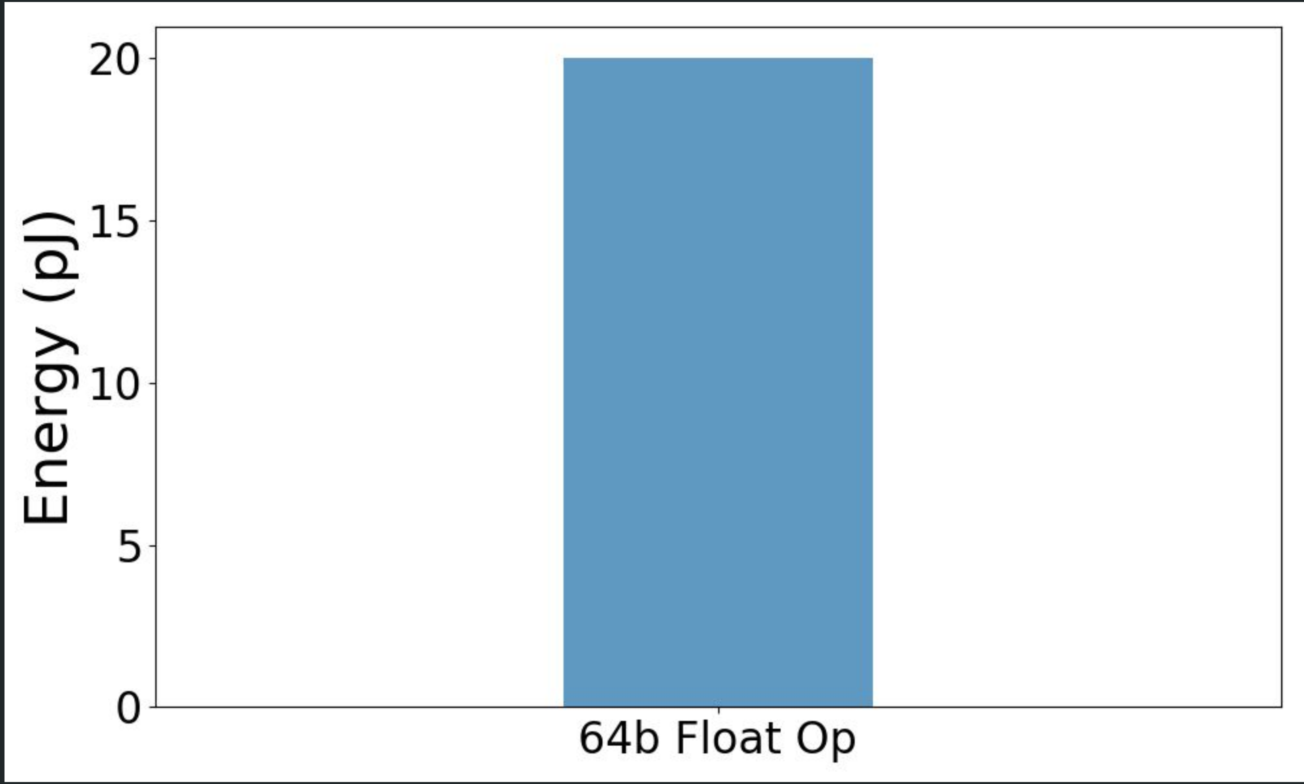
Multi-spectral Reuse Distance:
Divining Spatial Information from Temporal Data

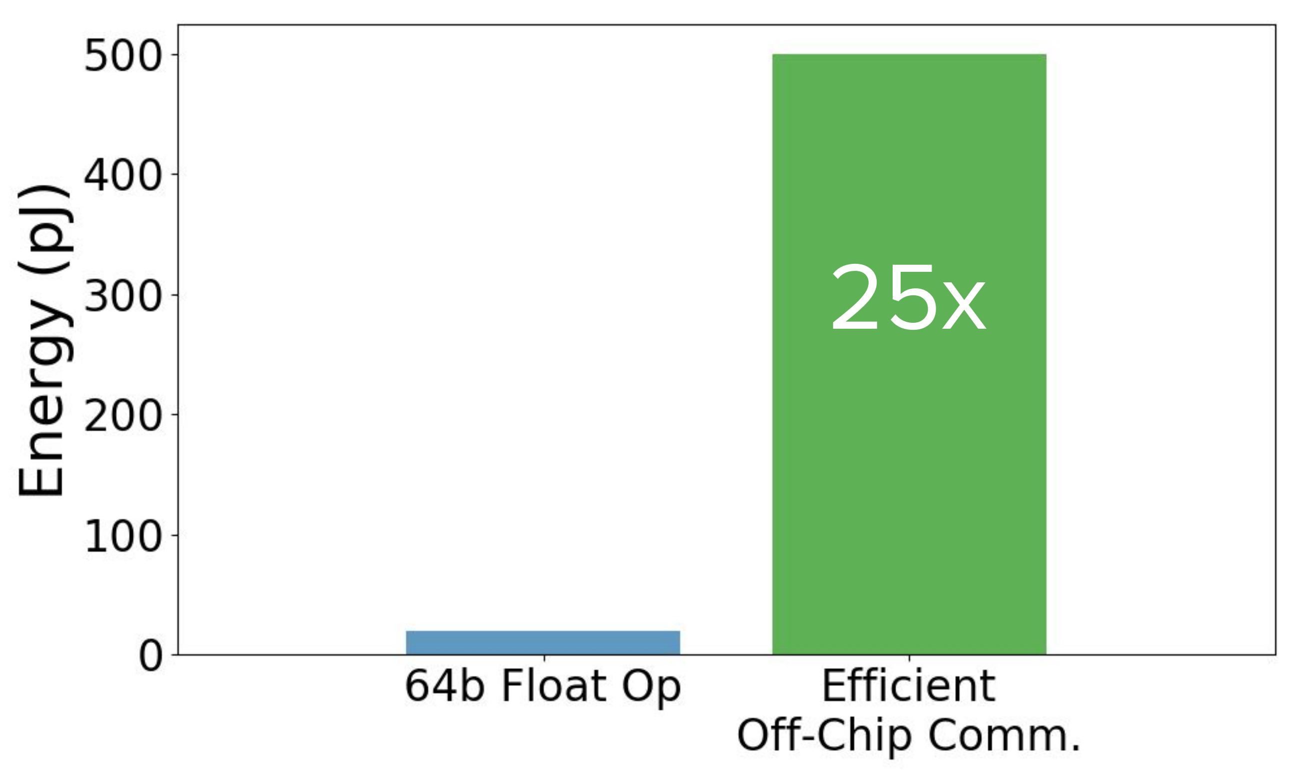
Anthony Cabrera*, Roger Chamberlain*, Jonathan Beard†
*Washington University in St. Louis, MO, USA
†Arm Research, Austin, TX, USA

HPEC `19, Waltham, MA, USA

The Data Movement Problem

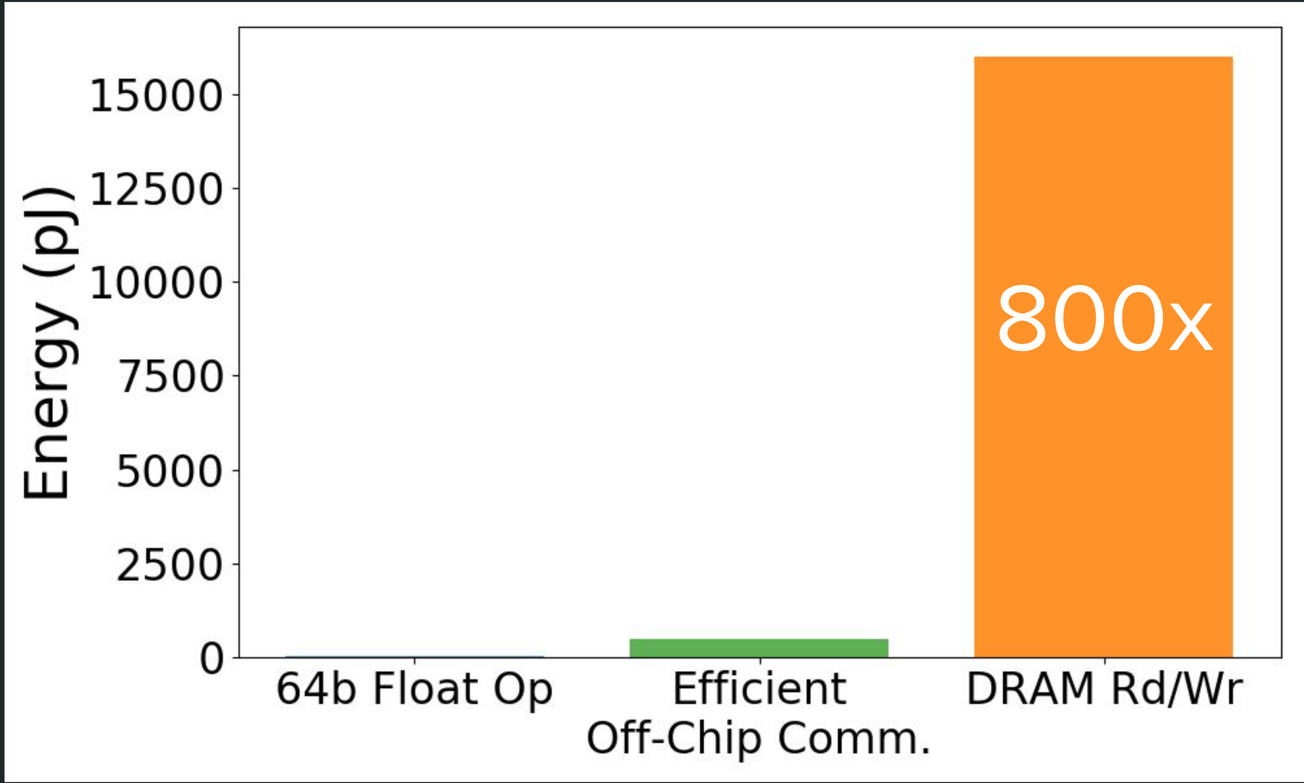


The Data Movement Problem



WJ Dally, HotPar '13 Keynote

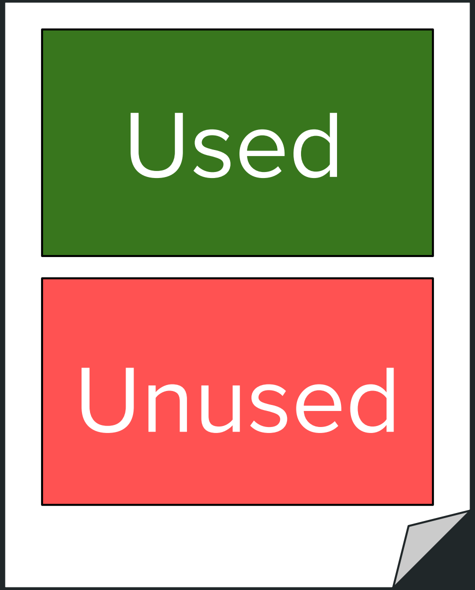
The Data Movement Problem



WJ Dally, HotPar '13 Keynote

Superfluous Data Movement Hurts

Paging data that never gets used



Our Contribution

- Develop a tool to inform the relationship between spatial and temporal locality

- Qualify spatial locality from multispectral reuse distance
  AND
  Quantify spatial locality from Earth Mover's Distance

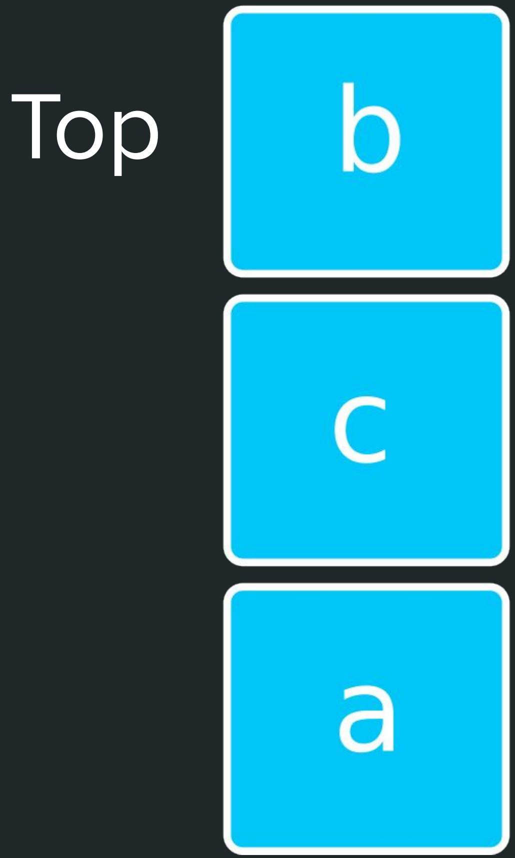
- Identify opportunities to reduce data movement
  AND
  Inform memory subsystem design/management

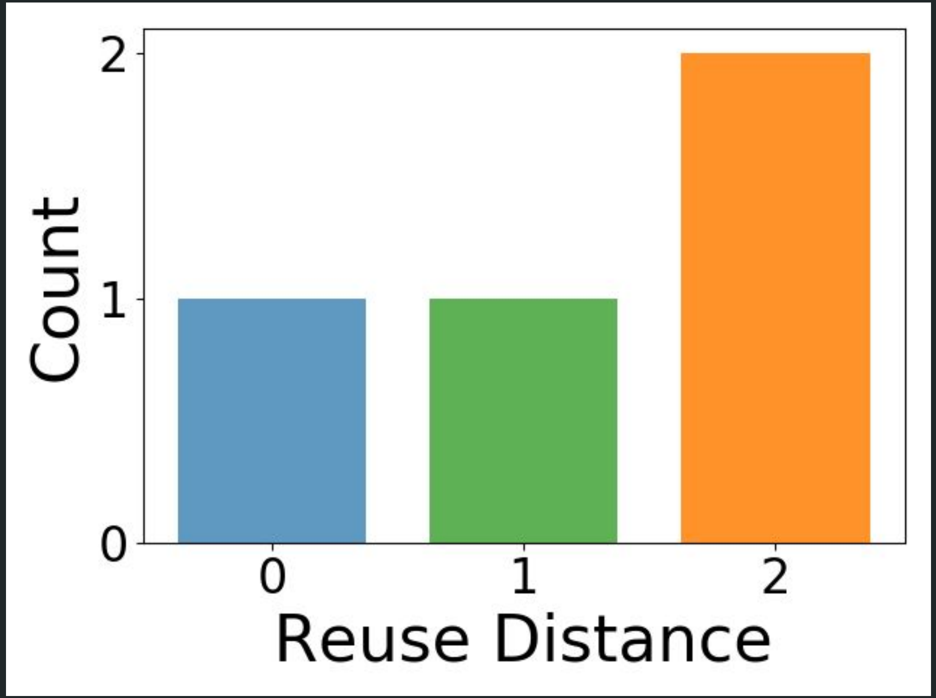
Method Overview

SPEC2006 Regions of Interest + DynamoRIO (The Dr. is in.) = Instruction Trace

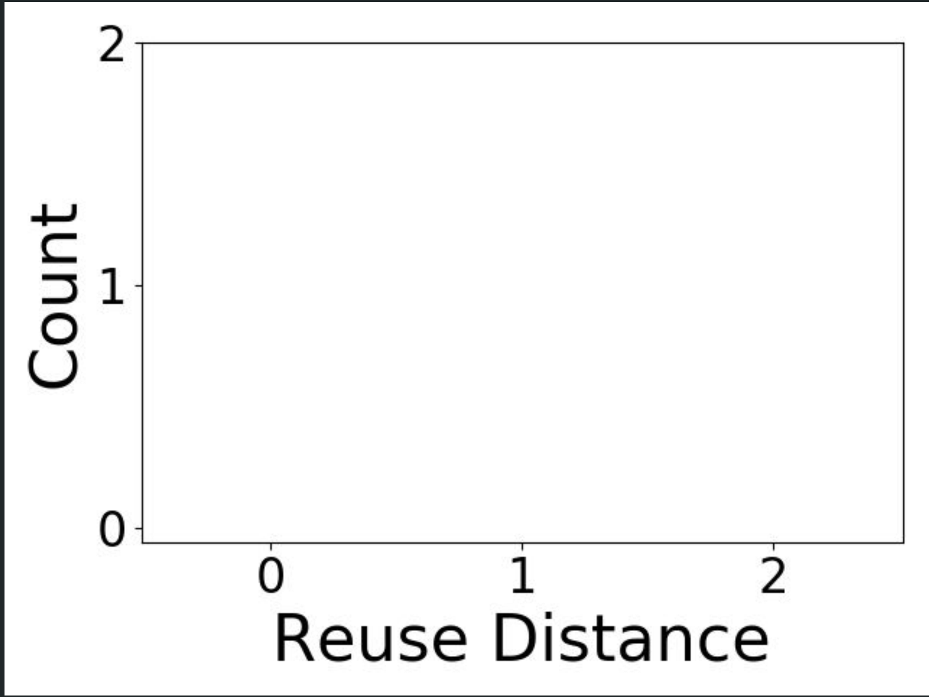
Instruction Trace + Reuse Distance @ 64B, 4KiB, 2MiB Granularities = Reuse Signatures, Earth Mover's Distance, Memory Footprint

Reuse Distance Primer: a b c a a c **b**





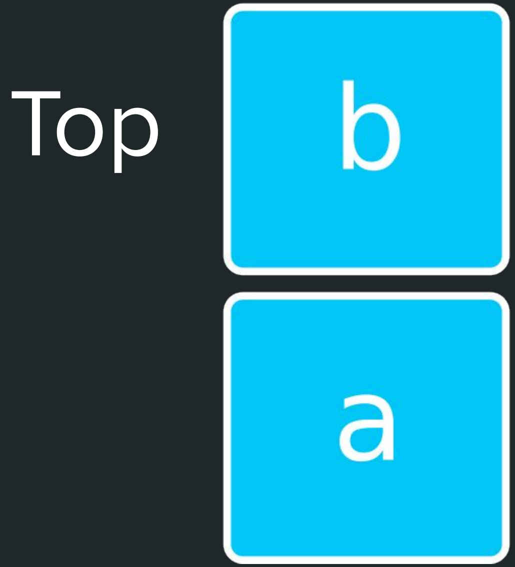
Reference Trace: **a**

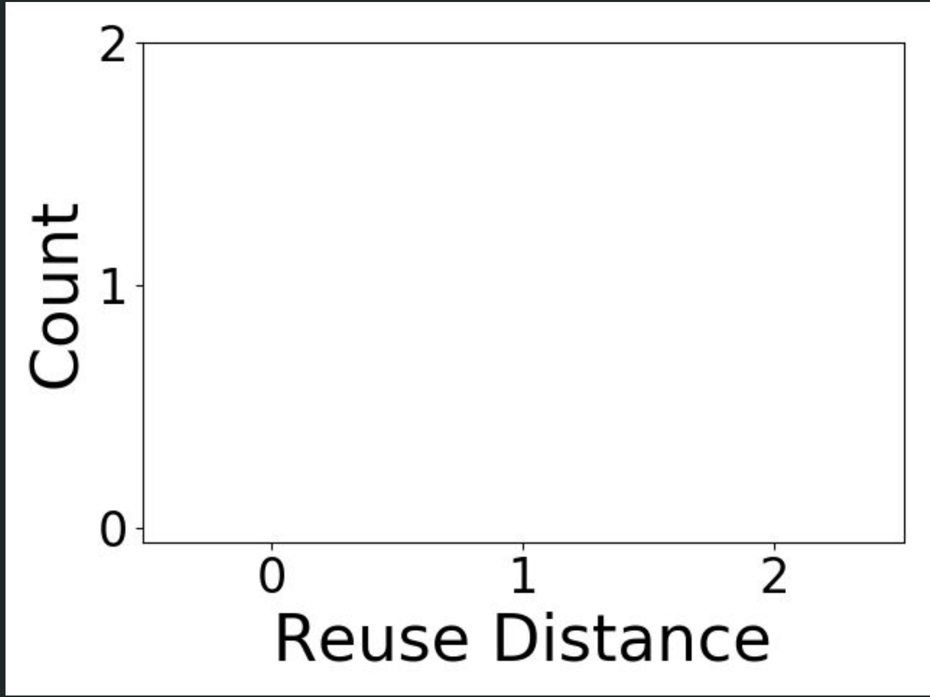


Top

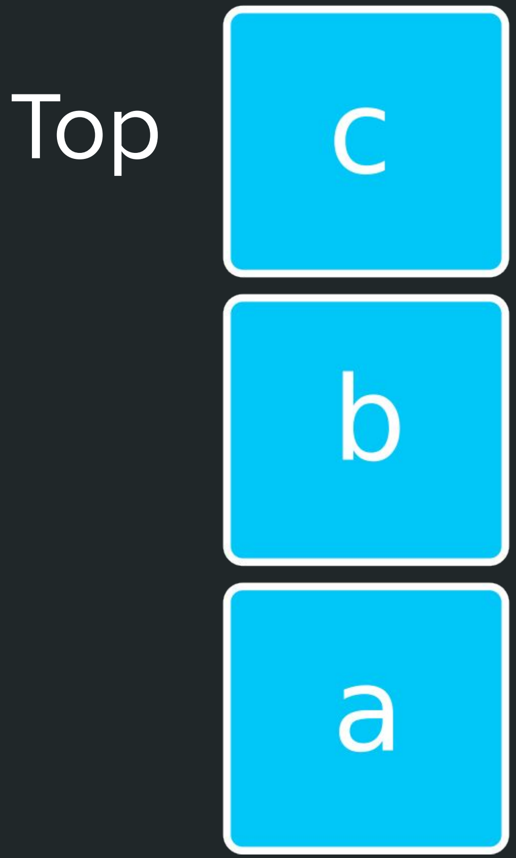
a

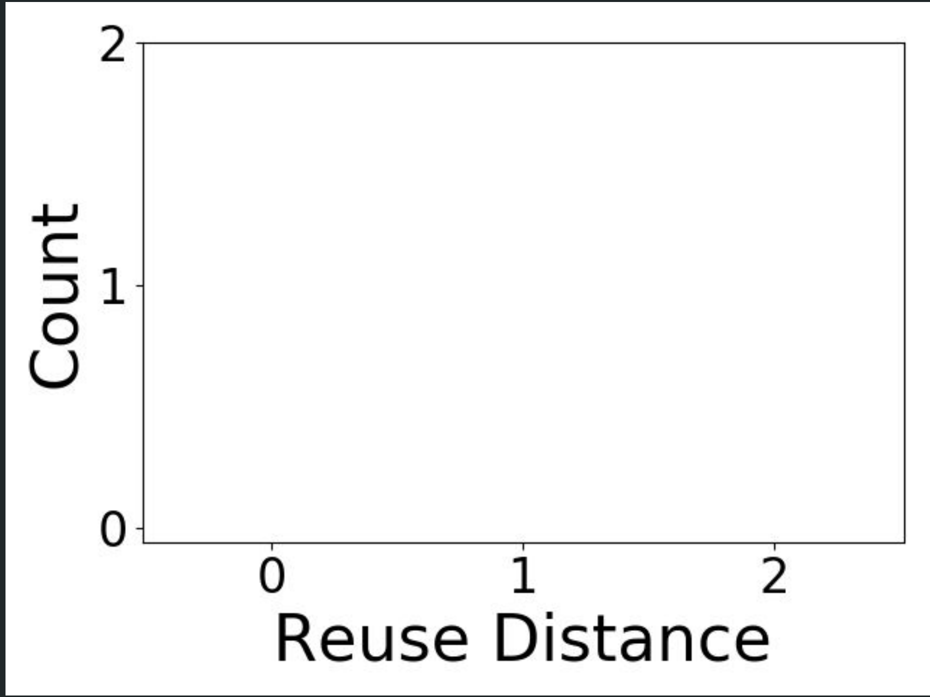
Reference Trace: a **b**





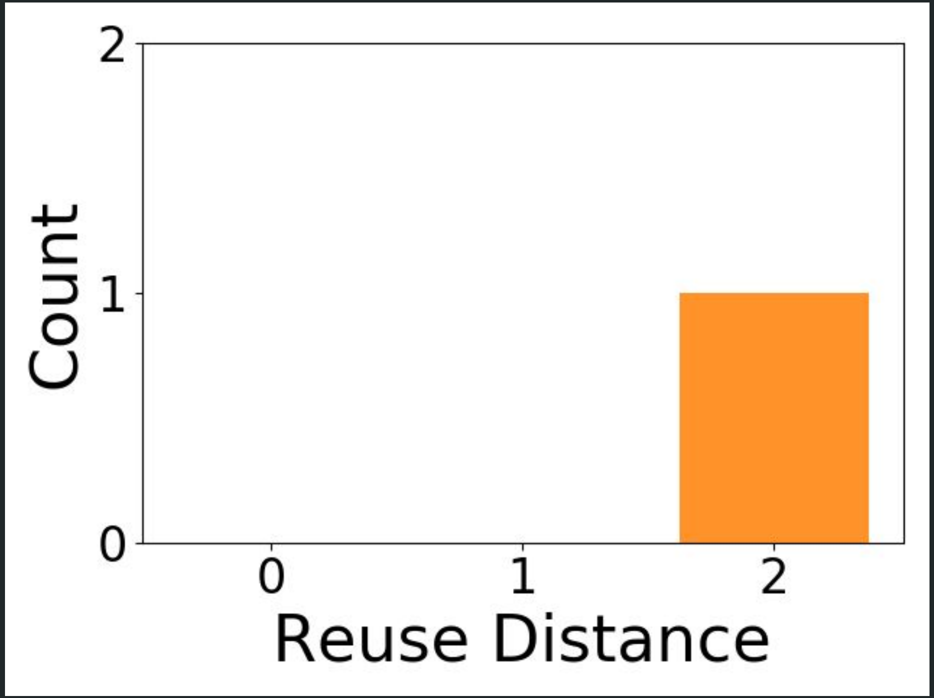
Reference Trace: a b **c**



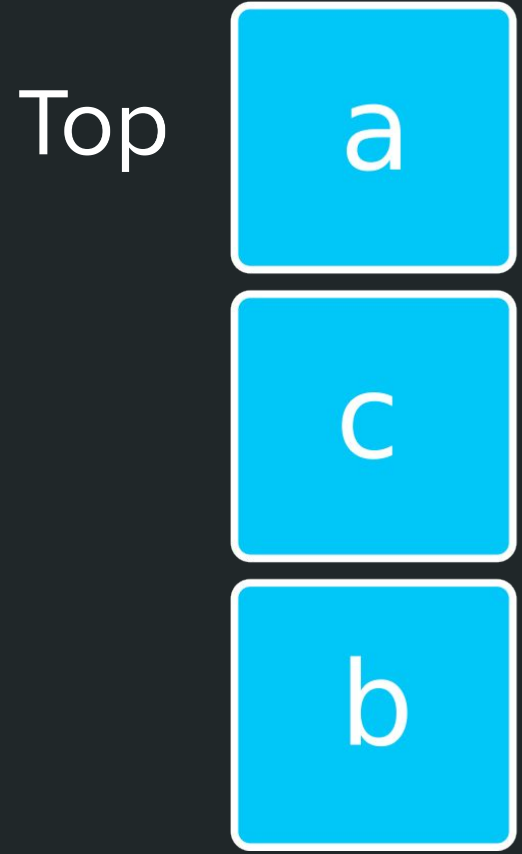


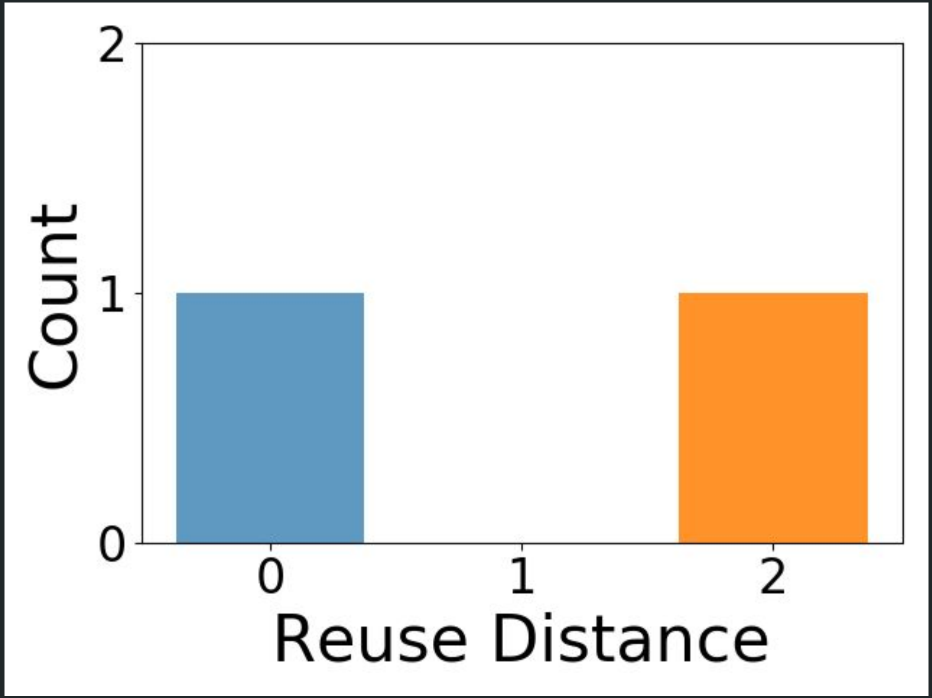
Reference Trace: a b c **a**

Top



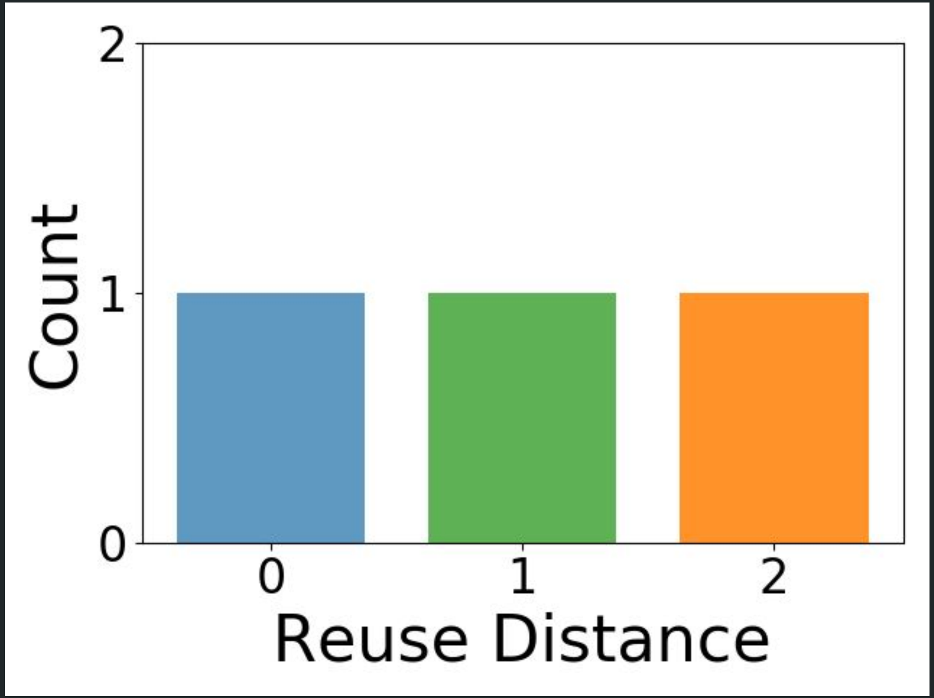
Reference Trace: a b c a **a**





Reference Trace: a b c a a **c**

Top



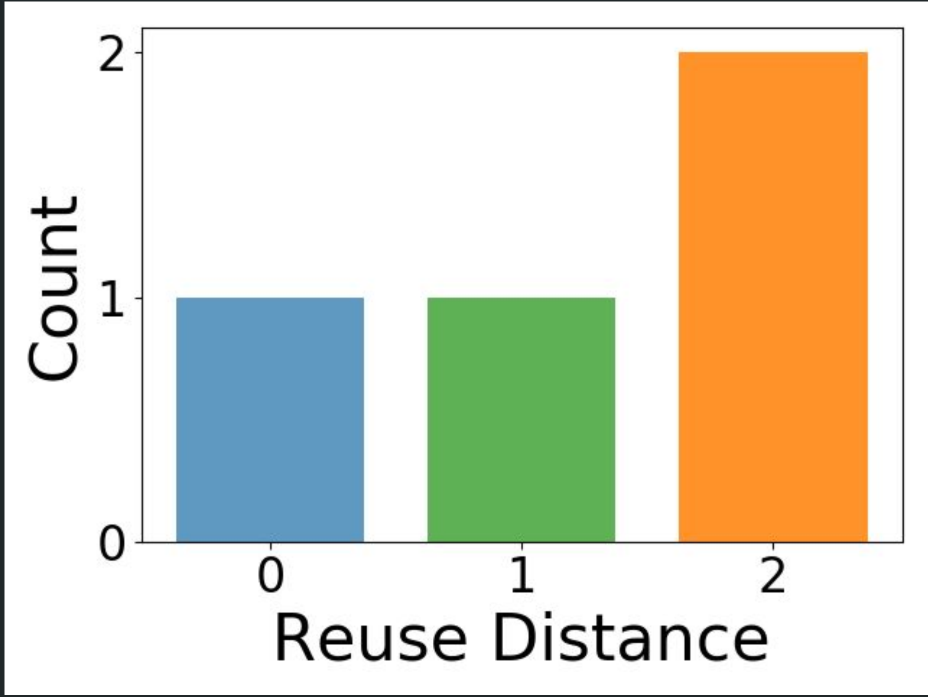
Reference Trace: a b c a a c **b**

Top

b

c

a



Reuse Distance *Granularity*

The size of the address blocks used in the reuse distance analysis

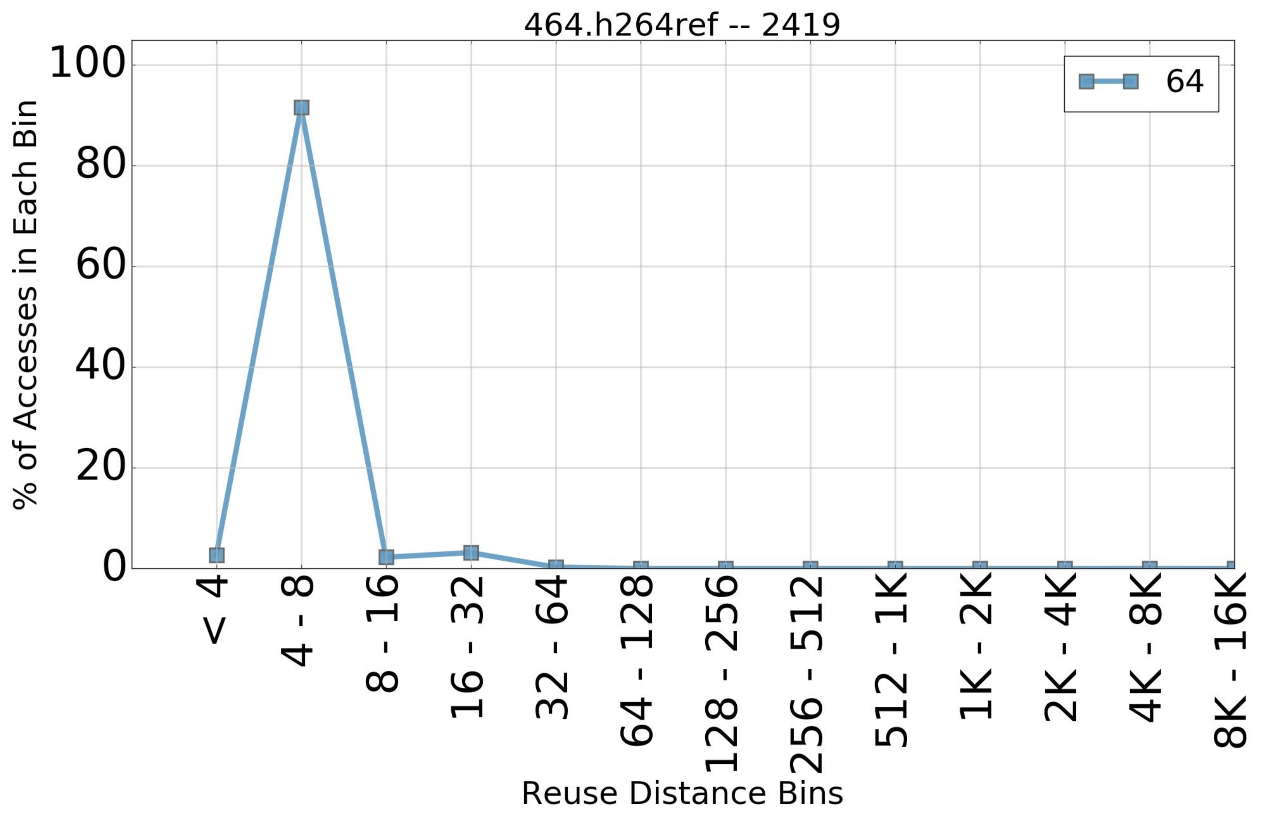
We vary granularity size in order to qualify and quantify spatial locality from the temporal data

Results

Spatially Dense (or not) Memory Access Patterns

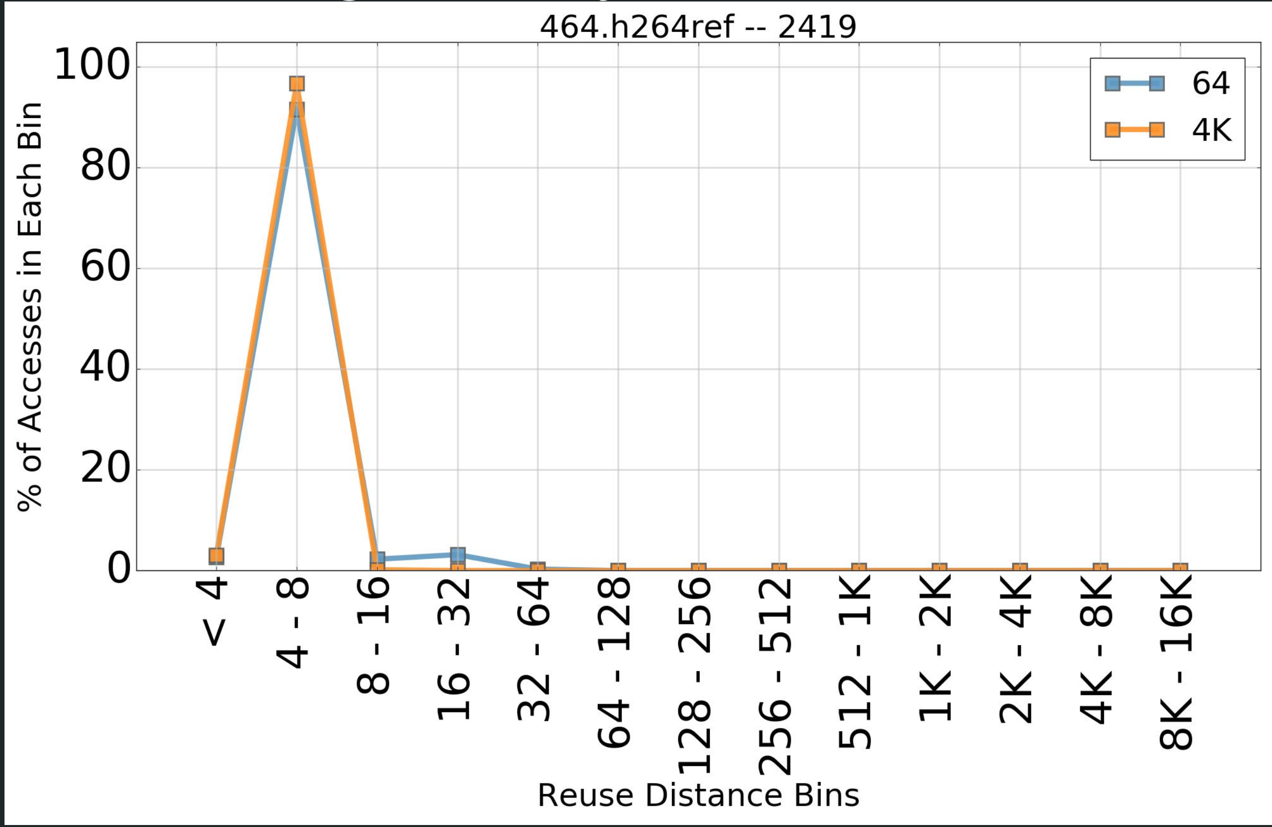
The Two Prototypical Behaviors

1) Mass Shifts left as granularity increases



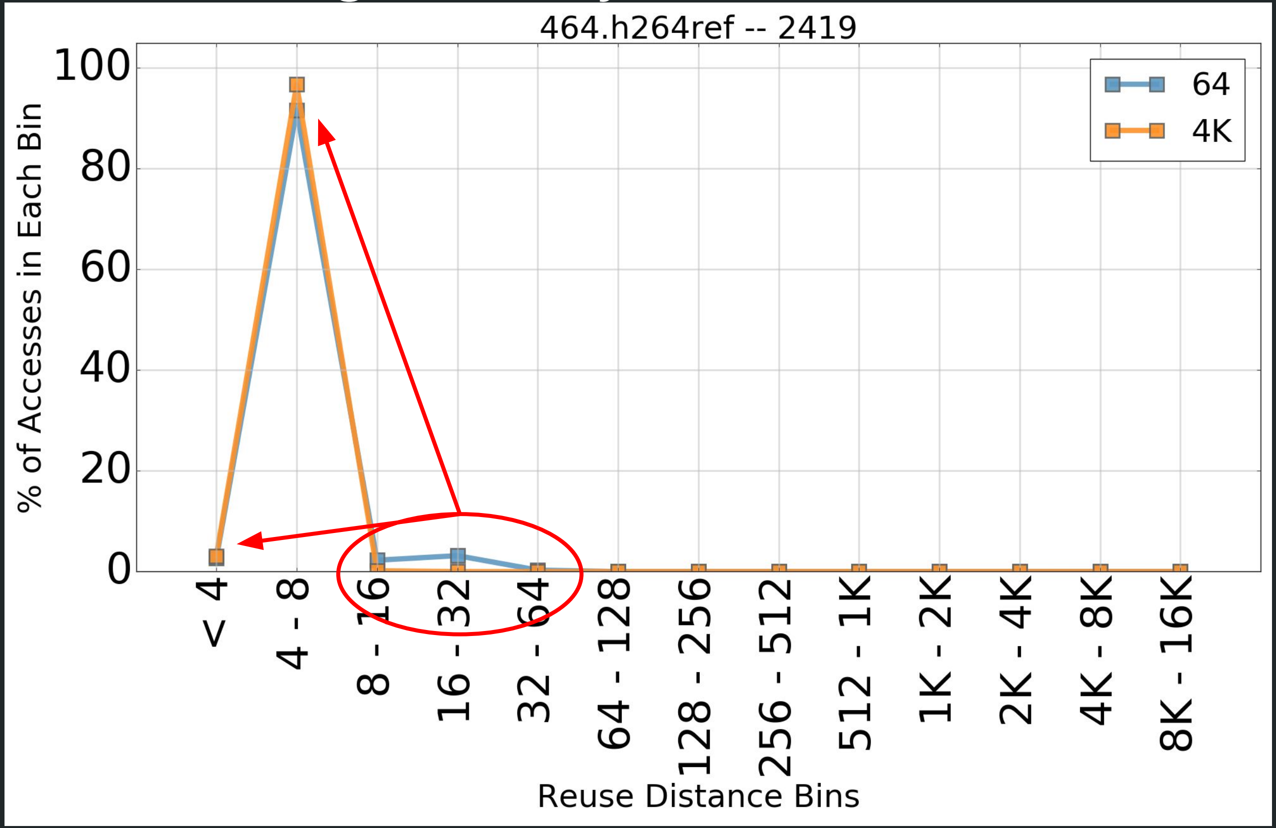
The Two Prototypical Behaviors

1) Mass Shifts left as granularity increases



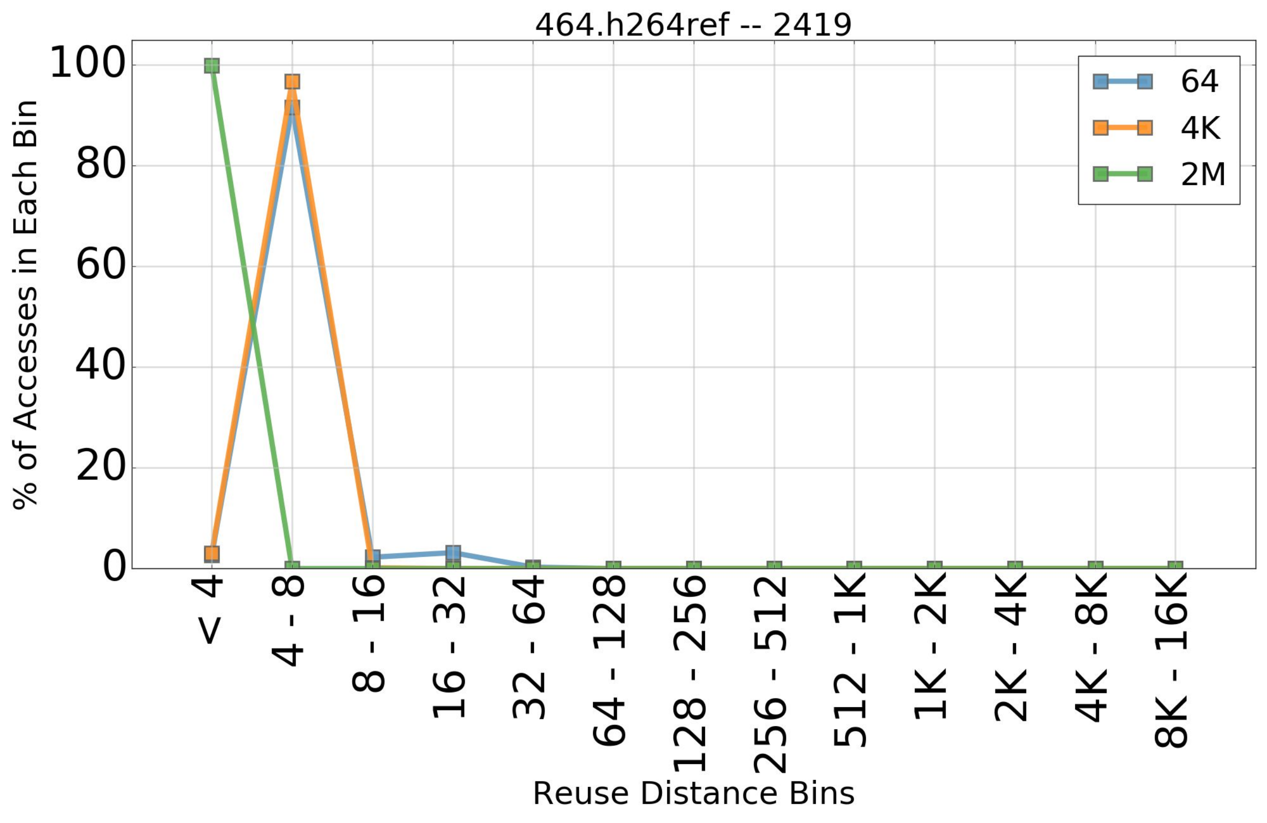
The Two Prototypical Behaviors

1) Mass Shifts left as granularity increases



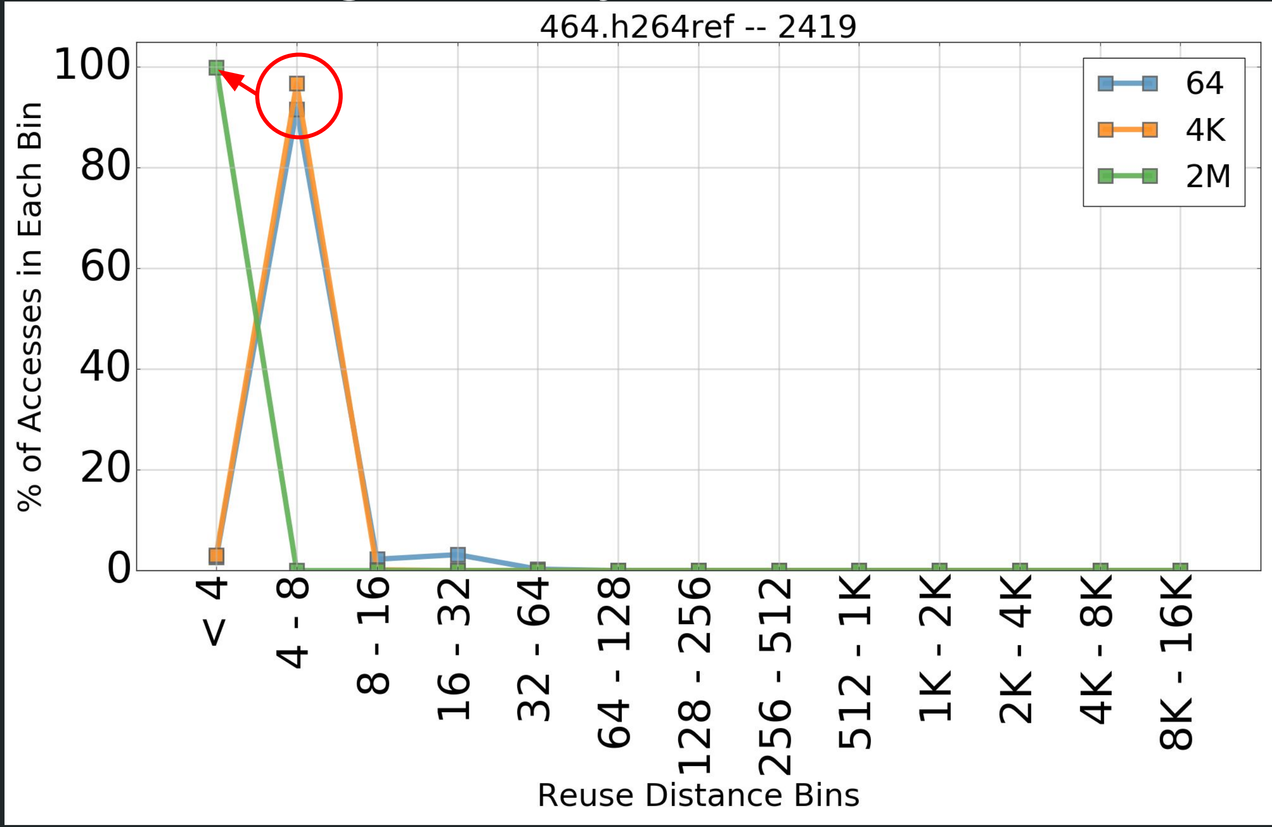
The Two Prototypical Behaviors

1) Mass Shifts left as granularity increases



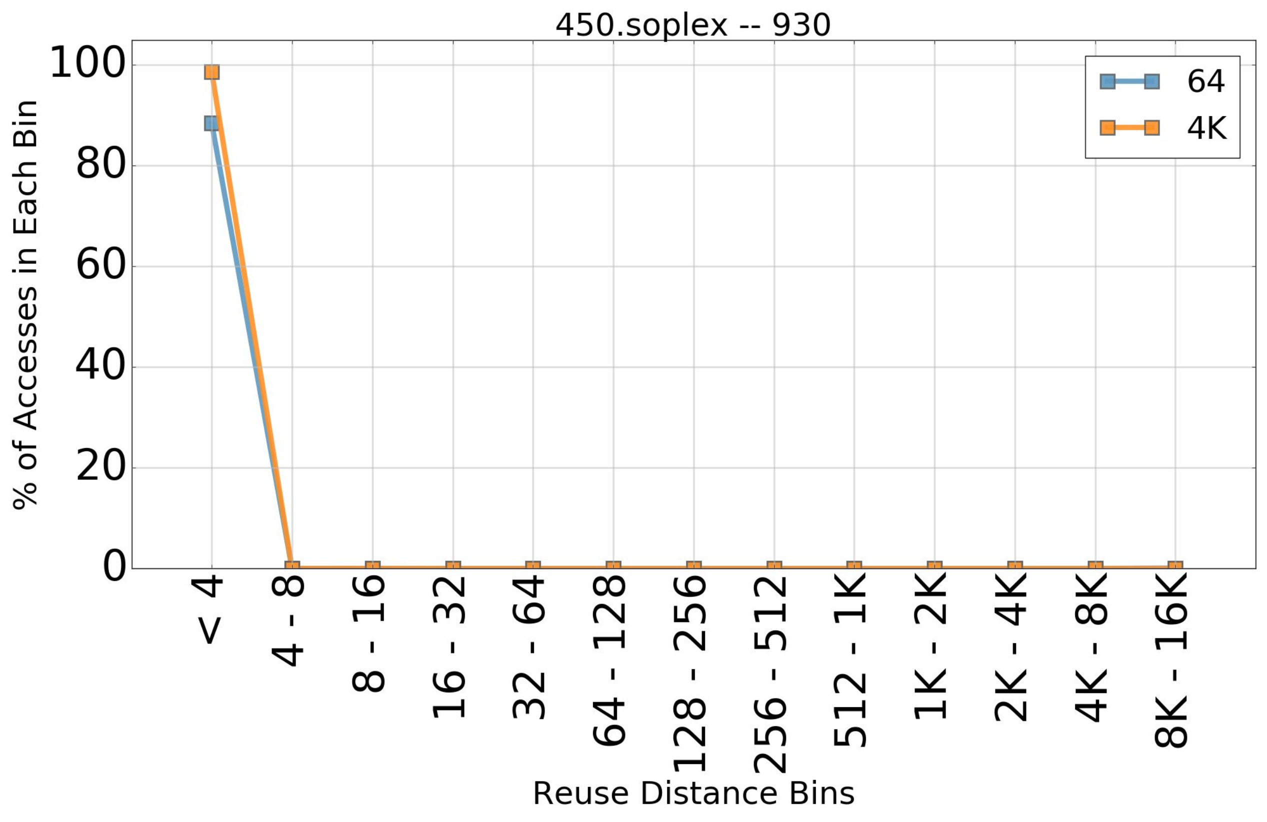
The Two Prototypical Behaviors

1) Mass Shifts left as granularity increases



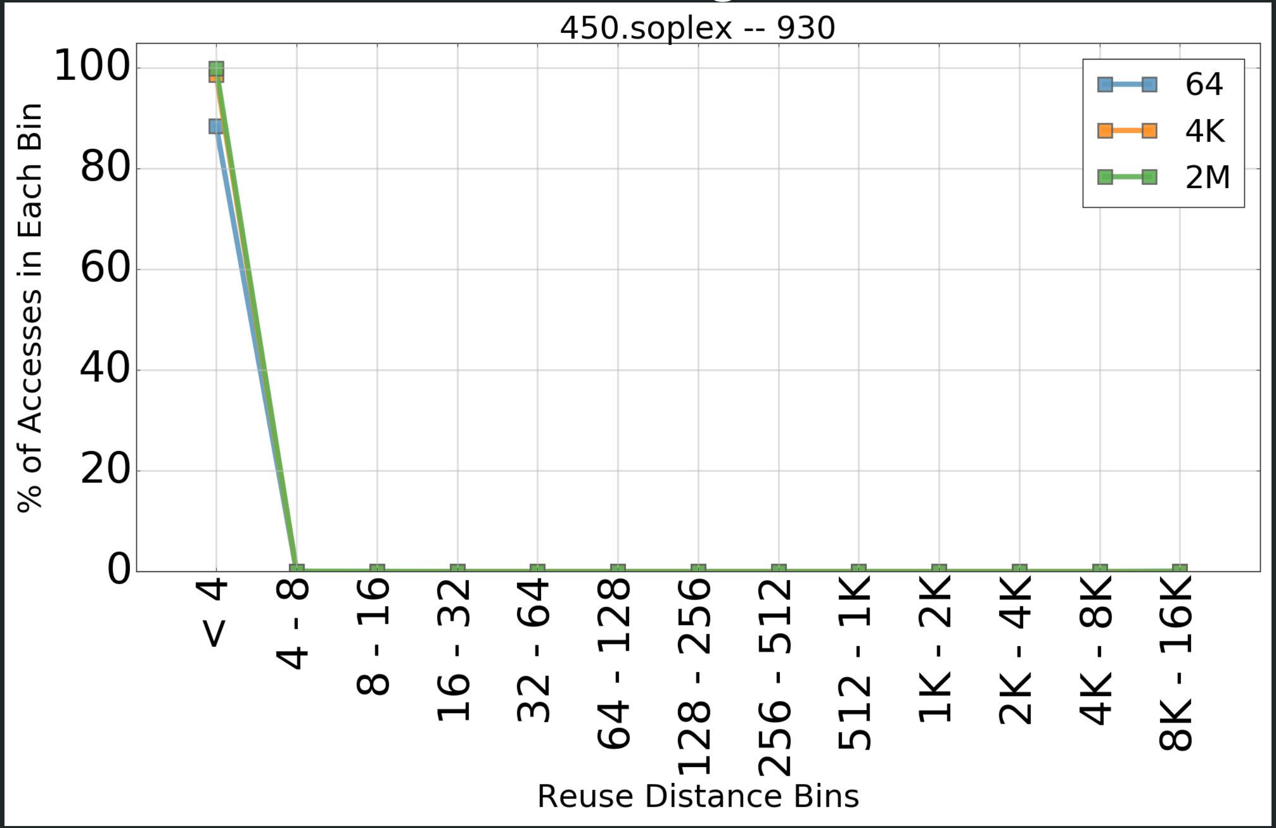
The Two Prototypical Behaviors

2) Mass remains the same across granularities



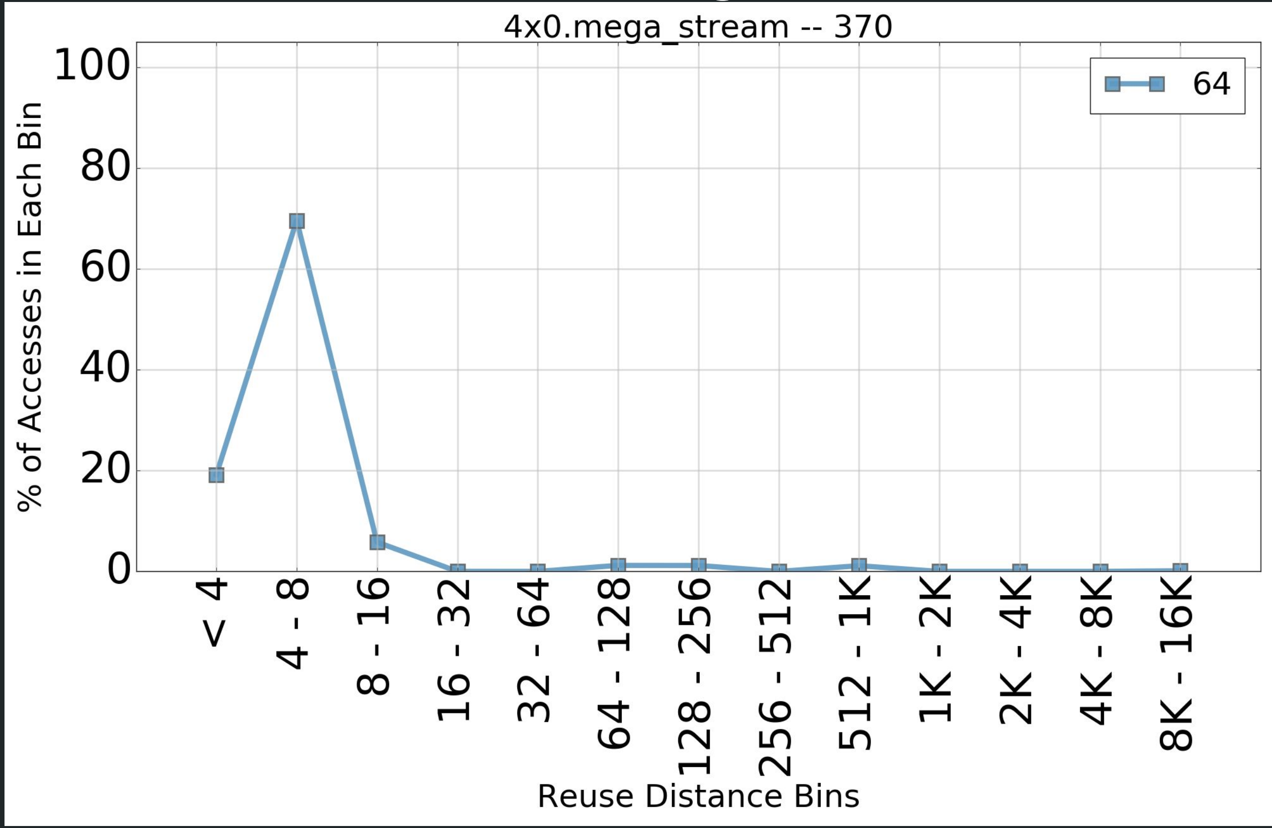
The Two Prototypical Behaviors

2) Mass remains the same across granularities



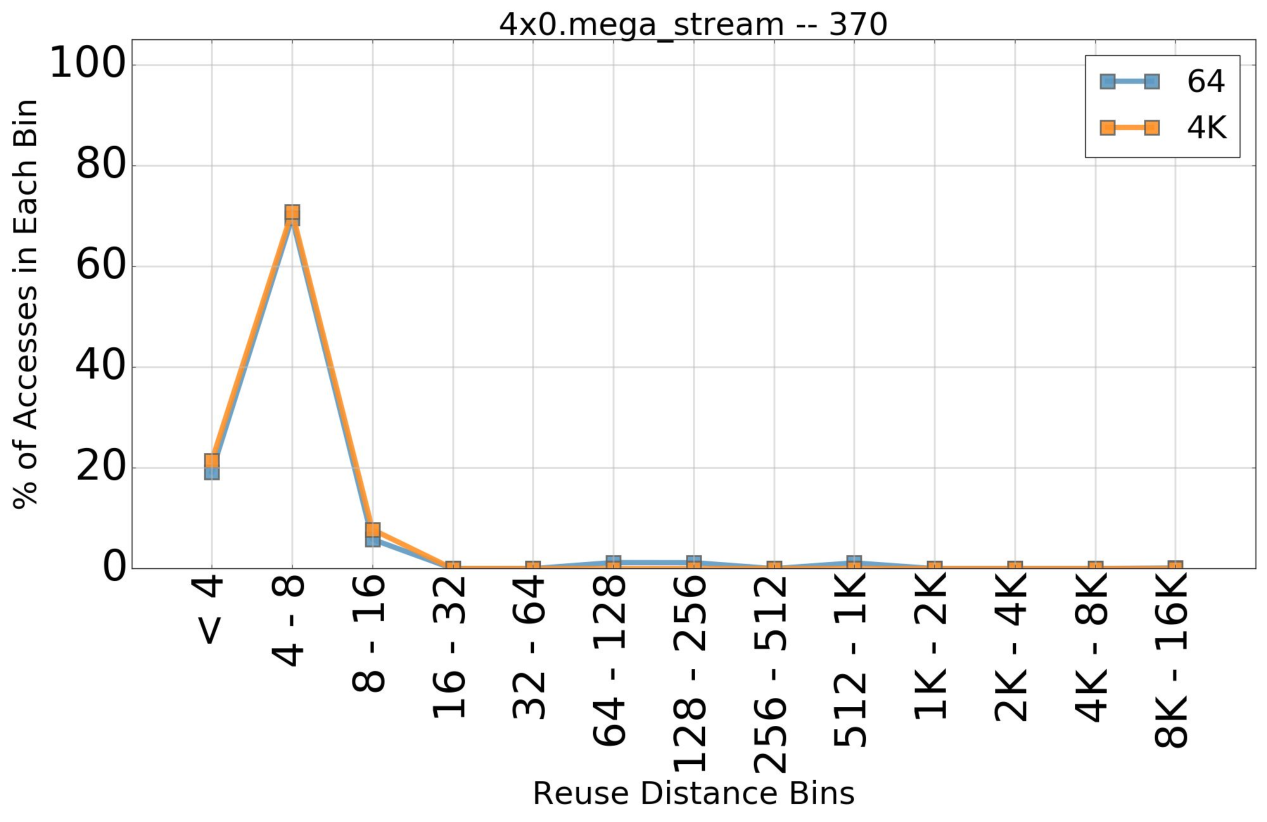
The Two Prototypical Behaviors

2) Mass remains the same across granularities



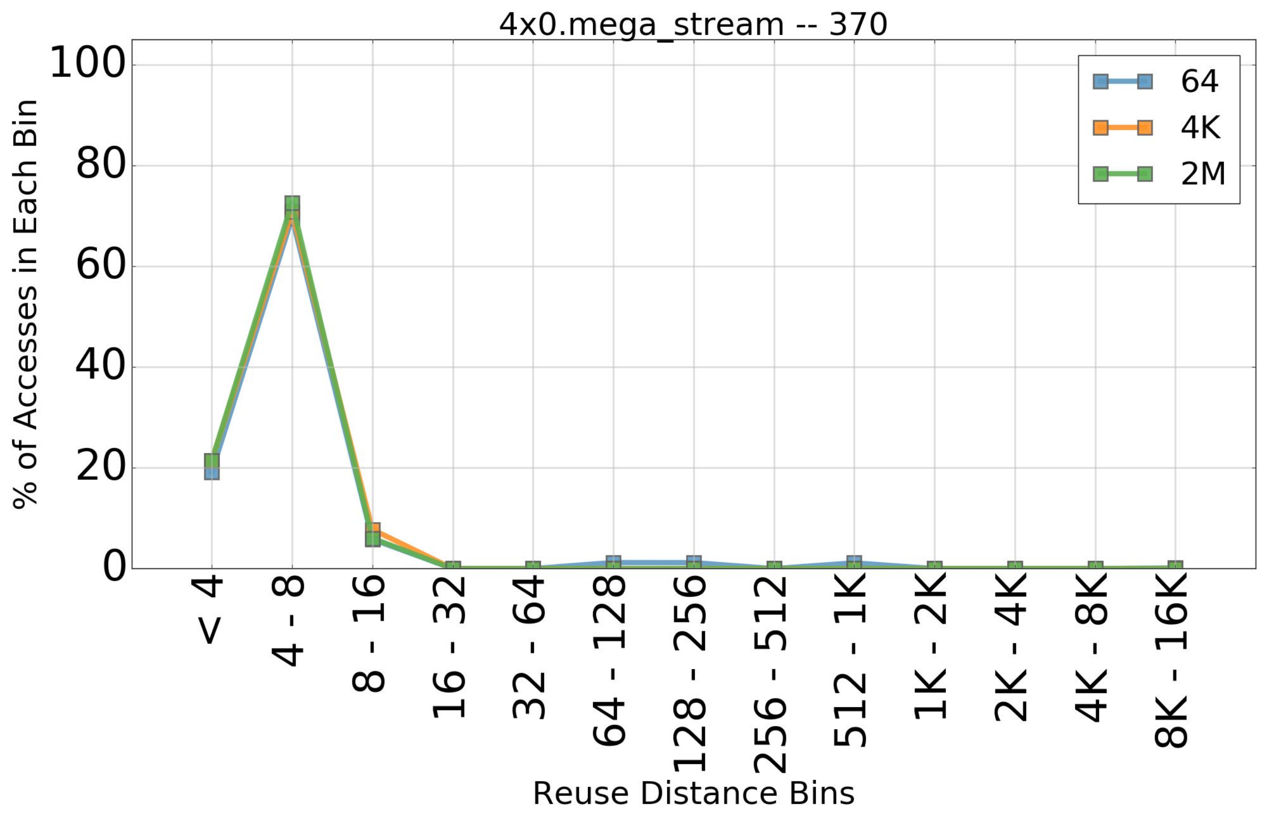
The Two Prototypical Behaviors

2) Mass remains the same across granularities



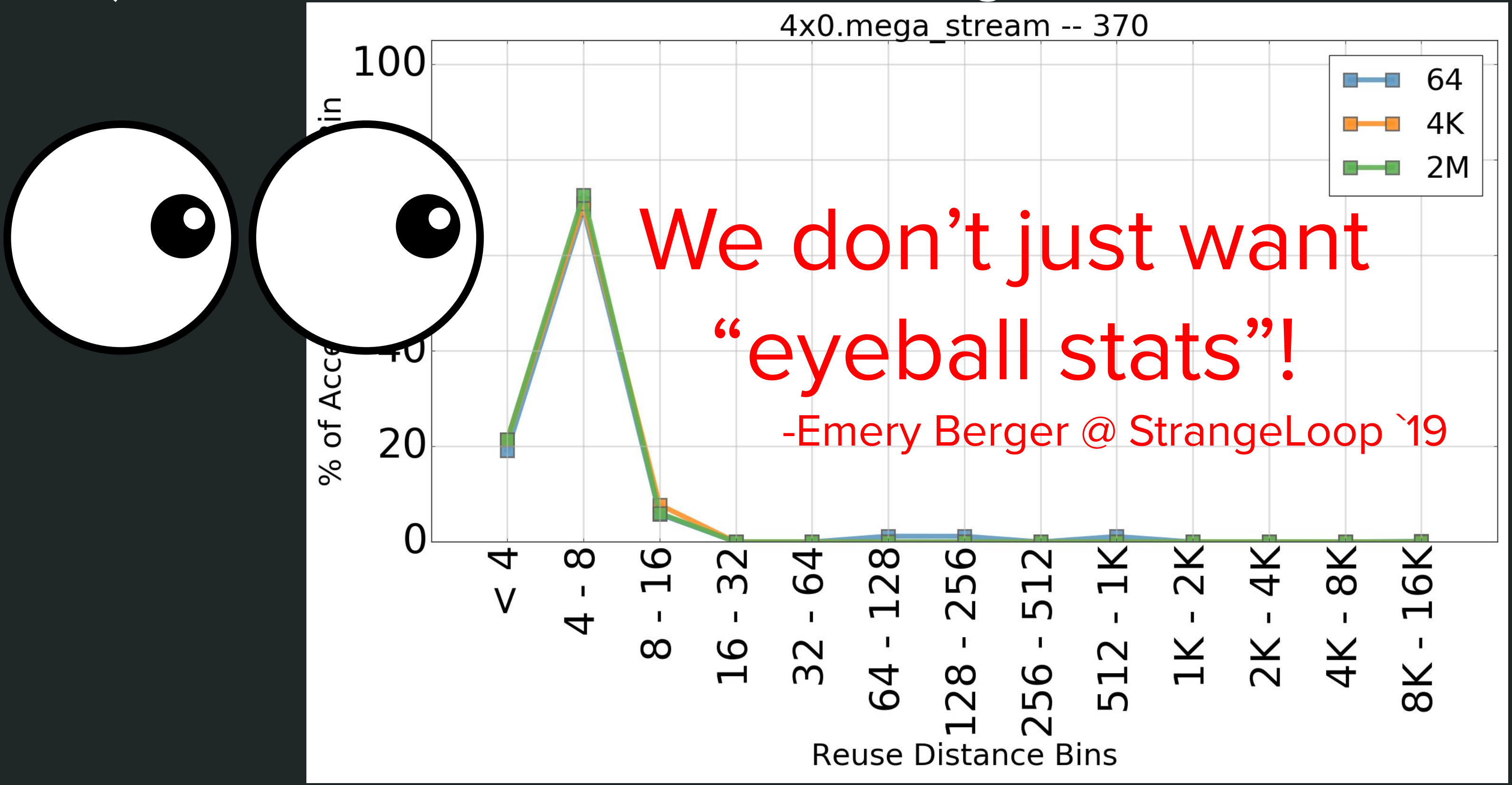
The Two Prototypical Behaviors

2) Mass remains the same across granularities



The Two Prototypical Behaviors

2) Mass remains the same across granularities



We don't just want "eyeball stats"!

-Emery Berger @ StrangeLoop `19

Earth Mover's Distance

minimize $$EMD = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} c_{ij}$$

Earth Mover's Distance

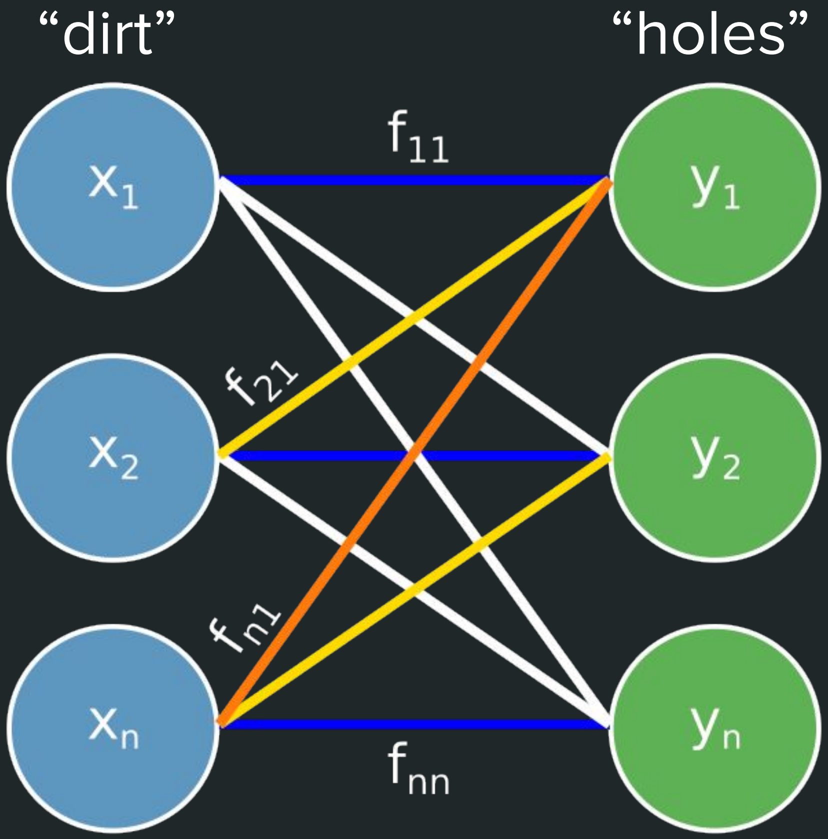
$$EMD = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} c_{ij}$$

$c_{ij} = 0$

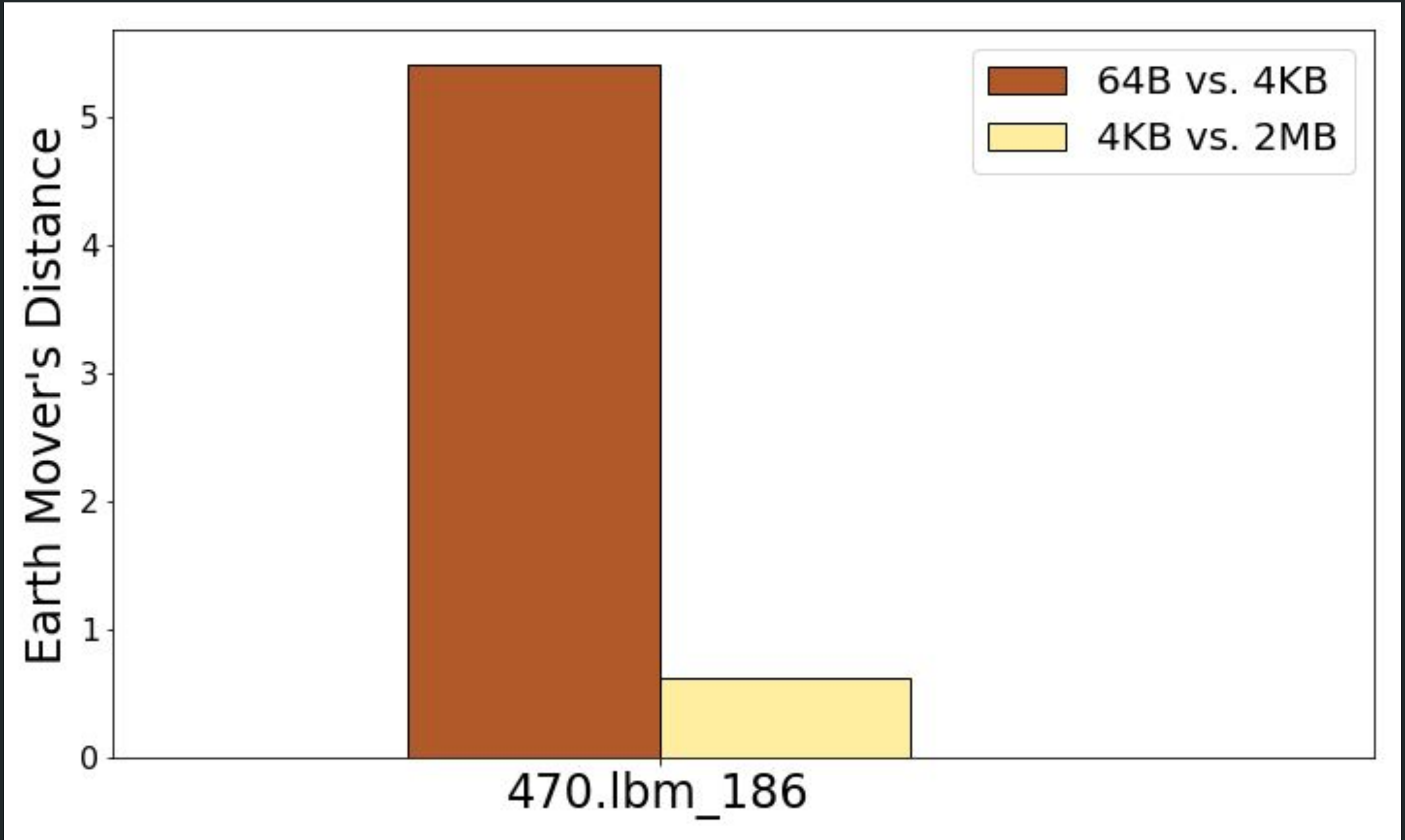
$c_{ij} = 1$

$c_{ij} = 2$

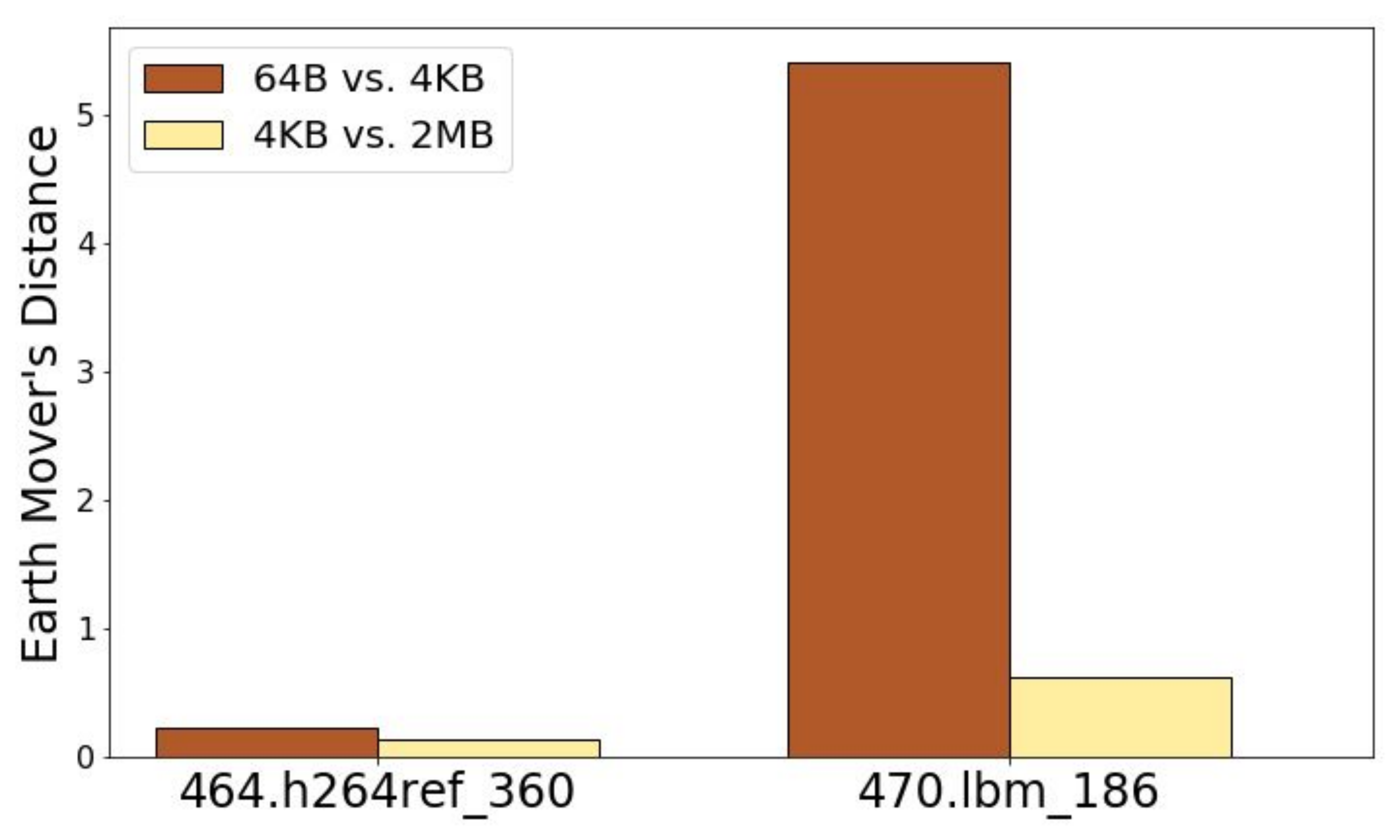
$c_{ij} = \infty$



Quantifying Spatial Locality with EMD



Quantifying Spatial Locality with EMD



Results

Spatially Dense (or not) Memory Accesses

Page Utilization

Memory Footprint =

$$S_{block\_granularity} \times N_{unique\_blocks}$$

$S_{block\_granularity}$ Size of reuse distance block granularity

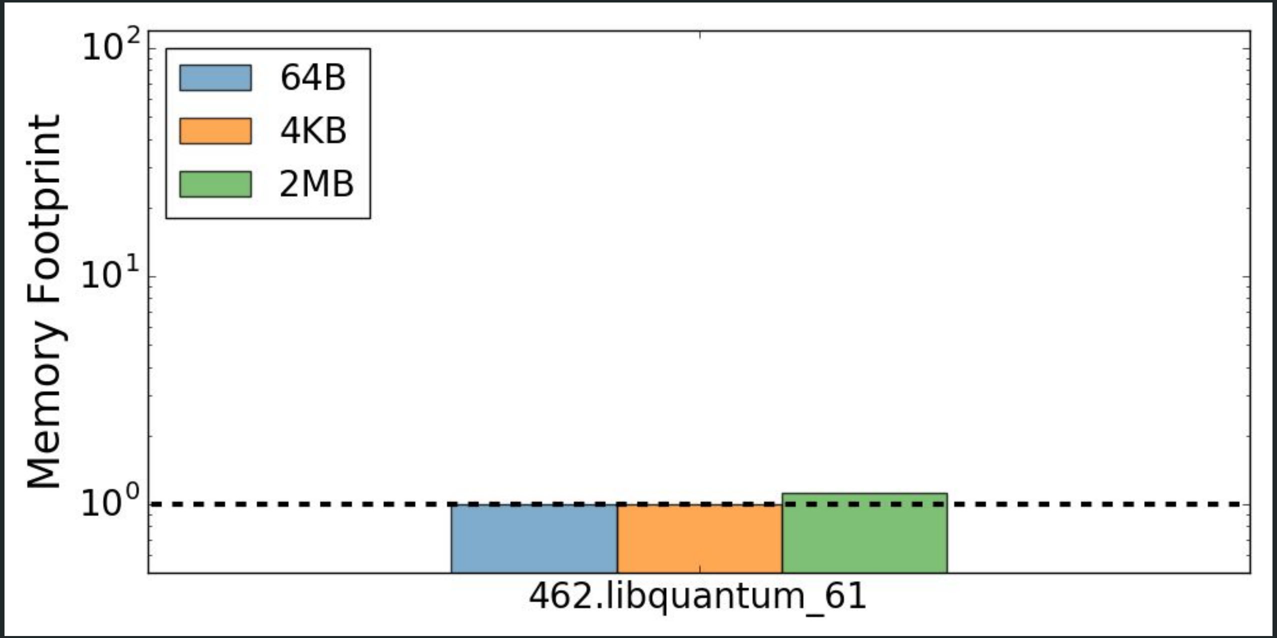
N_{unique_blocks} Number of unique blocks on stack after reuse distance analysis is complete

Memory Footprint Example

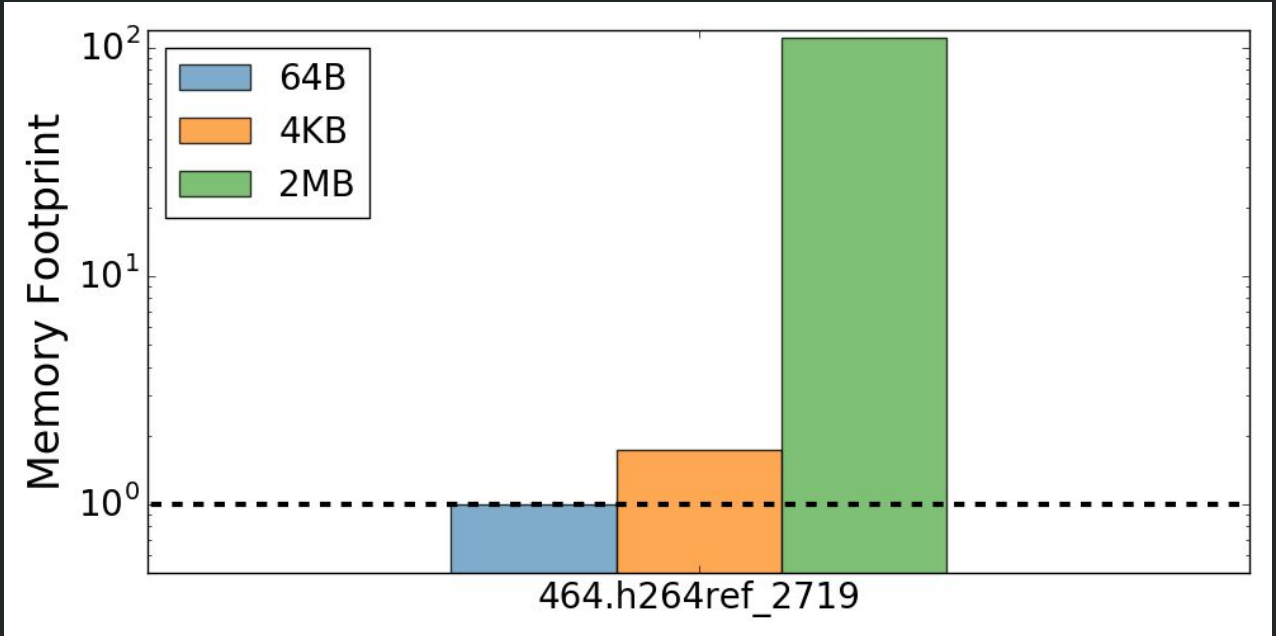
$$S_{block\_granularity} \times N_{unique\_blocks}$$

$$S_{block\_granularity} = 2MiB$$
$$N_{unique\_blocks} = 3$$
$$Memory\ Footprint = 6MiB$$

When is a page is fully utilized?



When isn’t a page is fully utilized?



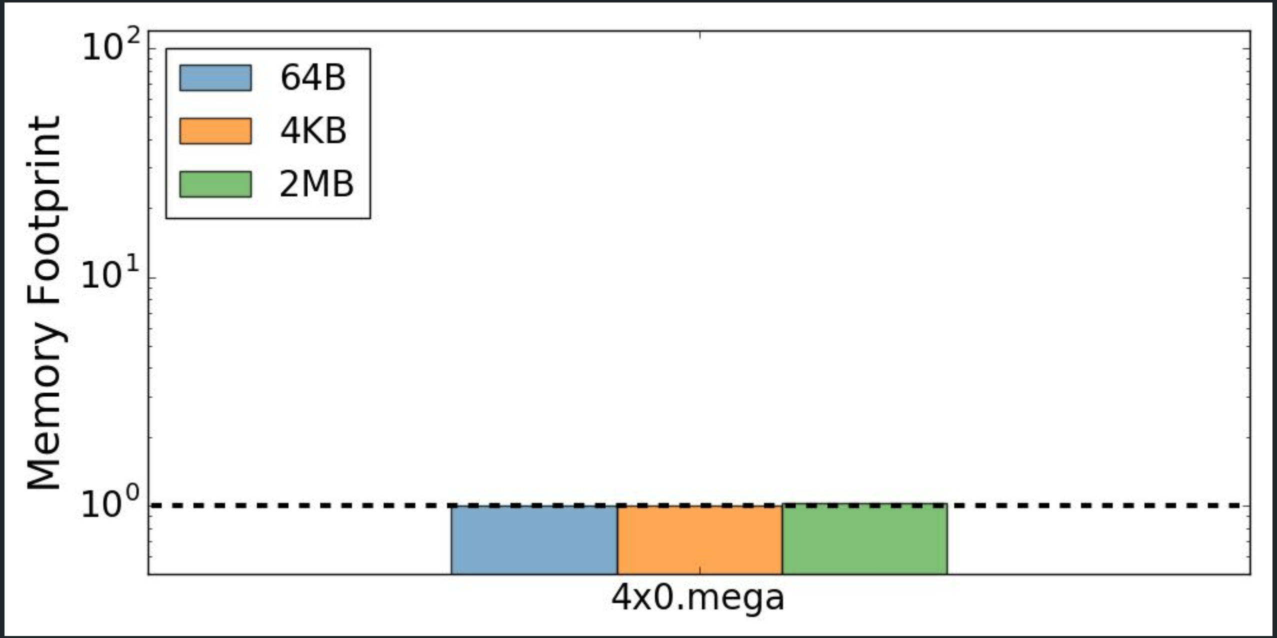
Results

Spatially Dense (or not) Memory Accesses

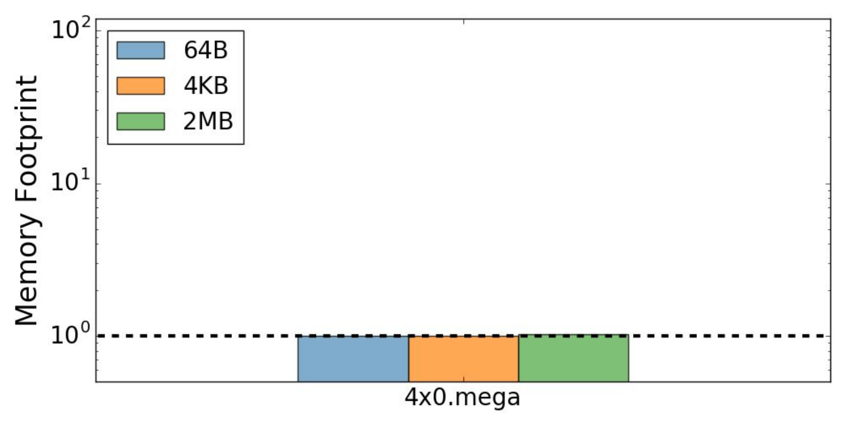
Page Utilization

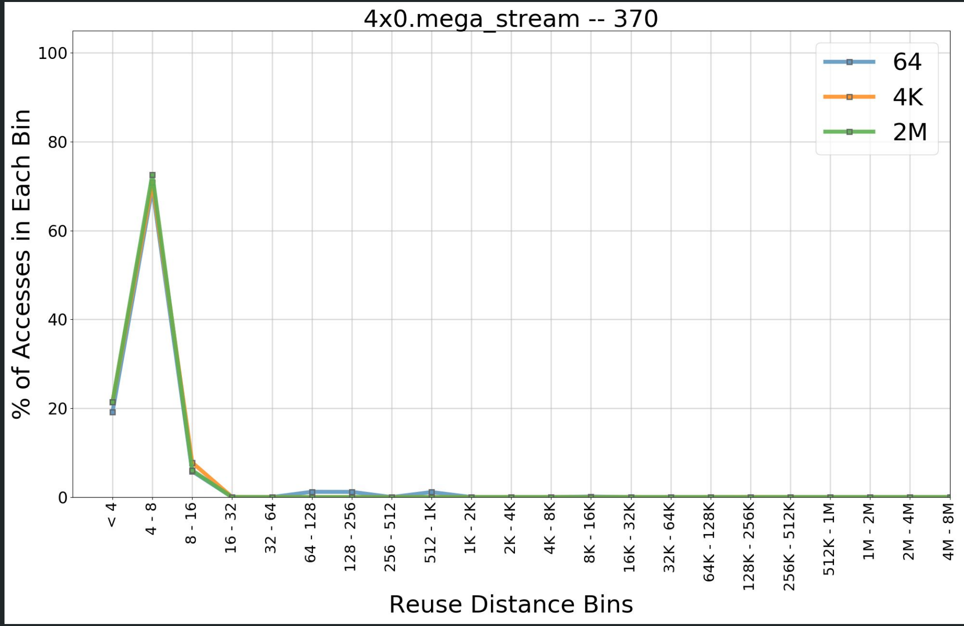
Data Layout Transformation (DLT)

Identify Opportunities for DLT



Identify Opportunities for DLT





Conclusion

- Infer both temporal and spatial data from reuse distance

- Quantify spatial locality with Earth Mover's distance

- Identify opportunities to reduce data movement
  AND
  Inform memory subsystem design/management

Contact Info

acabrera@wustl.edu

Backup

Outline

- Motivation
- Methodology
- Results & Discussion
- Conclusion

Motivational Stuff

Data movement is really painful and expensive

Spend 1000x less energy doing intense floating point op than reading/writing to DRAM

Paging hurts

- Think copy on write for even just half the page

Even cache lines aren’t fully utilized

- Cite JCB, chopping up sparse data one byte at a time

Motivational Stuff

It’s a heterogeneous world and memory is no exception

Heterogeneous compute systems

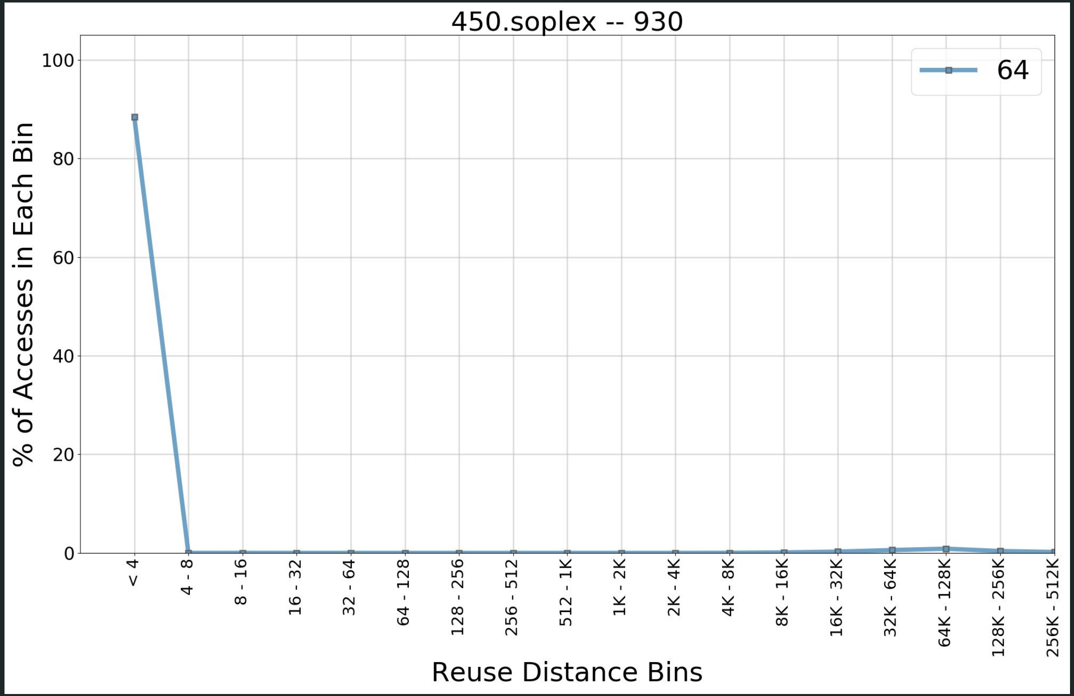
- E.g. CPUs, GPUs, FPGAs, DSPs, ASICs

have all sorts of memory types

- DRAM, SRAM, NVRAM, 3DRAM, HBM

The Two Prototypical Behaviors

2) Mass remains the same across granularities



How do we use it?

Use DynamoRIO to generate instruction trace around regions of interest (ROIs) identified in SPEC2006 benchmark

Perform reuse distance at different granularities

64B, 4KiB, and 2MiB

Earth Mover's Distance

$$X = x_1, \ldots, x_n$$
$$Y = y_1, \ldots, y_n$$

x_i

y_j

Earth Mover's Distance

minimize

$$EMD = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} c_{ij}$$

subject to

$$f_{ij} \geq 0$$

$$\sum_{j=1}^{n} f_{ij} = x_i, \quad x_i \in X$$

$$\sum_{i=1}^{n} f_{ij} = y_j, \quad y_j \in Y$$

Earth Mover's Distance

minimize $$EMD = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}c_{ij}$$

f_{ij}

c_{ij}

$$c_{ij} = j - i$$

Future Work

- Automate multi-spectral analysis on the fly for dynamic memory management