Analyzing the Energy and Power Consumption of Remote Memory Accesses in the OpenSHMEM Model

Siddhartha Jana[1], Oscar Hernandez[2], Stephen Poole[2], Chung-Hsing Hsu[2], Barbara Chapman[1]

[1]University of Houston, {sidjana, chapman}@cs.uh.edu

[2] Oak Ridge National Laboratory, {oscar, spoole, hsuc}@ornl.gov

First OpenSHMEM Workshop
Experiences, Implementations, and Tools

March 6, 2014

How *Green* are we ?

P = 15 Watts

5 secs 75 Joules

10 secs 150 Joules

How *Green* are we ?

P = 15 Watts

5 secs 75 Joules

10 secs 150 Joules

P = 600 milliWatts

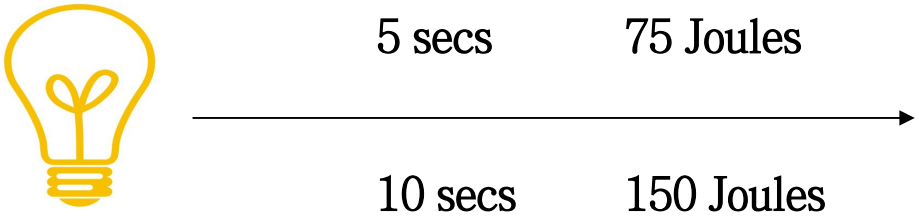
add R1 R2 1 cycles X Joules

mul R1 R2 10 cycles Y Joules

Y ≥ X

How *Green* are we ?

P = 15 Watts



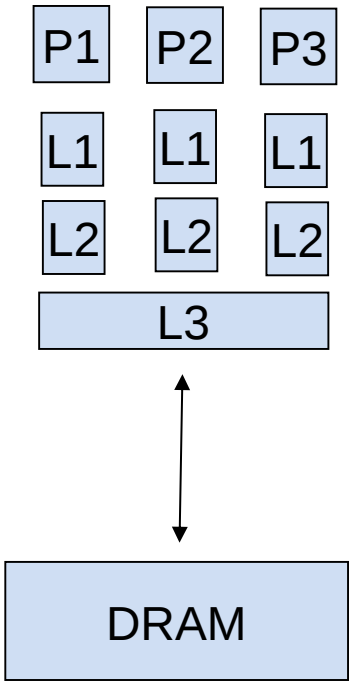
- Fastest supercomputer @ 50 PFLOPS
  [according to the TOP500 list]
- DOE Exascale limit: 20 MW
- 1 Million USD per Megawatt per year
- At current consumption : 20Million USD per year

Outline

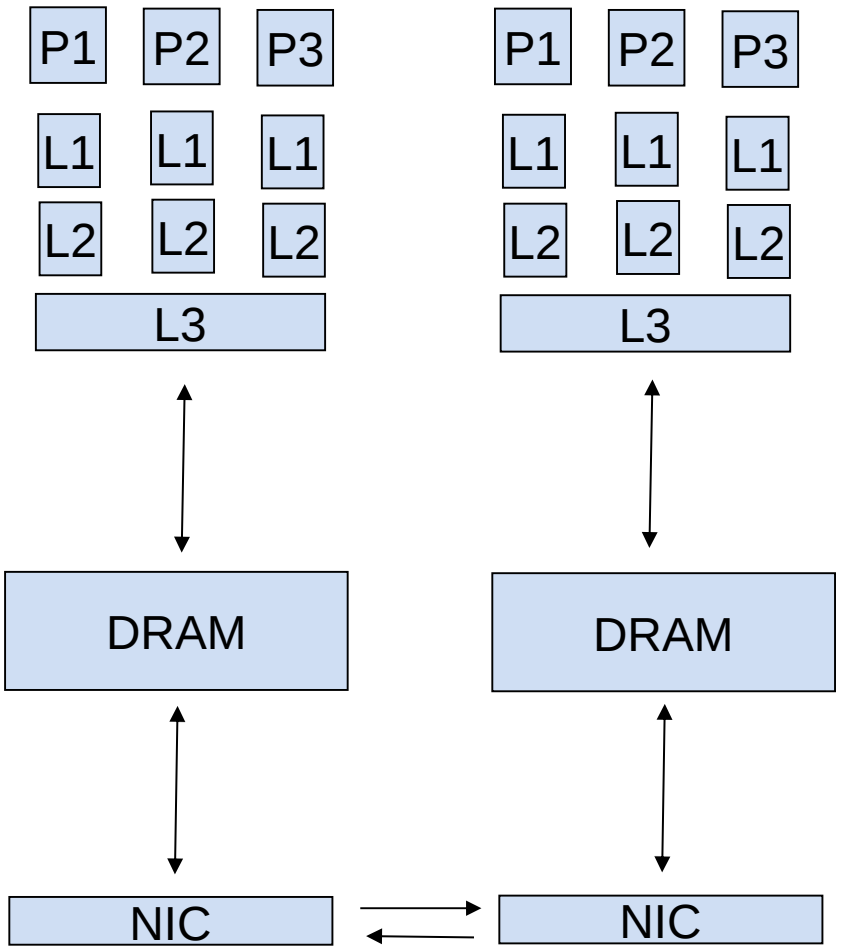
- Motivation
  - Impact on power and energy efficiency
- Energy and power efficiency - software stack
  - Synchronizations
  - Remote data transfers
- Impact of remote data transfers
  - Experimental Setup
  - Observations
- Conclusions and Future Work

Impact on power and energy efficiency

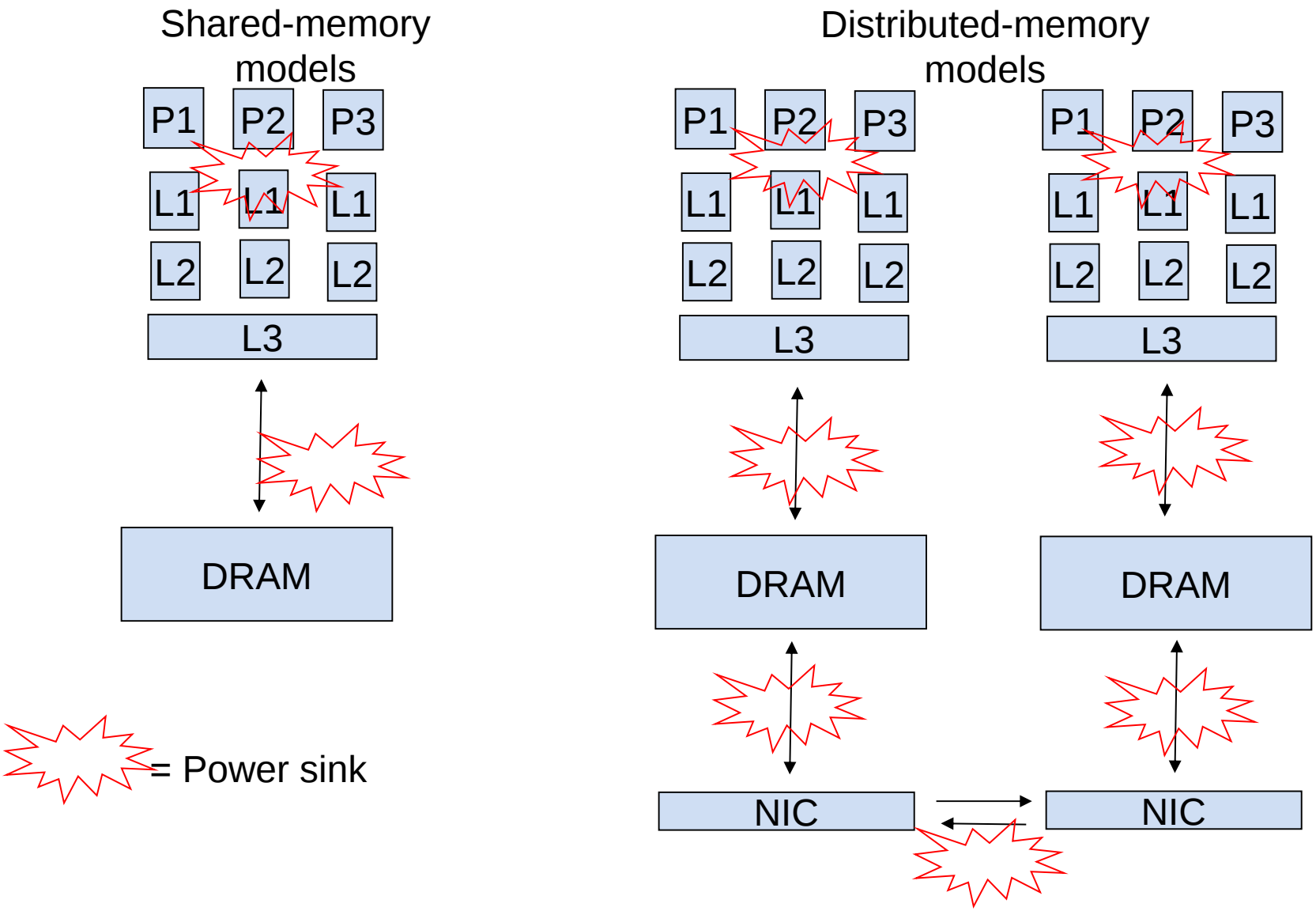
Shared-memory models



Distributed-memory models



Impact on power and energy efficiency



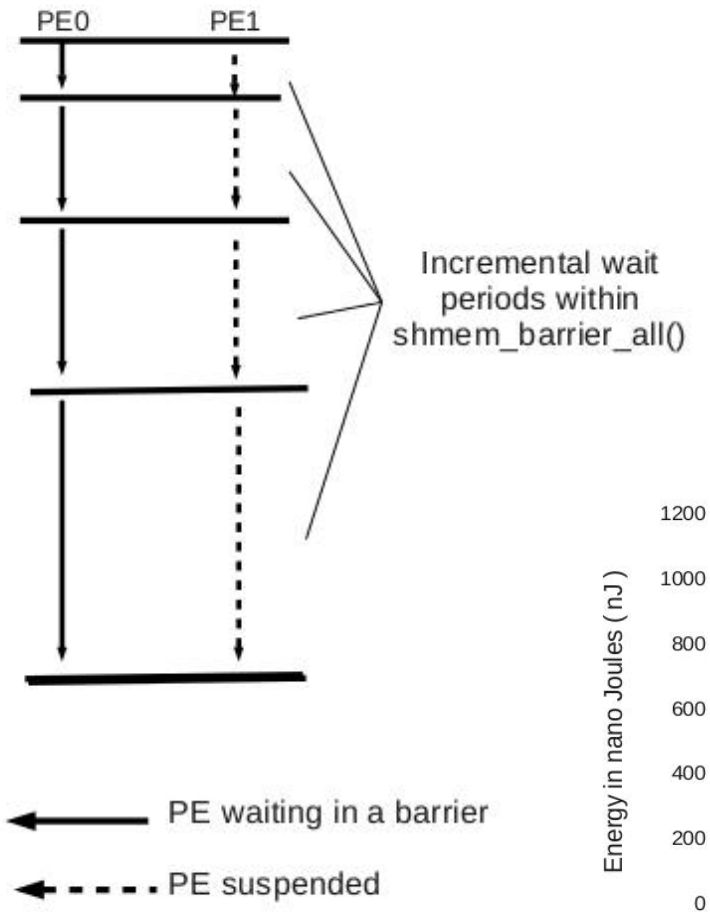
Impact on power and energy efficiency

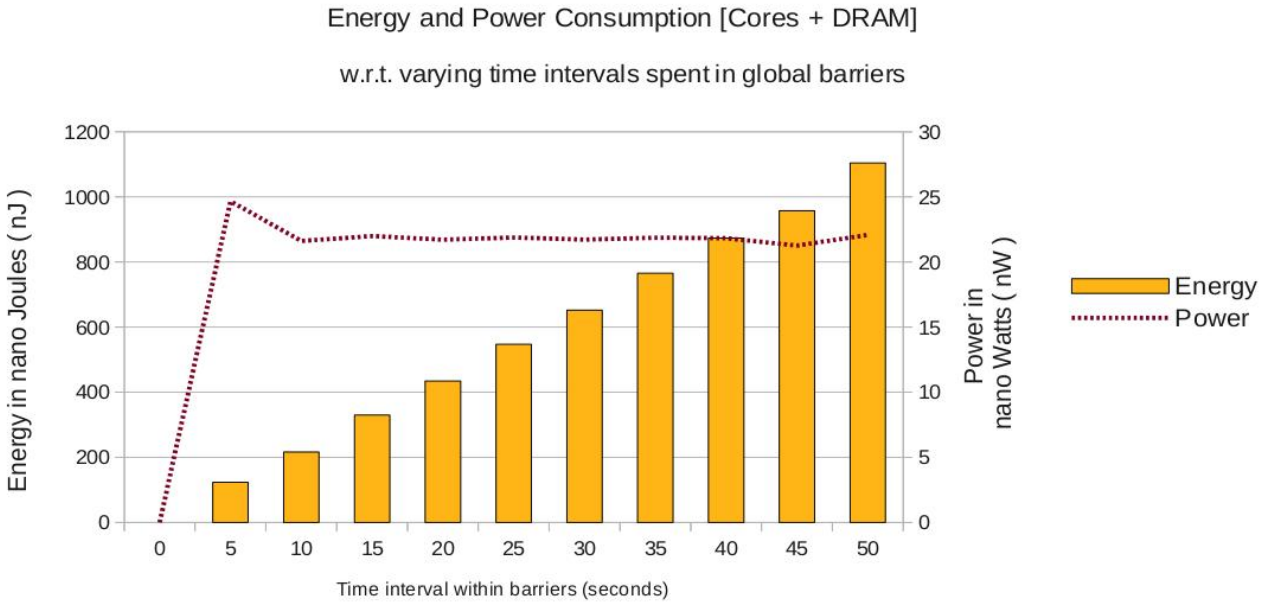
Impact across the hardware-software stack

<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 protocol
e.g. Message passing
(Eager, Rendezvous)
or Direct access</td></tr>
<tr><td>Implementation details
Polling, registration of memory,
reliability, reusability of memory,
caching, memory management,
fault-tolerance</td><td>Flow/congestion control,
routing protocols,
deadlock handling,
load-balancing, quality-of-service</td></tr>
<tr><td>Cache sizes, set-associativity,
cache-coherency protocol
memory bandwidth,
Hyperthreading, page-replacement</td><td>router-switch, organization,
network topology,
reliability, latency,
peak-bandwidth</td></tr>
</table>

Energy and power efficiency - software stack

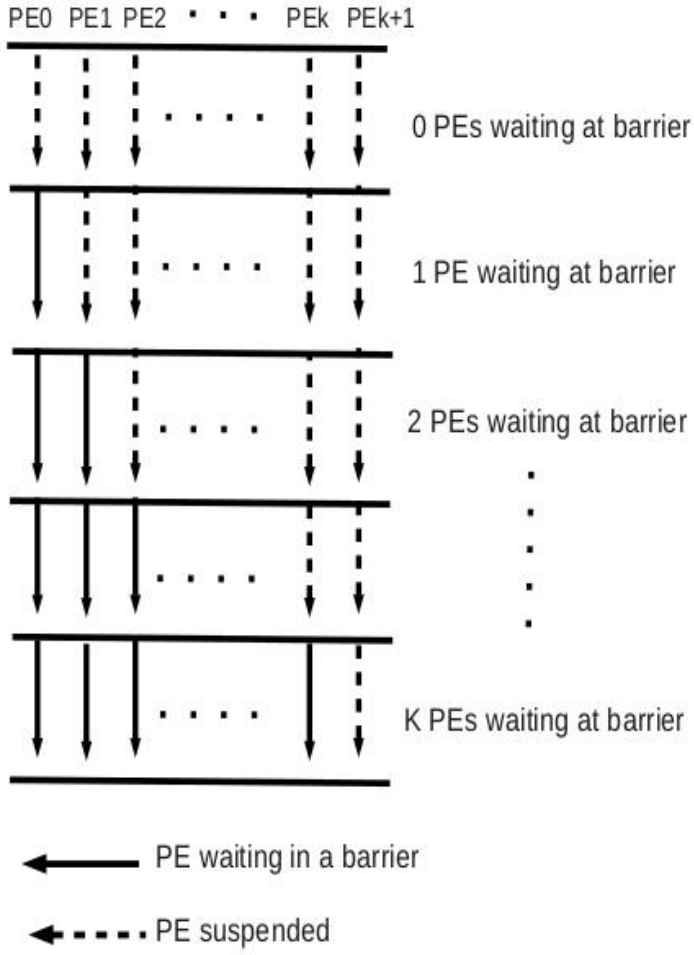
Synchronizations

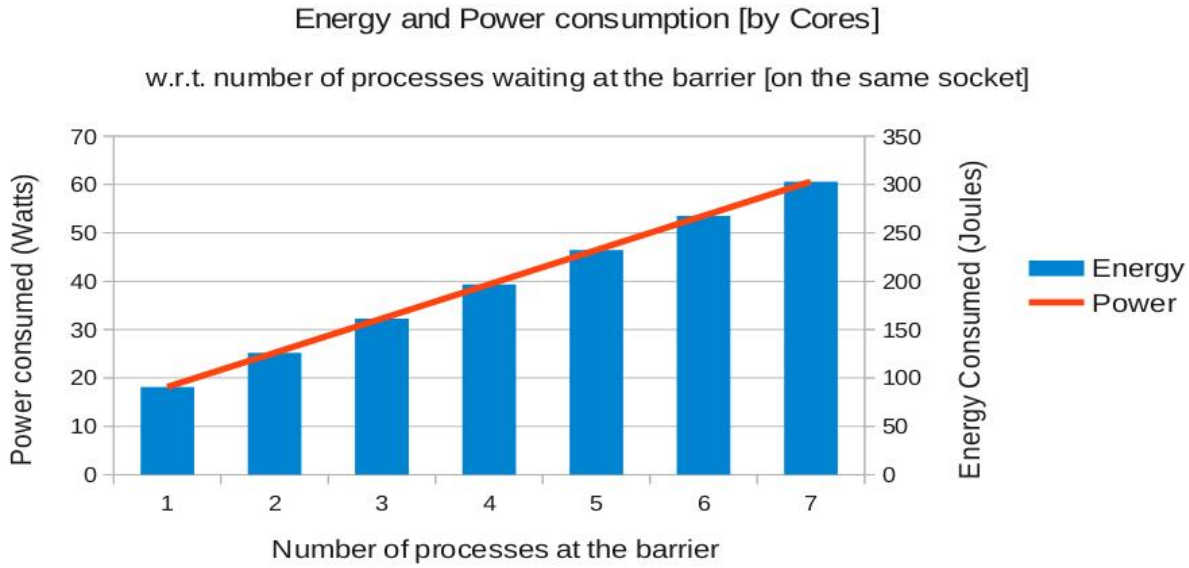




Energy and power efficiency - software stack

Synchronizations

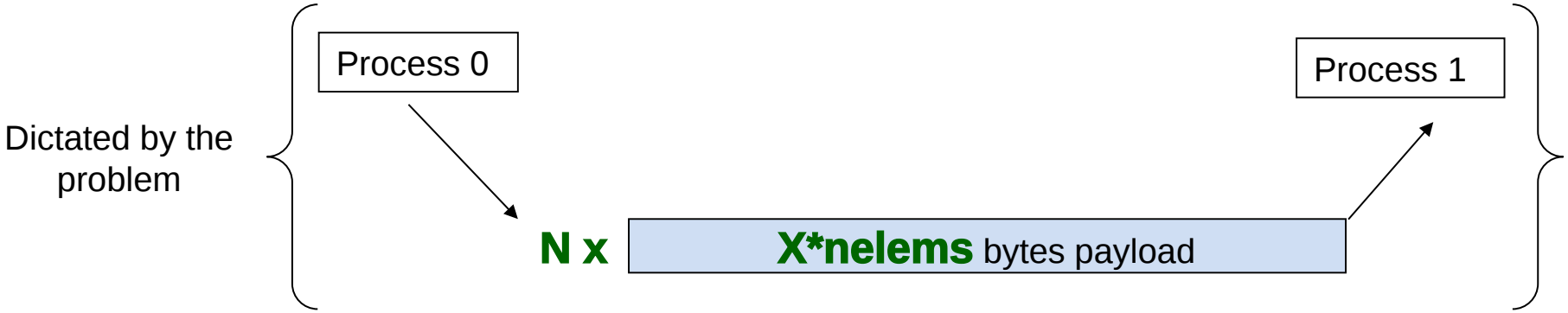




Energy and power efficiency - software stack

Remote data transfers

Consider the following case for a given problem:



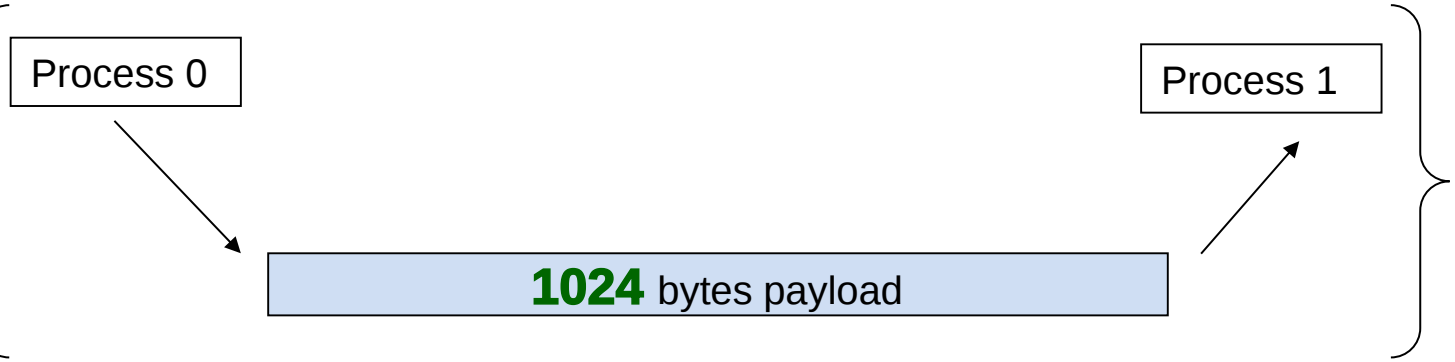
N * shmem_**putX** (void *target, const void *source, **size_t nelems**, int pe)

Energy and power efficiency - software stack

Remote data transfers

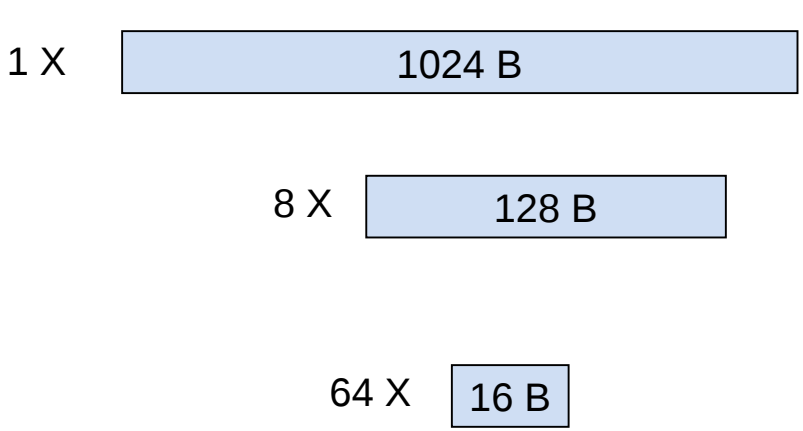
Consider the following case for a given problem:

Dictated by the problem



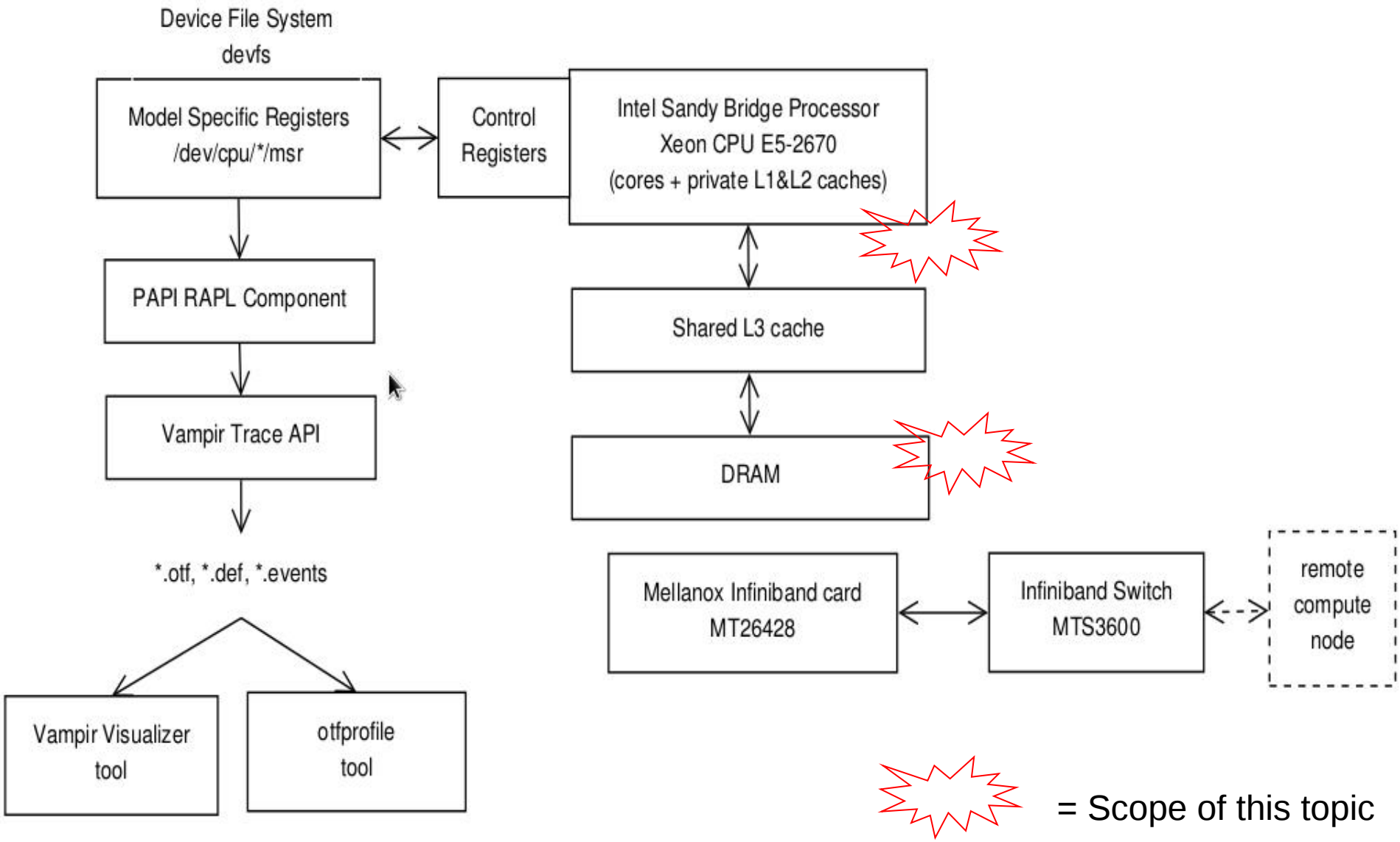
This can be transferred as:

Dictated by the algorithm design



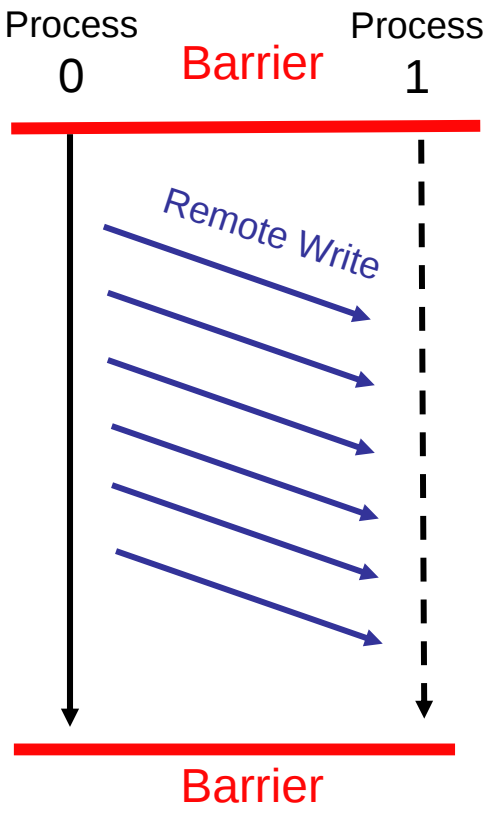
The Impact of Remote Transfers

Experimental Setup



The Impact of Remote Transfers

Microbenchmarks

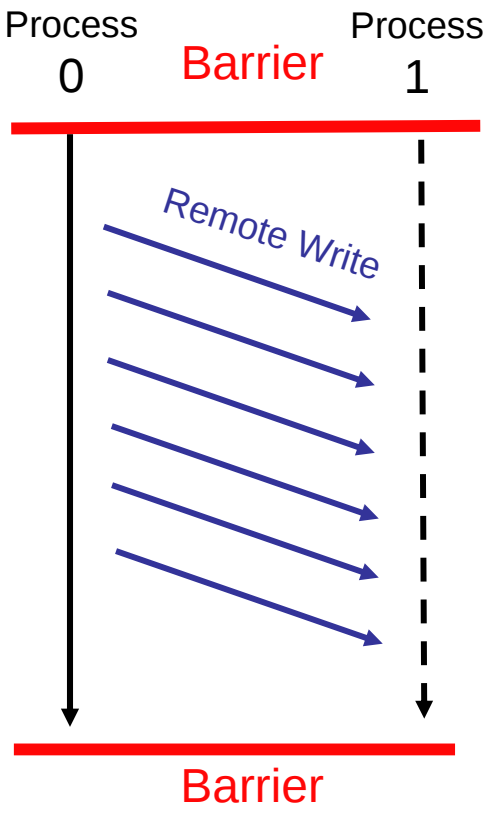



```
me = my_pe() ;
for(j=1 ; j<=MAX_WRK_SIZE ; j*=2)
{
   for(frag_cnt=MIN_MSG_NUM ;
            frag_cnt<=j ; frag_cnt*=2 )
   {
      bytes_per_frag = j / frag_cnt;
      shmem_barrier_all();
      // instrumentation start
      if(me == 0)
         for (it=0; it<frag_cnt; it++)
            shmem_putmem(  ....., bytes_per_frag, 1);
      shmem_barrier_all();
      // instrumentation end
   }
}
```

Remote Write

The Impact of Remote Transfers

Microbenchmarks



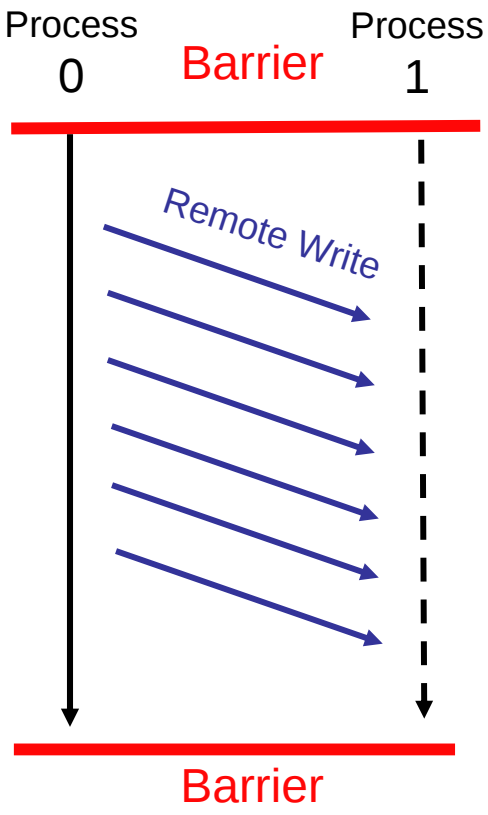

```
me = my_pe() ;
for(j=1 ; j<=MAX_WRK_SIZE ; j*=2)
{
   for(frag_cnt=MIN_MSG_NUM ;
            frag_cnt<=j ; frag_cnt*=2 )
   {
      bytes_per_frag = j / frag_cnt;
      shmem_barrier_all();
      // instrumentation start
      if(me == 0)
         for (it=0; it<frag_cnt; it++)
            shmem_putmem(  ....., bytes_per_frag, 1);
      shmem_barrier_all();
      // instrumentation end
   }
}
```

Barrier

Barrier

The Impact of Remote Transfers

Microbenchmarks



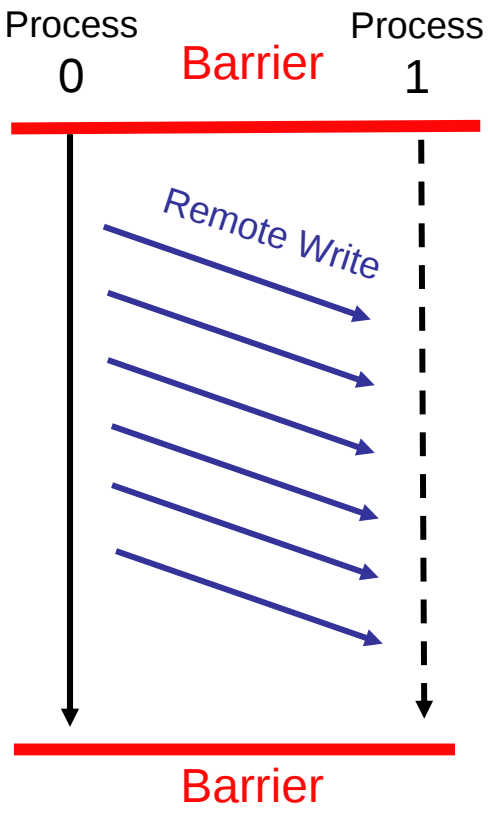
Msg Size

Number of transfers

```
me = my_pe() ;
for(j=1 ; j<=MAX_WRK_SIZE ; j*=2)
{
    for(frag_cnt=MIN_MSG_NUM ;
            frag_cnt<=j ; frag_cnt*=2 )
    {
        bytes_per_frag = j / frag_cnt;
        shmem_barrier_all();
        // instrumentation start
        if(me == 0)
            for (it=0; it<frag_cnt; it++)
                shmem_putmem(  ....., bytes_per_frag, 1);
        shmem_barrier_all();
        // instrumentation end
    }
}
```

The Impact of Remote Transfers

Microbenchmarks



Msg Size

Number of transfers

Barrier

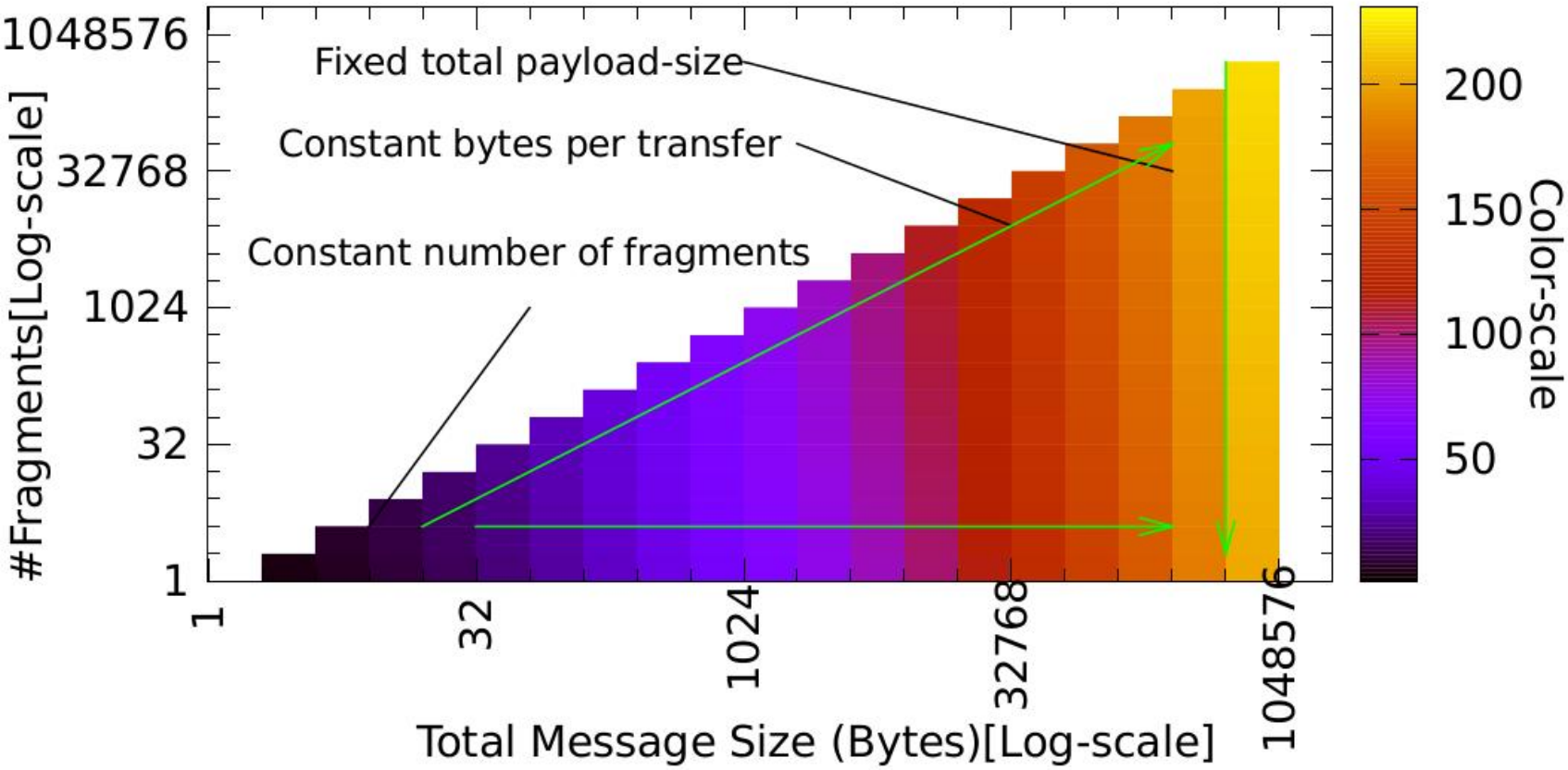
Remote Write

Barrier

```
me = my_pe() ;
for(j=1 ; j<=MAX_WRK_SIZE ; j*=2)
{
    for(frag_cnt=MIN_MSG_NUM ;
            frag_cnt<=j ; frag_cnt*=2 )
    {
        bytes_per_frag = j / frag_cnt;
        shmem_barrier_all();
        // instrumentation start
        if(me == 0)
            for (it=0; it<frag_cnt; it++)
                shmem_putmem(  ....., bytes_per_frag, 1);
        shmem_barrier_all();
        // instrumentation end
    }
}
```

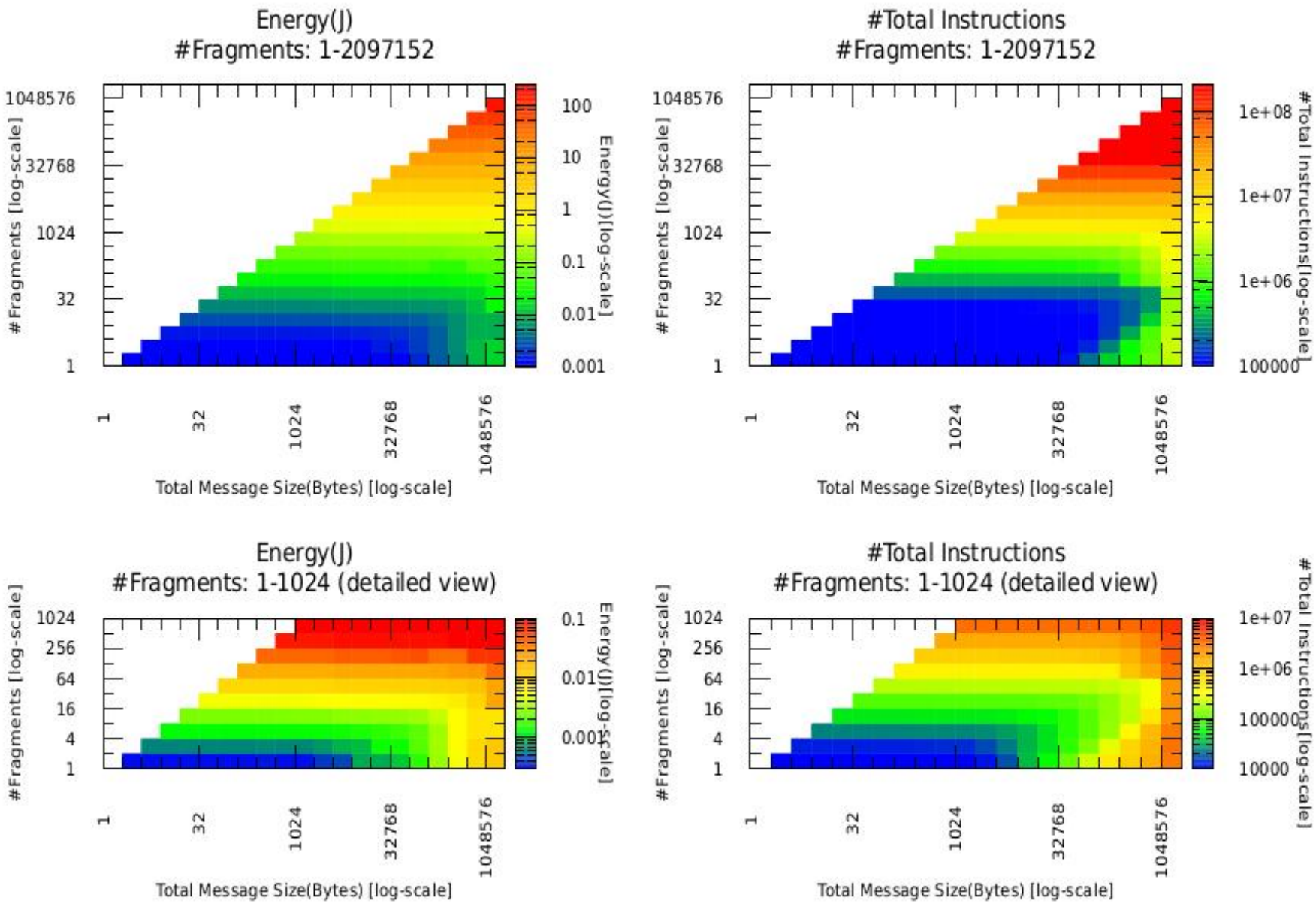
The Impact of Remote Transfers

A note on reading the plots



Impact of remote data transfers

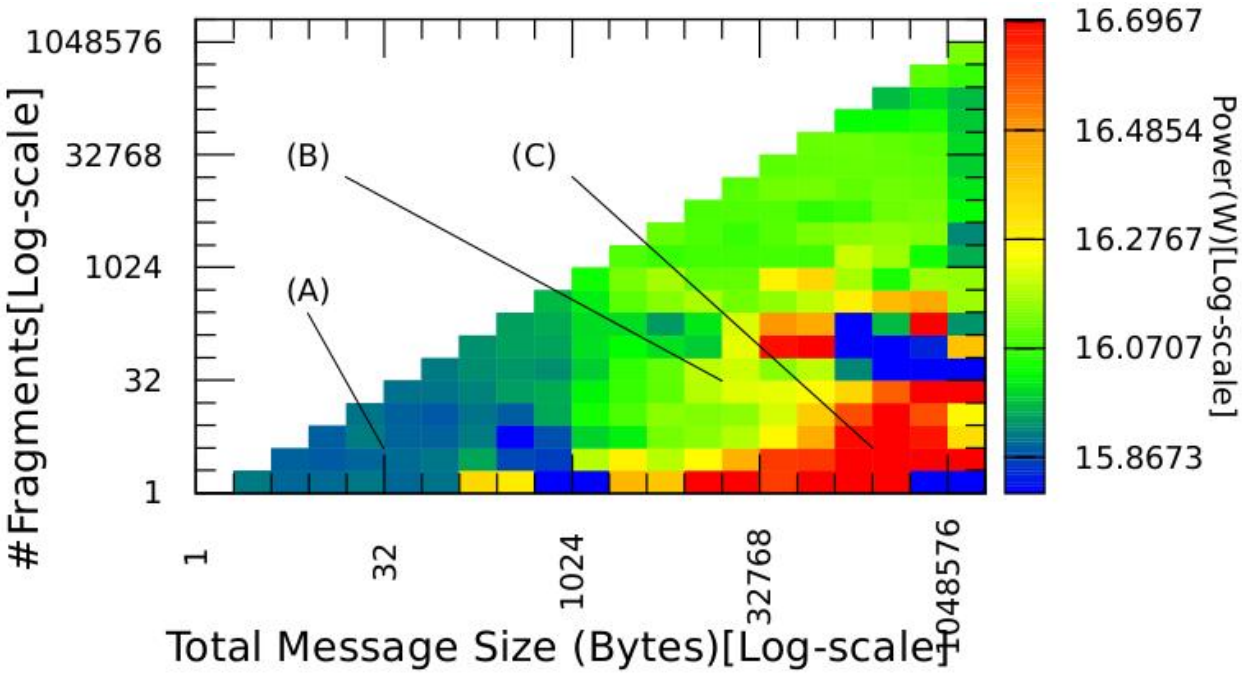
Energy consumption



The Impact of Remote Transfers

Cores - Power consumption

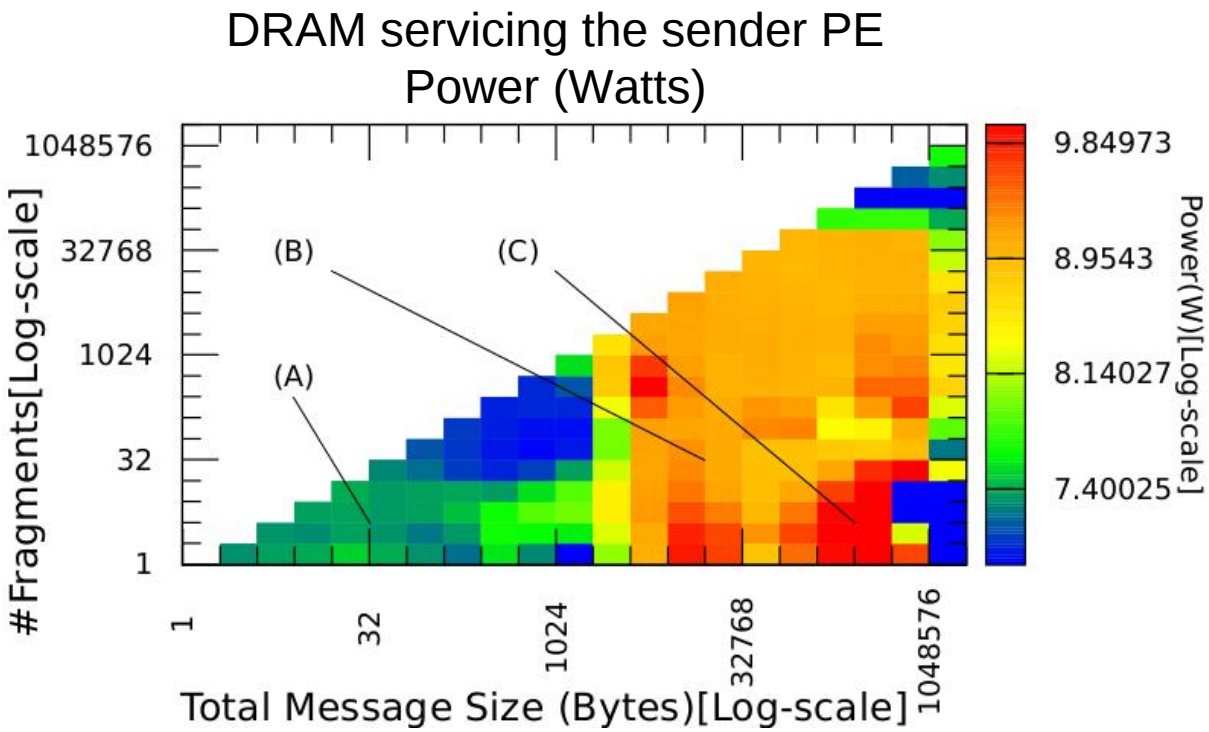
CPU Core servicing the sender PE
Power (Watts)



(A): Small data(< 1KB), (B): Medium data (1KB - 64KB), (C): Bulk data (>64KB)

The Impact of Remote Transfers

DRAM pressure due to CPU memops



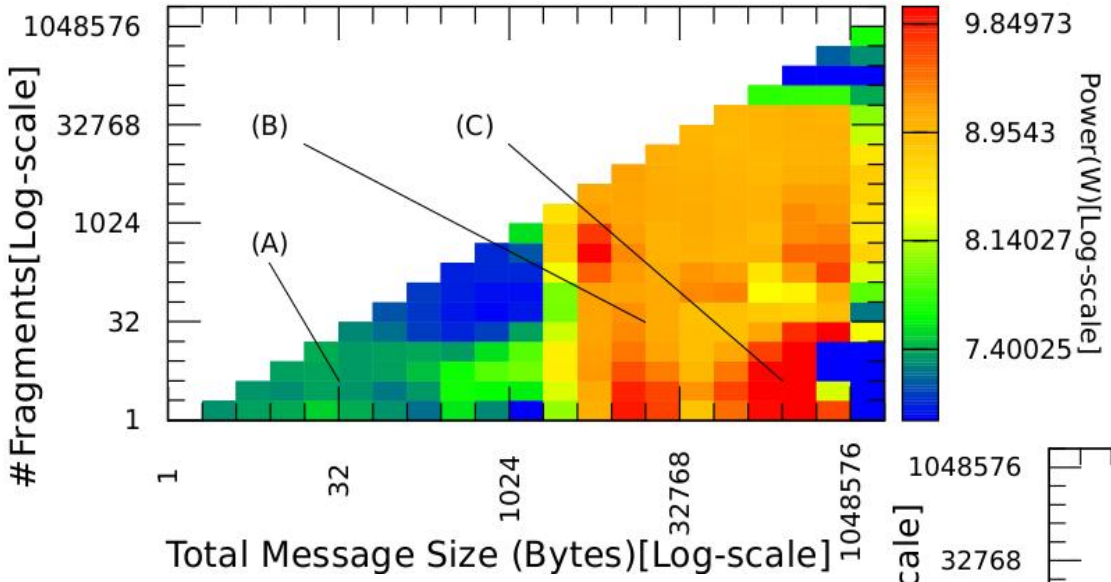
(A): Small data(< 1KB), (B): Medium data (1KB - 64KB), (C): Bulk data (>64KB)

The Impact of Remote Transfers

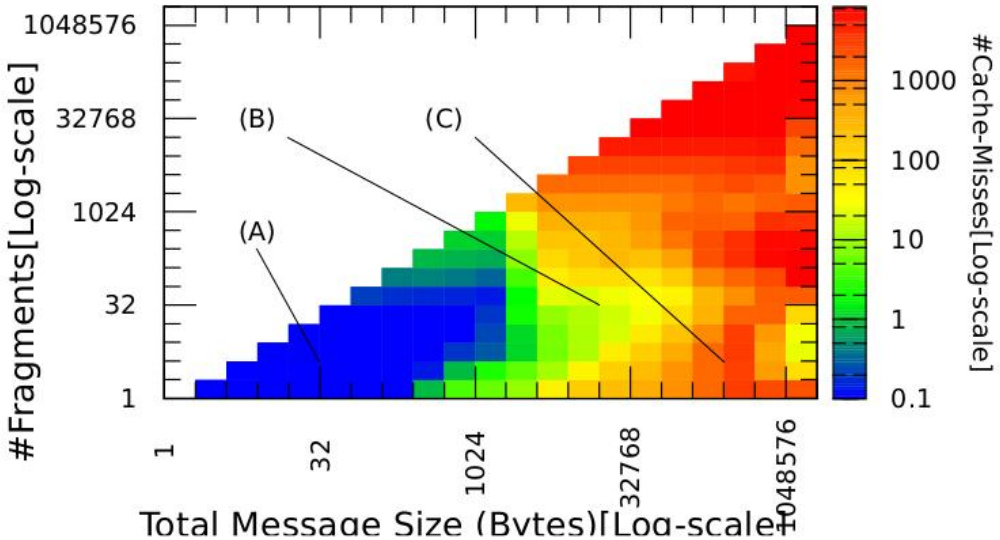
DRAM pressure due to CPU memops

Impact due to memory - hierarchy

DRAM



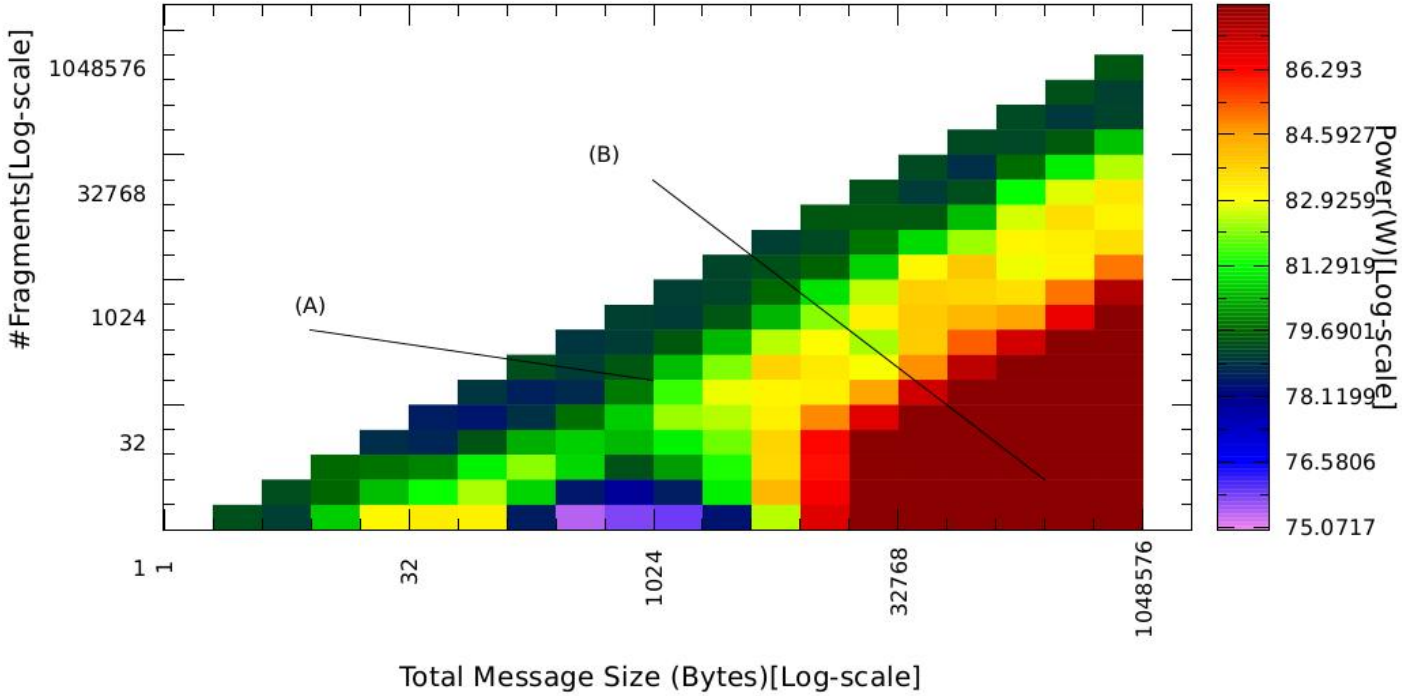
L3 - cache misses



(A): Small data(< 1KB), (B): Medium data (1KB - 64KB), (C): Bulk data (>64KB)

Impact of remote data transfers

DRAM - power consumption
at the source and target PE



Conclusions and Future Work

- Data movement is costly
  - Two design factors that impact power
    - Data transfer size
    - Number of explicit message transfers
  - Number of Instructions impact energy

- Design of Impact of programming practices
  - A rising concern !

Acknowledgments

- Funding supported by the U.S. DOD
- Resources supported by ORNL, TN
- Facility supported by U.S. DOE through Contract No. DE-AC05-00OR22725.
- Project supported by HPCTools @ University of Houston
- Thanks are due to support teams of:
  - PAPI @ University of Tennessee, Knoxville
  - VampirTrace @ Technische Universität Dresden
  - PowerPack @ Virginia Tech

Thank You !

Questions ?