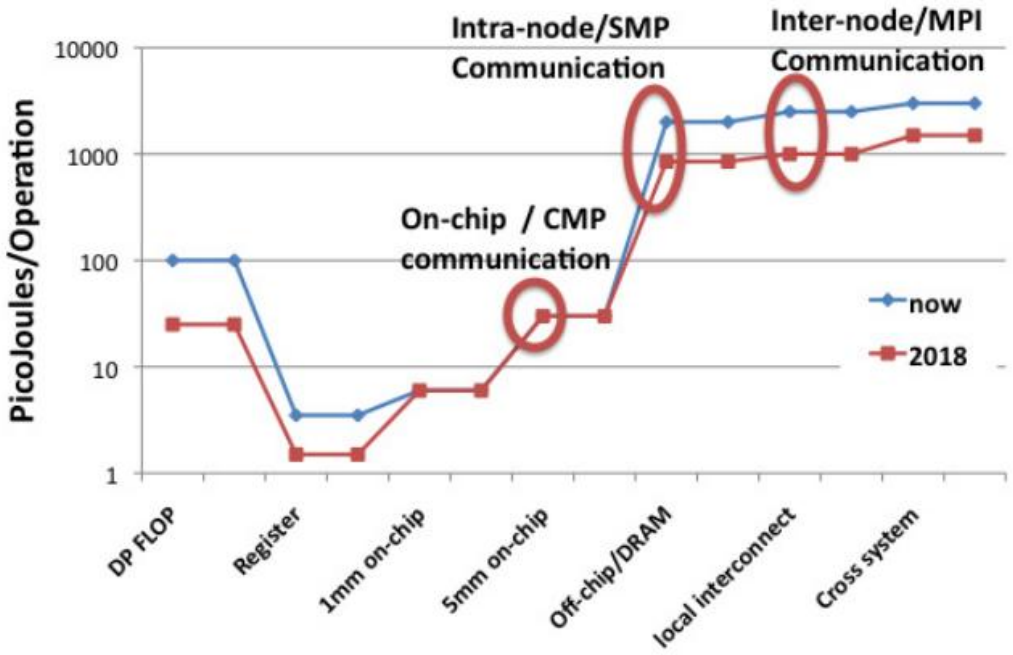
Analysis of Energy and Performance of Code Transformations for PGAS–based Data Access Patterns

Siddhartha Jana (UH), Joseph Schuchart (TU–D),
Barbara Chapman (UH)

PGAS, 2014
Oct 10, 2014

Why Talk About Energy?



Engineering FLOPs is not a design constraint – data movement presents the most daunting engineering and computer architecture challenge

Source: "*Exascale Computing Technology Challenges*", John Shalf, Sudip Dosanjh, and John Morrison

Outline

- Motivation
- Energy Cost Factors Across the H/W - S/W Stack
- Design of Data Access Patterns
- Access Pattern Transformations
- Analysis of Empirical Results
- Conclusions and Future Work

Motivation

- Energy savings is clearly a concern

  • DVFS techniques to achieve energy savings
  - Vishnu et al. (2013)

  • Energy study of point-to-point and collective operations
  - Venkatesh et al., Jana et al. (2013~14)

- What kernel characteristics should we be looking into?

- What layer within the software stack should we target?

Factors impacting energy consumption

Impact across the hardware-software stack

Intra-node Constraints	**Inter-node Constraints**
e.g. Cache sizes, set-associativity, cache-coherency protocol memory bandwidth, Hyperthreading, page-replacement	e.g. router-switch, organization, network topology, reliability, latency, peak-bandwidth

Factors impacting energy consumption

Impact across the hardware-software stack

Implementation Details e.g. Polling, registration of memory, reliability, reusability of memory, caching, memory management, fault-tolerance	**Flow / Congestion control** e.g. routing protocols, deadlock handling, load-balancing, quality-of-service
Intra-node Constraints e.g. Cache sizes, set-associativity, cache-coherency protocol memory bandwidth, Hyperthreading, page-replacement	**Inter-node Constraints** e.g. router-switch, organization, network topology, reliability, latency, peak-bandwidth

Factors impacting energy consumption

Impact across the hardware-software stack

Choice of Transport Layer e.g. TCP, OpenFabrics, shared memory	**Communication Protocols** e.g. Message passing (Eager, Rendezvous) or Direct access
Implementation Details e.g. Polling, registration of memory, reliability, reusability of memory, caching, memory management, fault-tolerance	**Flow / Congestion control** e.g. routing protocols, deadlock handling, load-balancing, quality-of-service
Intra-node Constraints e.g. Cache sizes, set-associativity, cache-coherency protocol memory bandwidth, Hyperthreading, page-replacement	**Inter-node Constraints** e.g. router-switch, organization, network topology, reliability, latency, peak-bandwidth

Factors impacting energy consumption

Impact across the hardware-software stack

<table>
<tr><td colspan="2">Communication Kernel Characteristics
e.g. total size of the data-payload transferred,
the number of calls initiated to service the transfers</td></tr>
<tr><td>Choice of Transport Layer
e.g. TCP, OpenFabrics,
shared memory</td><td>Communication Protocols
e.g. Message passing
(Eager, Rendezvous)
or Direct access</td></tr>
<tr><td>Implementation Details
e.g. Polling, registration of memory,
reliability, reusability of memory,
caching, memory management,
fault-tolerance</td><td>Flow / Congestion control
e.g. routing protocols,
deadlock handling,
load-balancing, quality-of-service</td></tr>
<tr><td>Intra-node Constraints
e.g. Cache sizes, set-associativity,
cache-coherency protocol
memory bandwidth,
Hyperthreading, page-replacement</td><td>Inter-node Constraints
e.g. router-switch, organization,
network topology,
reliability, latency,
peak-bandwidth</td></tr>
</table>

Factors impacting energy consumption

Impact across the hardware-software stack

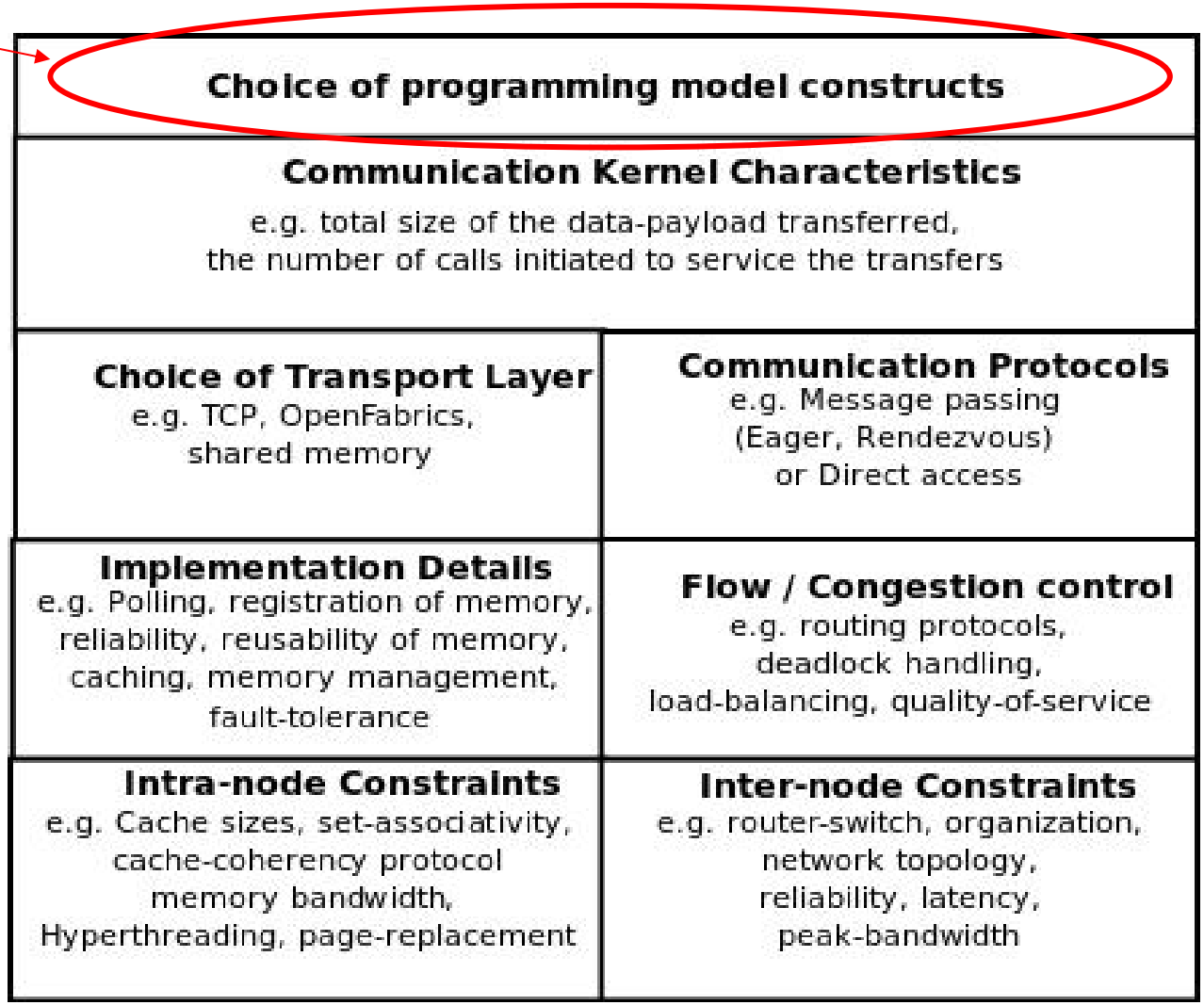
<table>
<tr><td colspan="2">Choice of programming model constructs</td></tr>
<tr><td colspan="2">Communication Kernel Characteristics
e.g. total size of the data-payload transferred,
the number of calls initiated to service the transfers</td></tr>
<tr><td>Choice of Transport Layer
e.g. TCP, OpenFabrics,
shared memory</td><td>Communication Protocols
e.g. Message passing
(Eager, Rendezvous)
or Direct access</td></tr>
<tr><td>Implementation Details
e.g. Polling, registration of memory,
reliability, reusability of memory,
caching, memory management,
fault-tolerance</td><td>Flow / Congestion control
e.g. routing protocols,
deadlock handling,
load-balancing, quality-of-service</td></tr>
<tr><td>Intra-node Constraints
e.g. Cache sizes, set-associativity,
cache-coherency protocol
memory bandwidth,
Hyperthreading, page-replacement</td><td>Inter-node Constraints
e.g. router-switch, organization,
network topology,
reliability, latency,
peak-bandwidth</td></tr>
</table>

Factors impacting energy consumption

Impact across the hardware-software stack

Scope of this work

Past work



Factors impacting energy consumption

Impact across the hardware-software stack

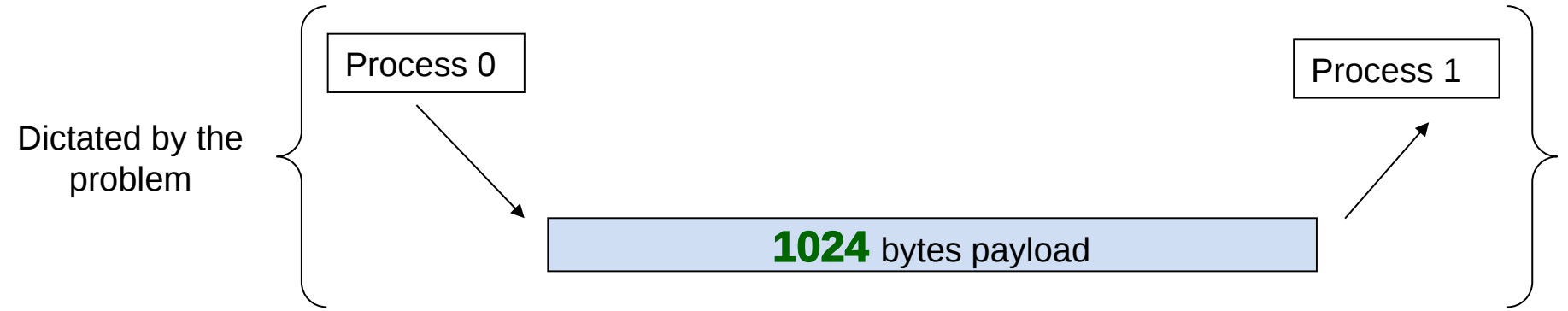
Choice of programming model constructs

Characteristics of a communication kernel

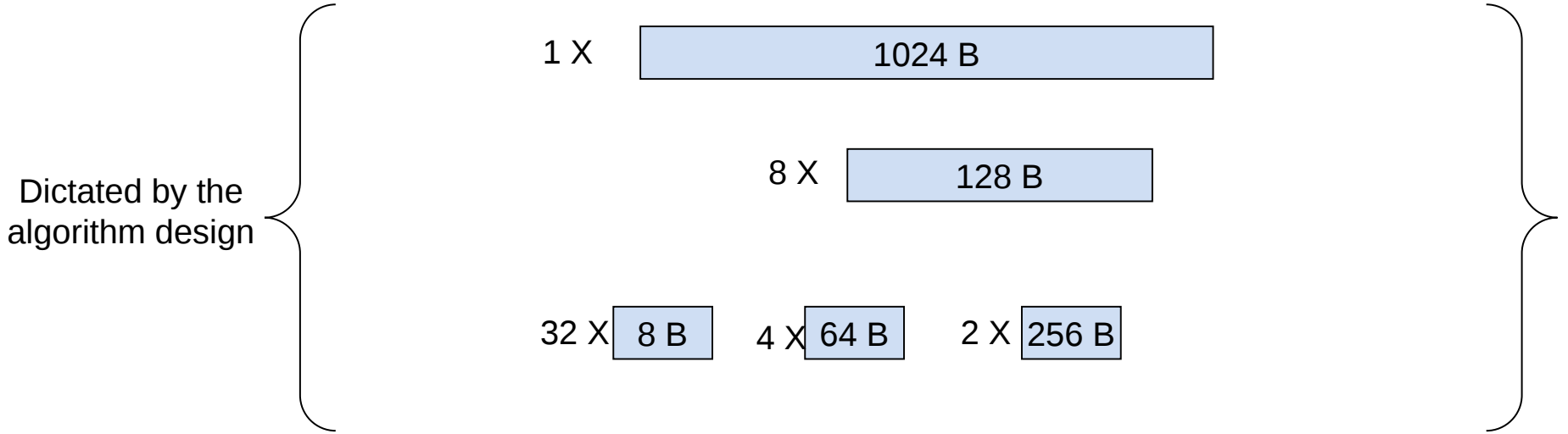
Characteritics of individual RDMA PUTs

Characteristics of a Communication Kernel

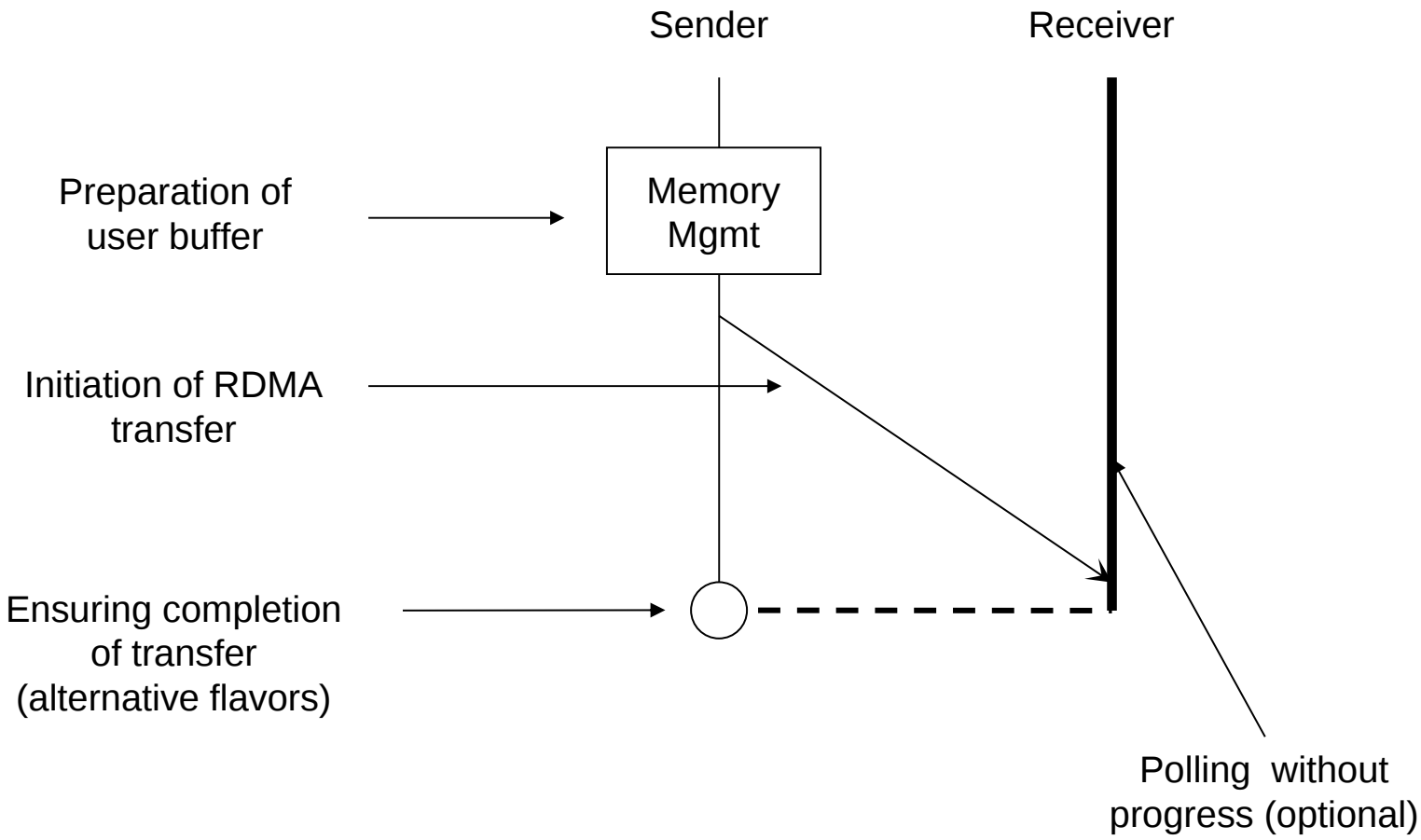
Consider the following case:



This can be transferred as:

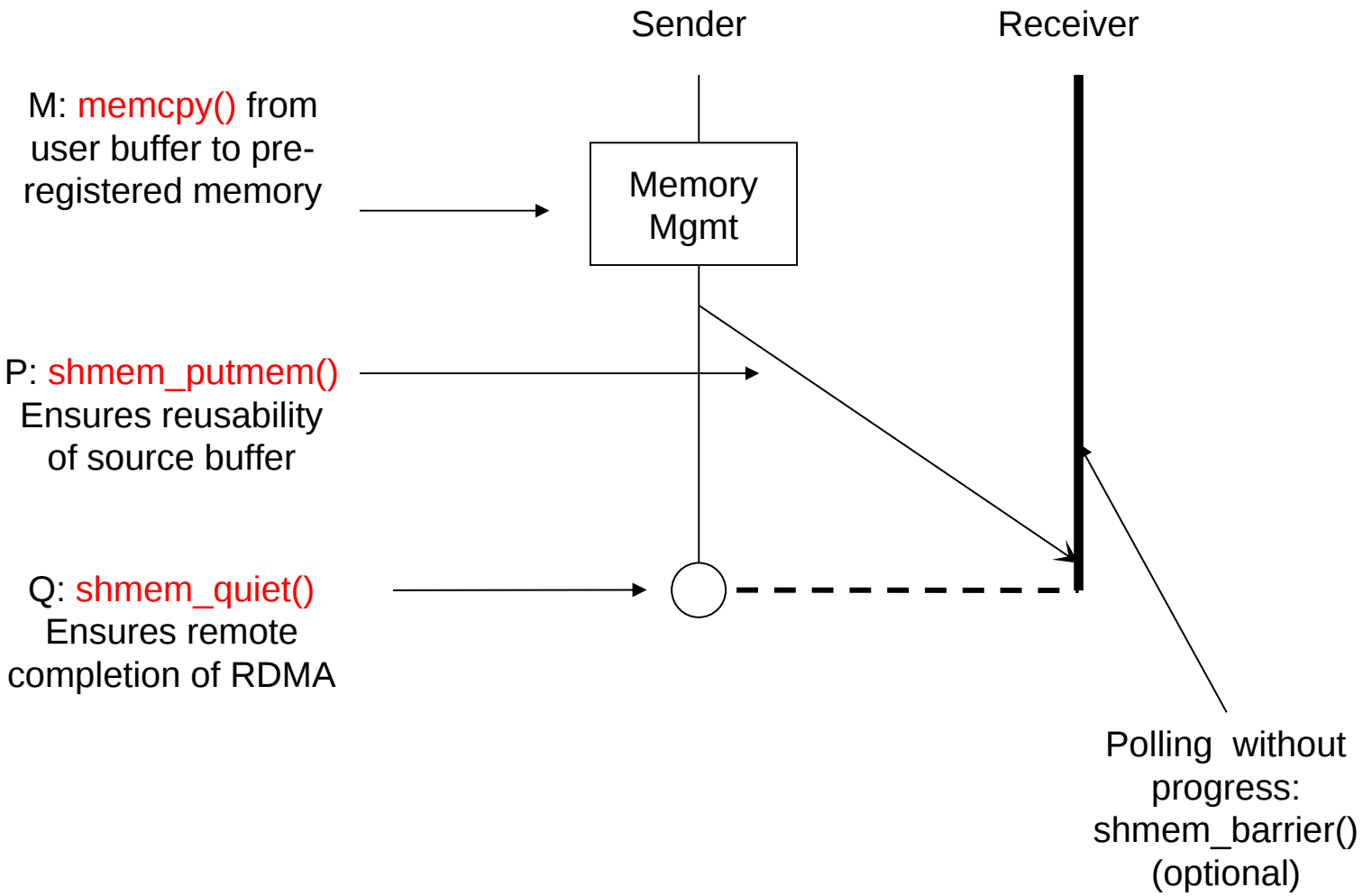


Costs Associated With RDMA Write Operations

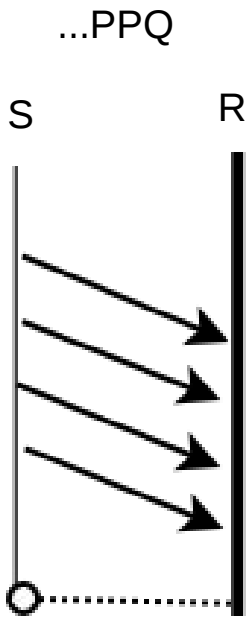


Costs Associated With RDMA Write Operations

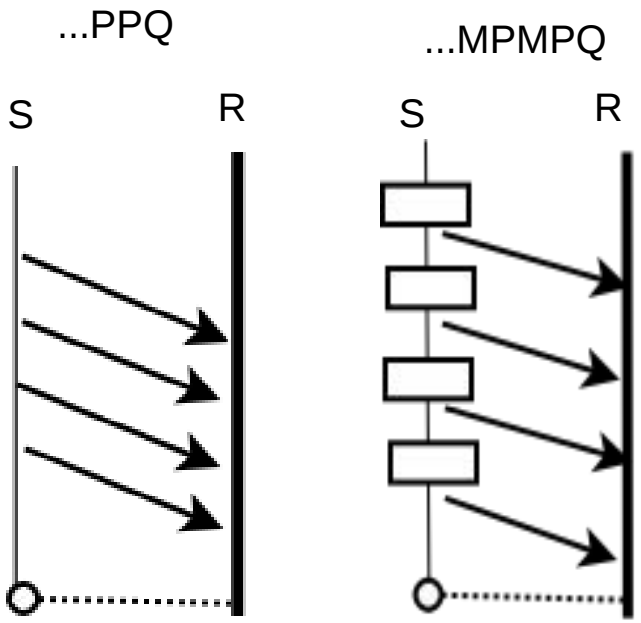
In terms of OpenSHMEM



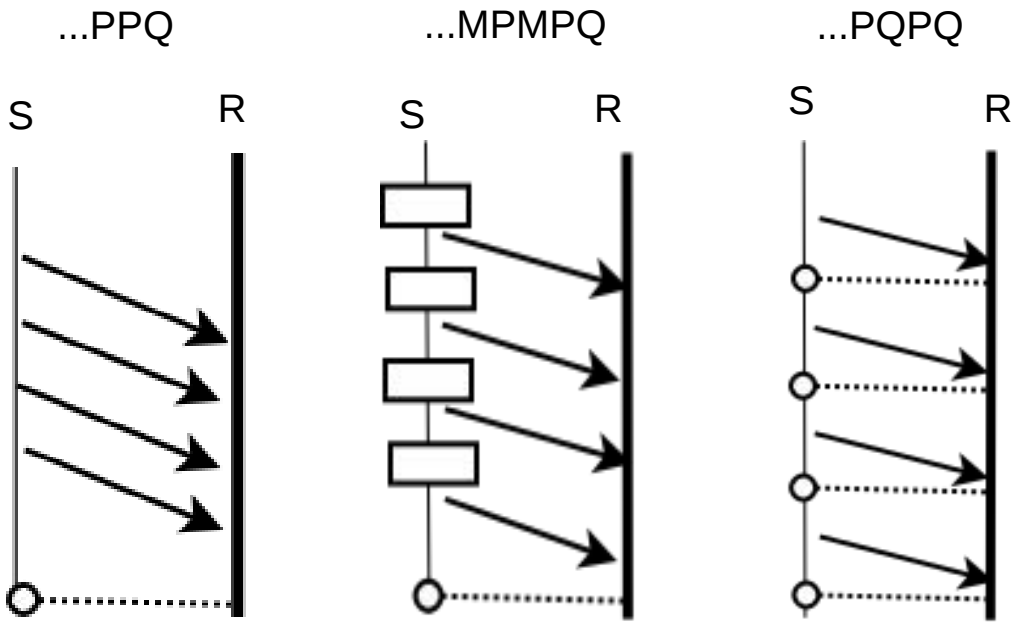
Analyzing Data Access Patterns



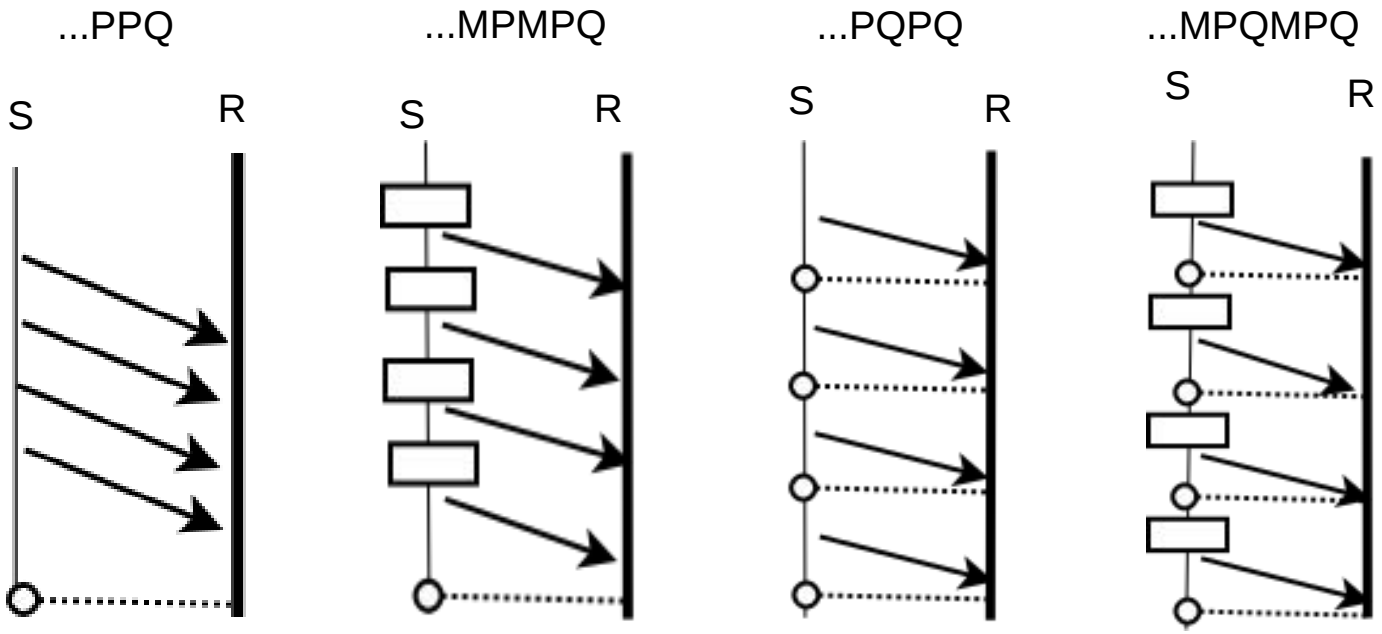
Analyzing Data Access Patterns



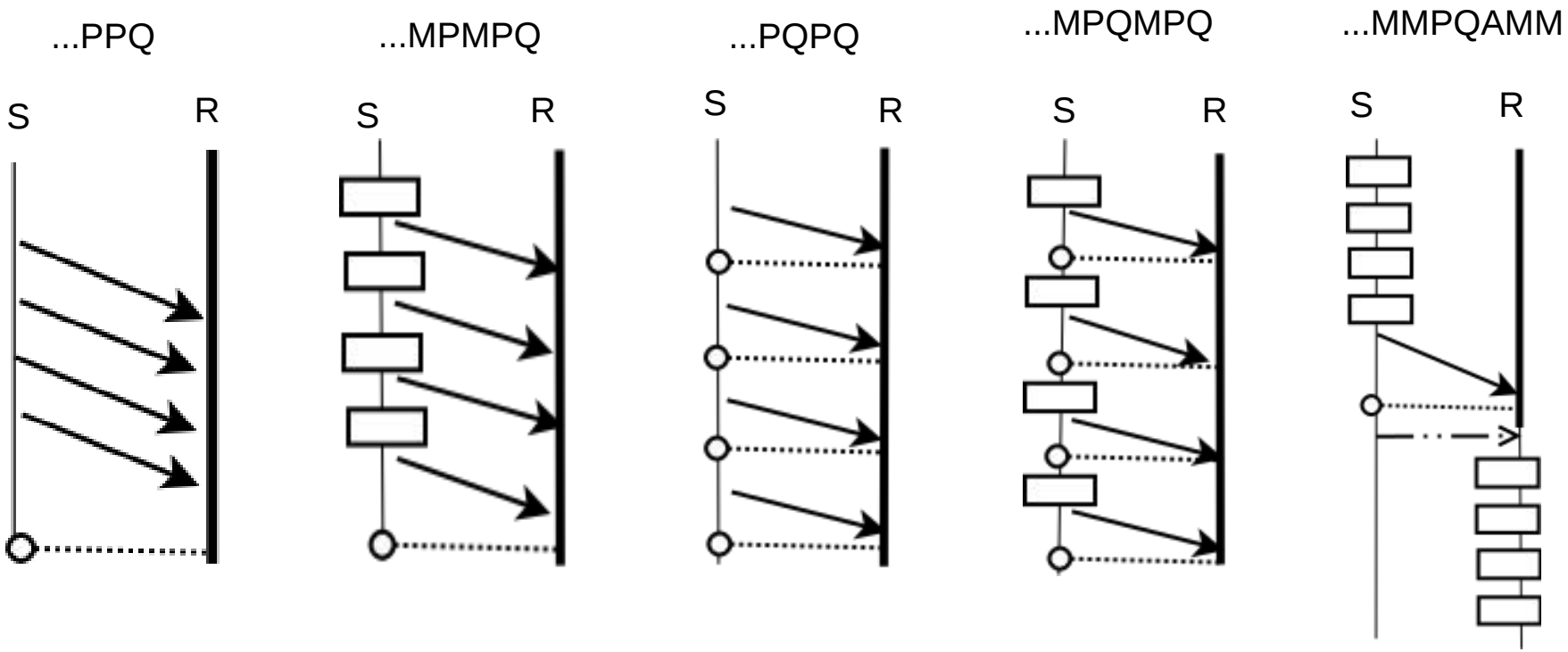
Analyzing Data Access Patterns



Analyzing Data Access Patterns

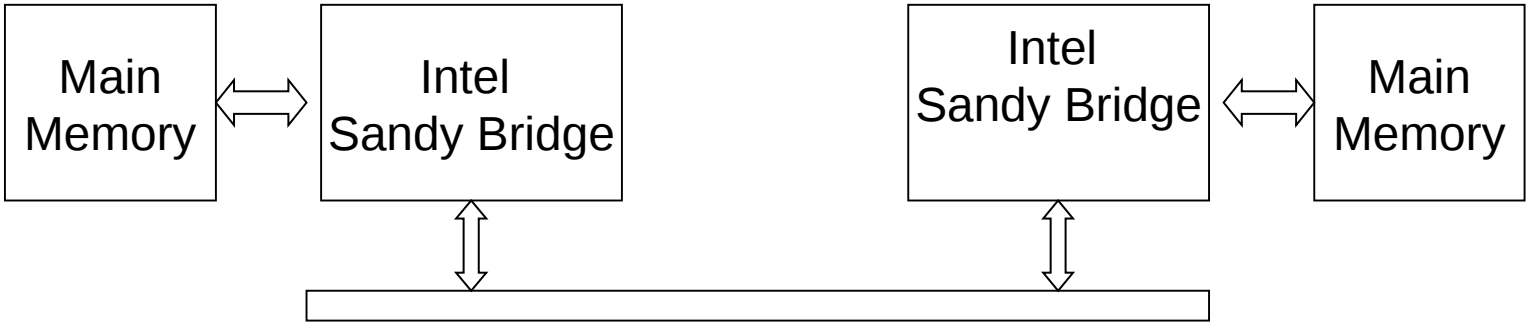


Analyzing Data Access Patterns



Experimental Setup

- CPU: Intel Sandy Bridge E5-2690
- NIC: Mellanox MT27500 : Connect-X
- One process (OpenSHMEM PE) per node
- Mellanox Scalable SHMEM



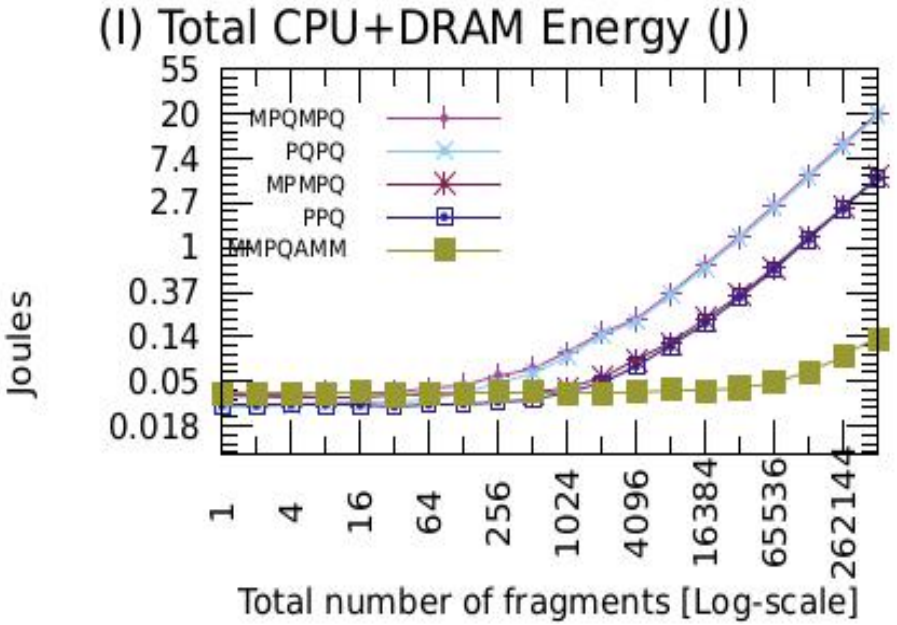
Infiniband (Mellanox)

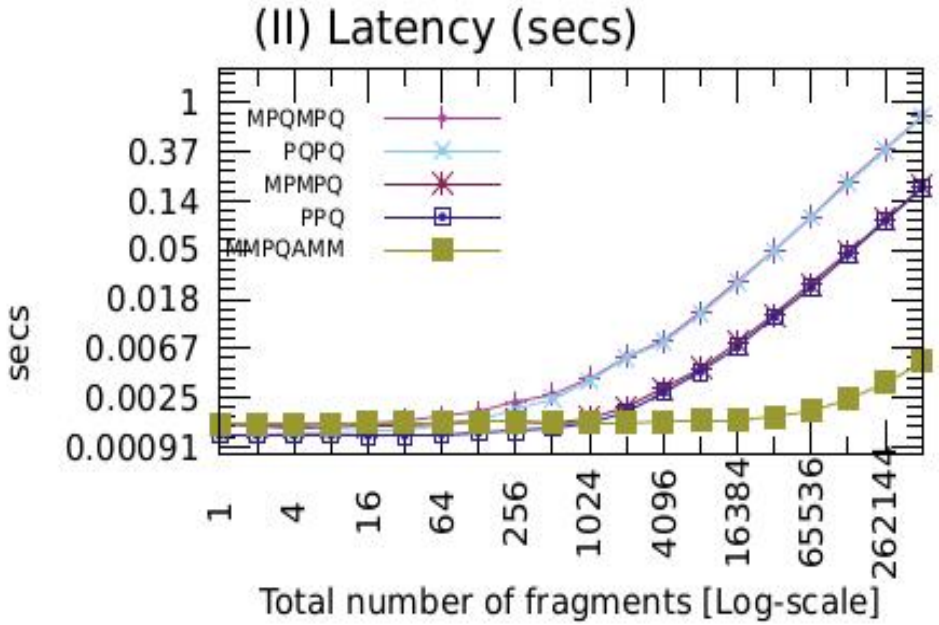
- Use of Voltage Regulators + FPGA
- Power measured at various levels of granularity
  - CPU, DRAM, Blade
  - 1KHz granularity

Peformance Characteristics of Data Access Patterns

Lower is better

For medium ~ bulk transfers:
MPQMPQ ~ PQPQ > MPMPQ ~ PPQ > MMPQAMM
Blocking versions > Non-blocking semantics > Aggregation



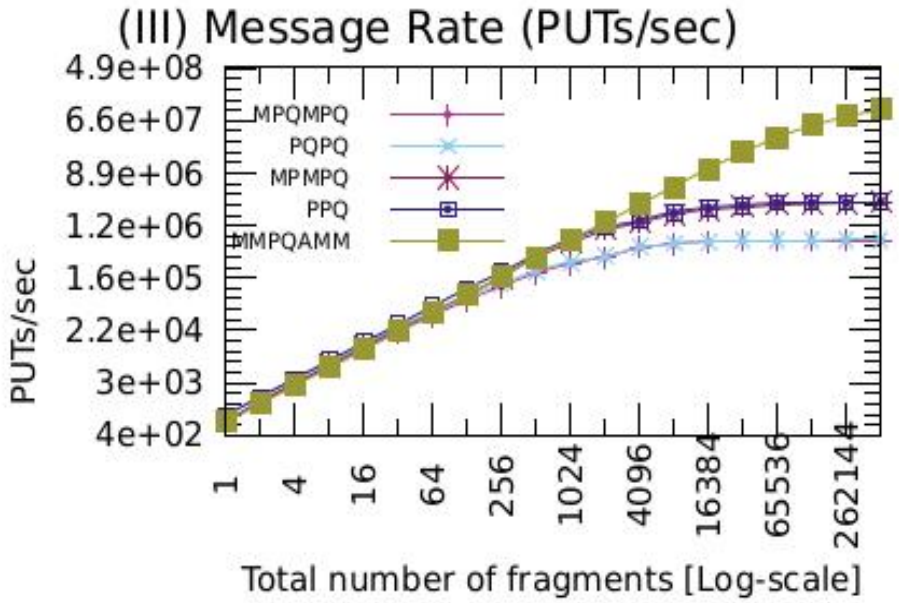


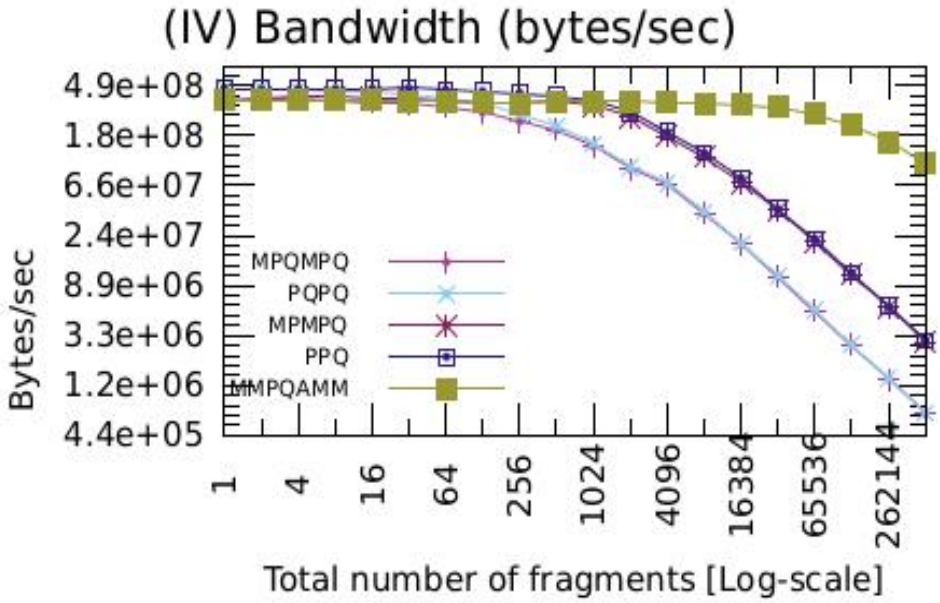
Total size of data payload = 0.5MB
Number of fragments (N) = Number of discrete user buffers
Size of a data buffer = 0.5MB/ N
Measurements in log-scale

Peformance Characteristics of Data Access Patterns

Higher is better

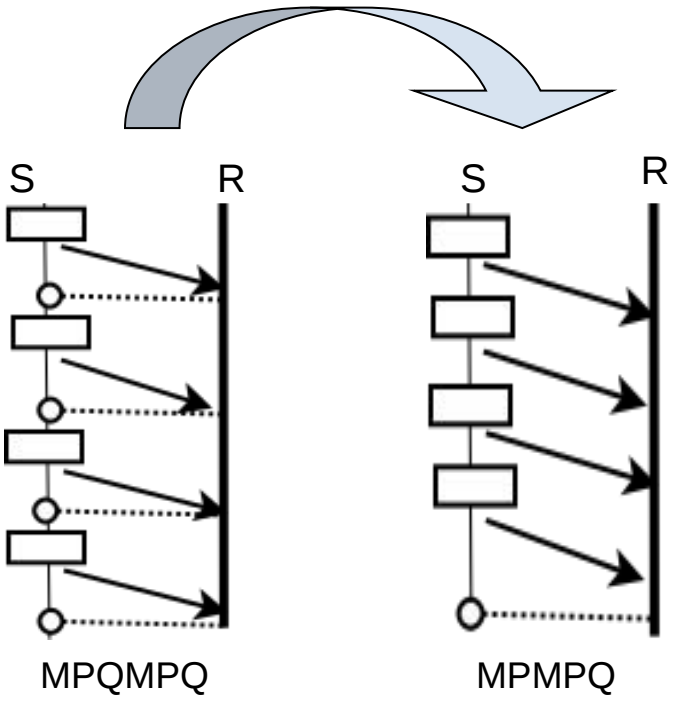
For medium ~ bulk transfers:
MPQMPQ ~ PQPQ < MPMPQ ~ PPQ < MMPQAMM
Blocking versions < Non-blocking semantics < Aggregation



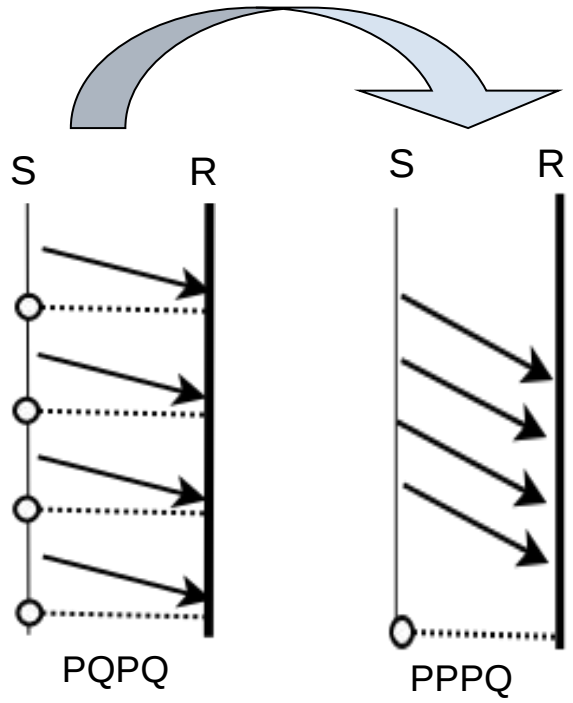


Total size of data payload = 0.5MB
Number of fragments (N) = Number of discrete user buffers
Size of a data buffer = 0.5MB/ N
Measurements in log-scale

Cost Savings Using Non-blocking operations



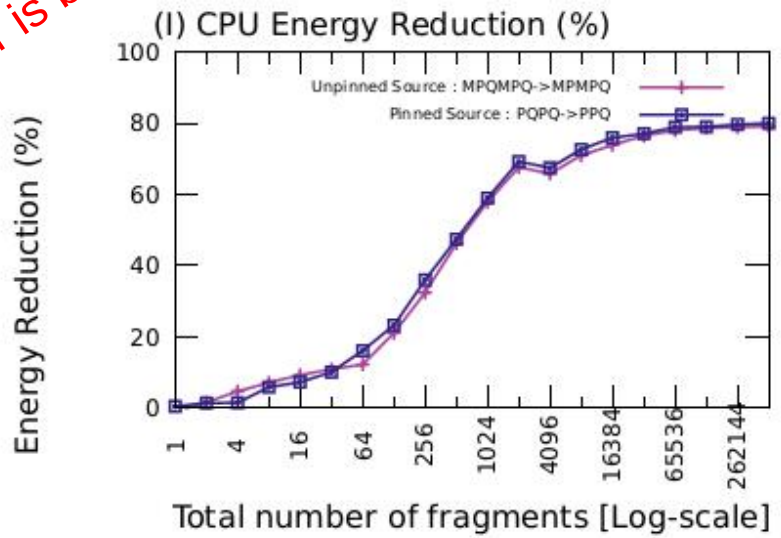
Case1: With unregistered buffers

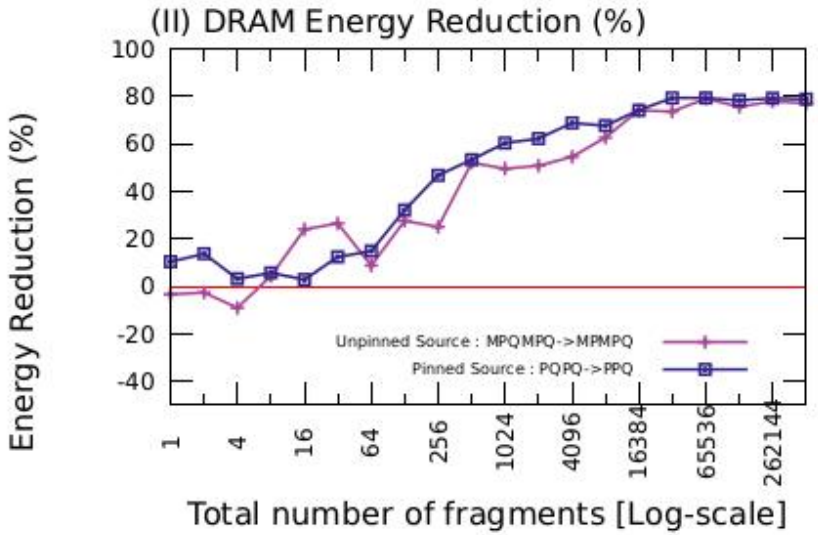


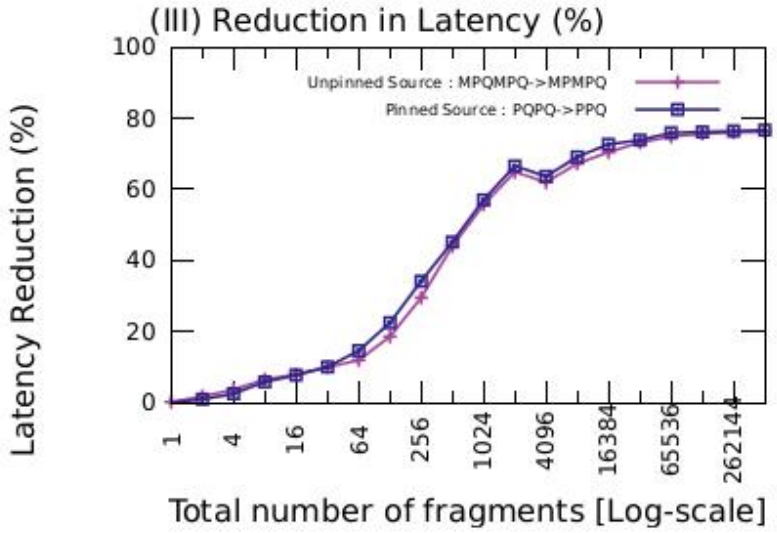
Case2: With registered buffers

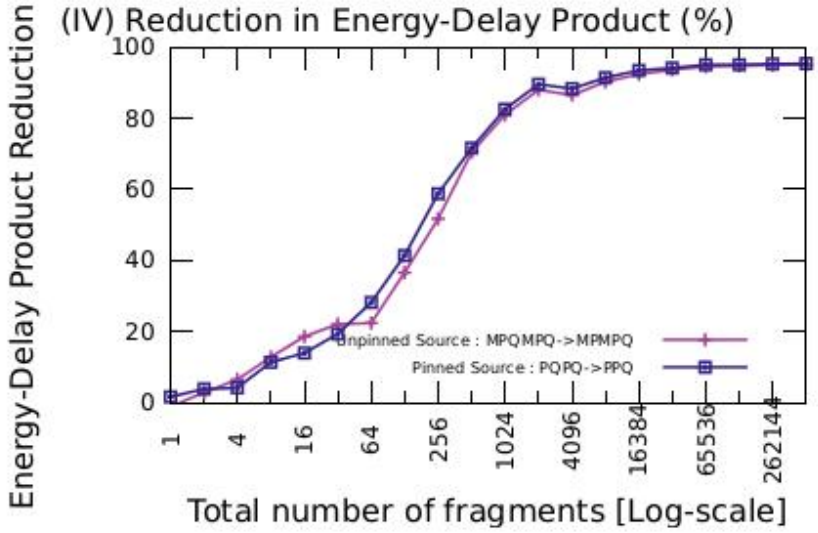
Impact of Using Non-blocking operations

Higher is better

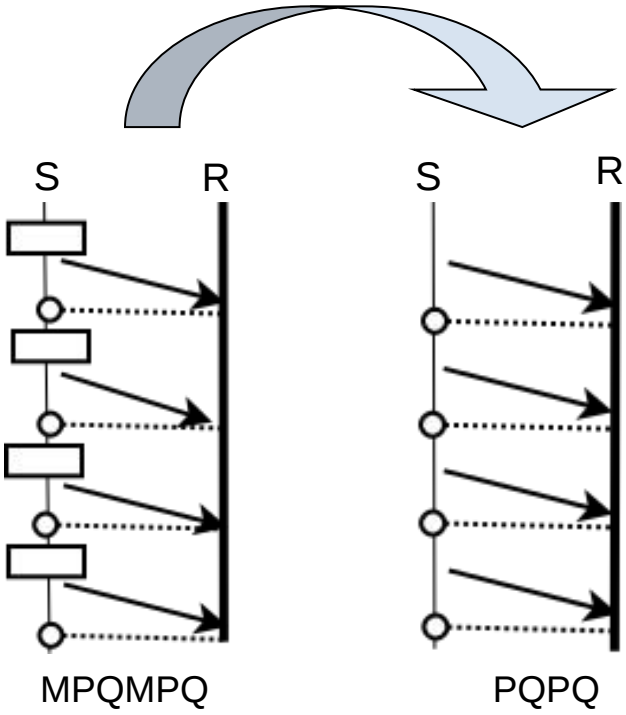




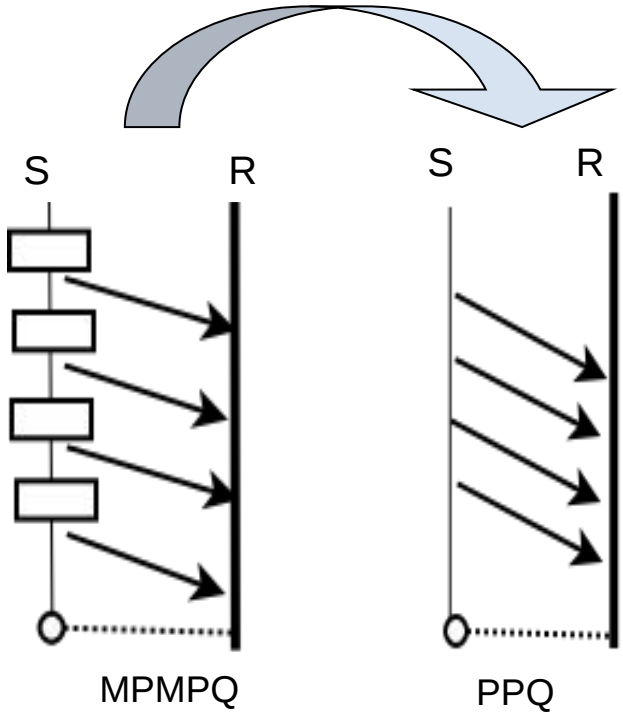




Costs Savings Using NIC-Registered Buffers



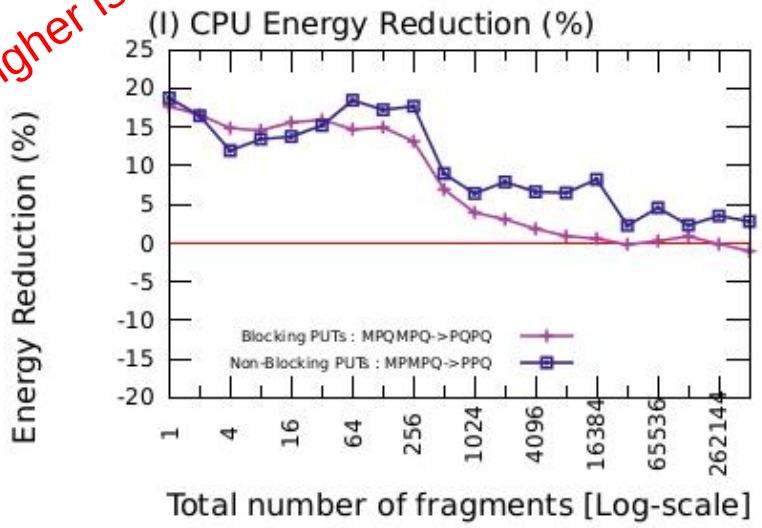
Case1: With blocking PUTs

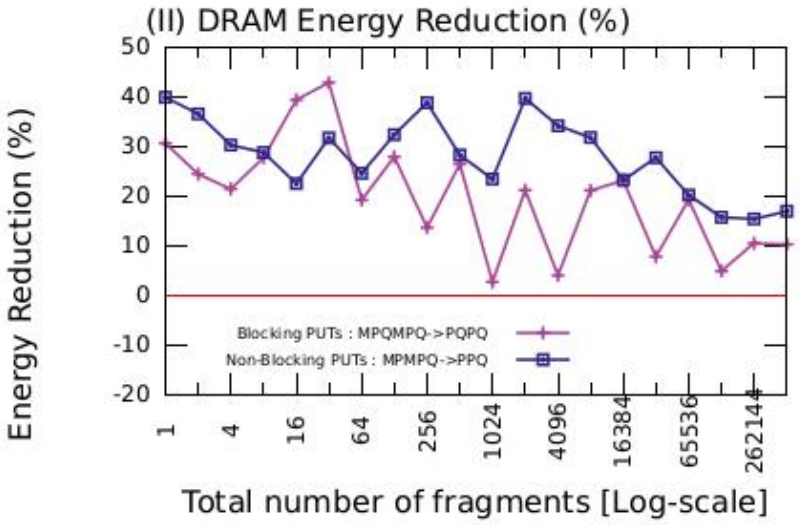


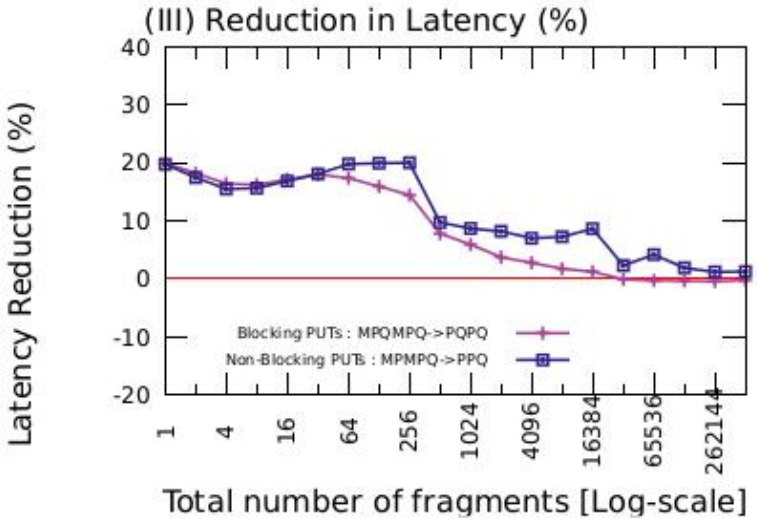
Case2: With non-blocking PUTs

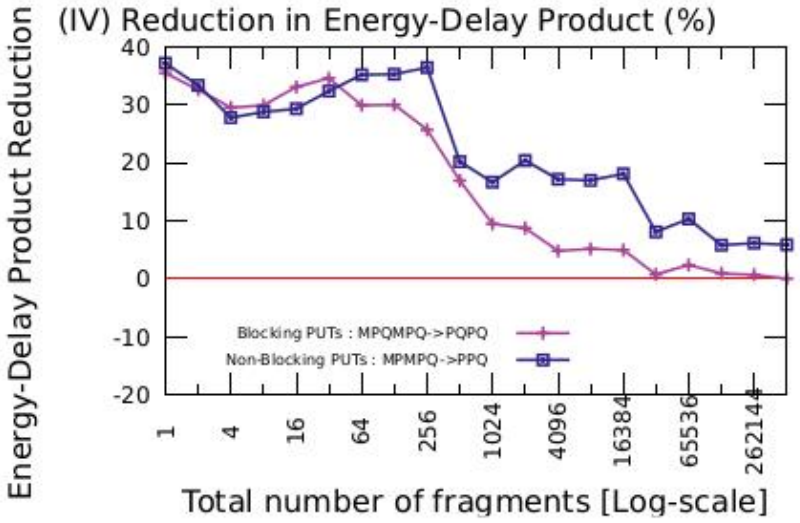
The Impact of Using NIC-Registered Buffers

Higher is better

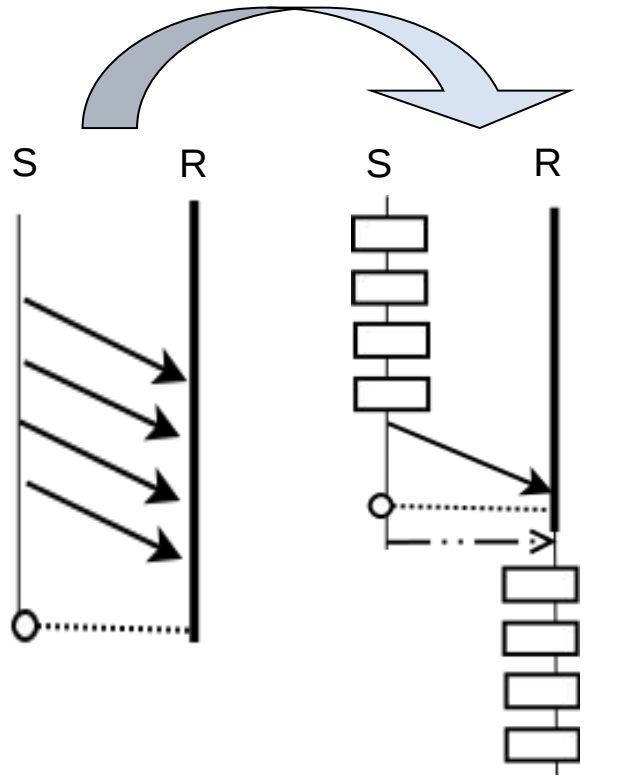








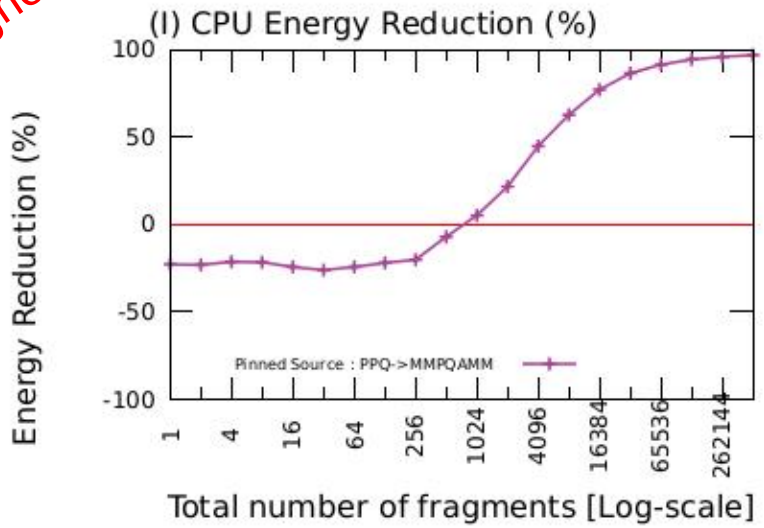
Costs Savings Using Aggregated Source Buffers

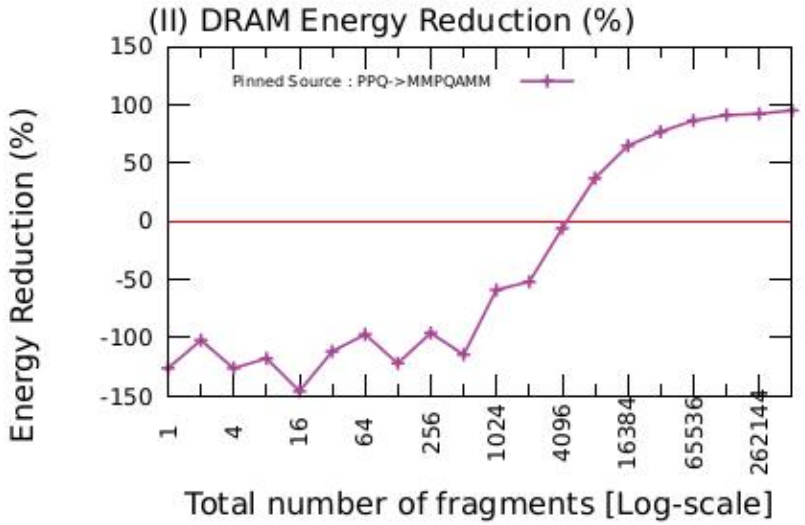


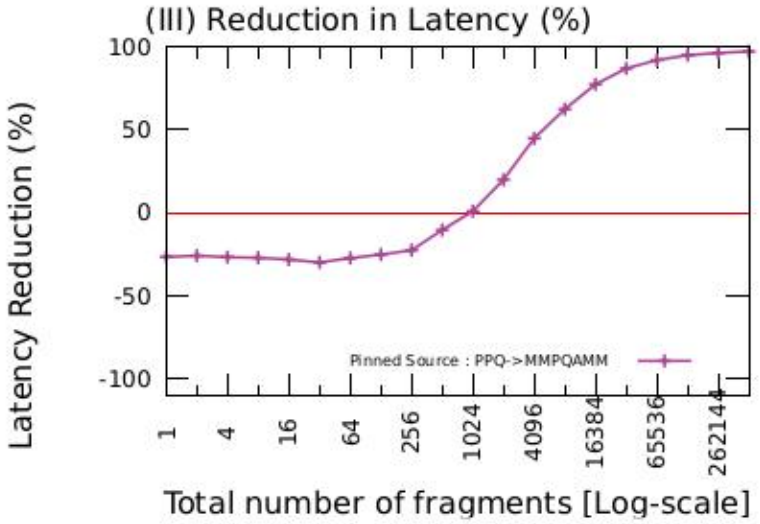
PPQ MMPQAMM

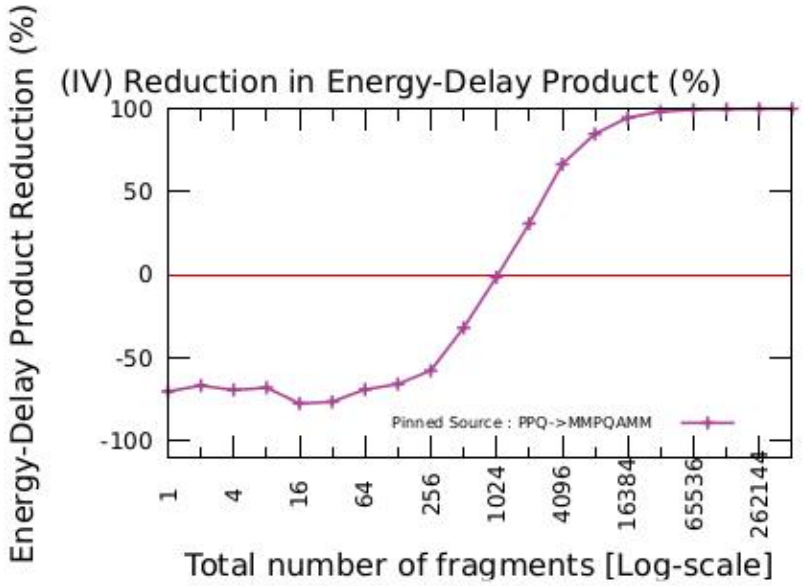
Higher is better

The Impact of Using Aggregated Source Buffers









Conclusions and Future Work

- Energy-based metrics similar to latency based results

- Analysis of different data access patterns
    - Aggregation
    - Blocking
    - Buffer registration

- Potential cost savings in converting in transformation of these patterns

Analysis of Energy and Performance of Code Transformations for PGAS-based Data Access Patterns

Speaker: Siddhartha Jana (UH)
Contact: sidjana@cs.uh.edu
University of Houston