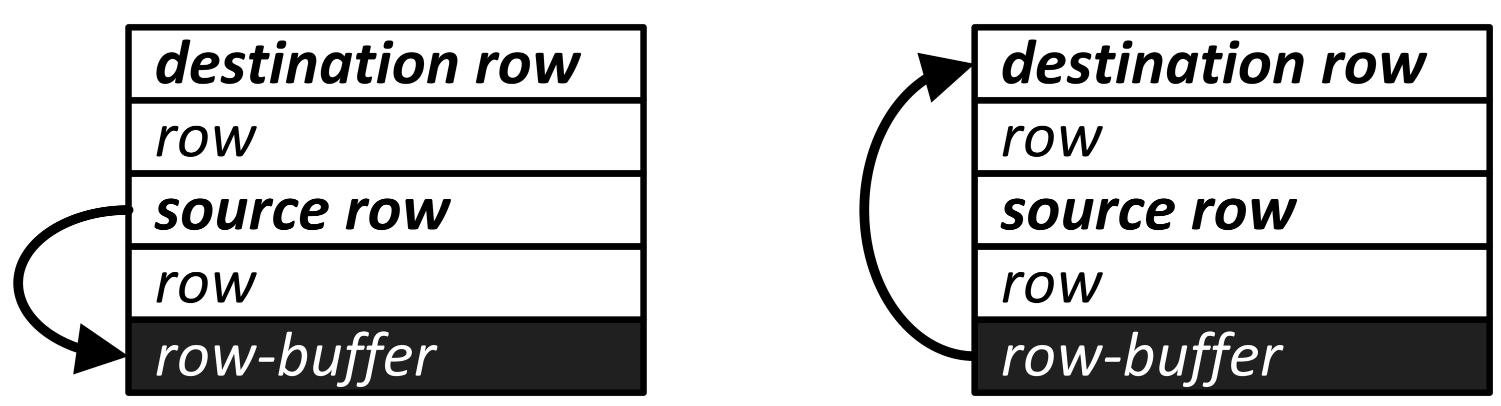
- A DRAM chip consists of ***rows*** and a ***row-buffer***
  - To access any data from a row, the *entire* row must first be loaded into the row-buffer



Key Observation: A DRAM chip *already* supports bulk data-transfer between any of its rows and the row-buffer – i.e., *the entire row is loaded into the row-buffer all at once*

5. Mechanism: RowClone

- **RowClone:** A two-step mechanism of copying one row to another row – *all within a DRAM chip*
  - At a low cost of only **0.026%** increase in die-size



Step 1. Load source row into row-buffer

Step 2. Load row-buffer into destination row

- Raw benefits when copying a single 4KB page:
  - **11.5x** reduction in latency
  - **74.4x** reduction in energy
- Performance improvements:
  - **14%** increase in average IPC for 8 benchmarks
  - **27%** increase in system throughput for 8-core workloads (copy/initialize-intensive benchmarks)

Intel Science & Technology Center for Cloud Computing

Carnegie Mellon University · Georgia Tech · intel · Princeton University · UC Berkeley