Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Memory System Performance is Messy

Pouya Esmaili-Dokht
and Victor Xirau

Barcelona Supercomputing Center

It's the Memory, Stupid!

25th of February 2026.

Collaborators: Three teams in BSC and Micron Technology



A Mess of Memory System Benchmarking, Simulation and Application Profiling

Pouya Esmaili-Dokht*†, Francesco Sgherzi*†, Valéria Soldera Girelli*†, Isaac Boixaderas*, Mariana Carmin*, Alireza Monemi*, Adrià Armejach*†, Estanislao Mercadal*, Germán Llort*, Petar Radojković*, Miquel Moreto*†, Judit Giménez*†, Xavier Martorell*†, Eduard Ayguadé*†, Jesus Labarta*†, Emanuele Confalonieri‡, Rishabh Dubey‡, Jason Adlard‡

Barcelona Supercomputing Center*, Unversitat Politecnica de Catalunya†, Micron Technology‡

{pouya.esmaili, francesco.sgherzi, valeria.soldera, isaac.boixaderas, mariana.carmin, alireza.monemi, adria.armejach}@bsc.es

{lau.mercadal, german.llort, petar.radojkovic, miquel.moreto, judit, xavier.martorell, eduard.ayguade, jesus.labarta}@bsc.es

{econfalo, rdubeya, jadlard}@micron.com

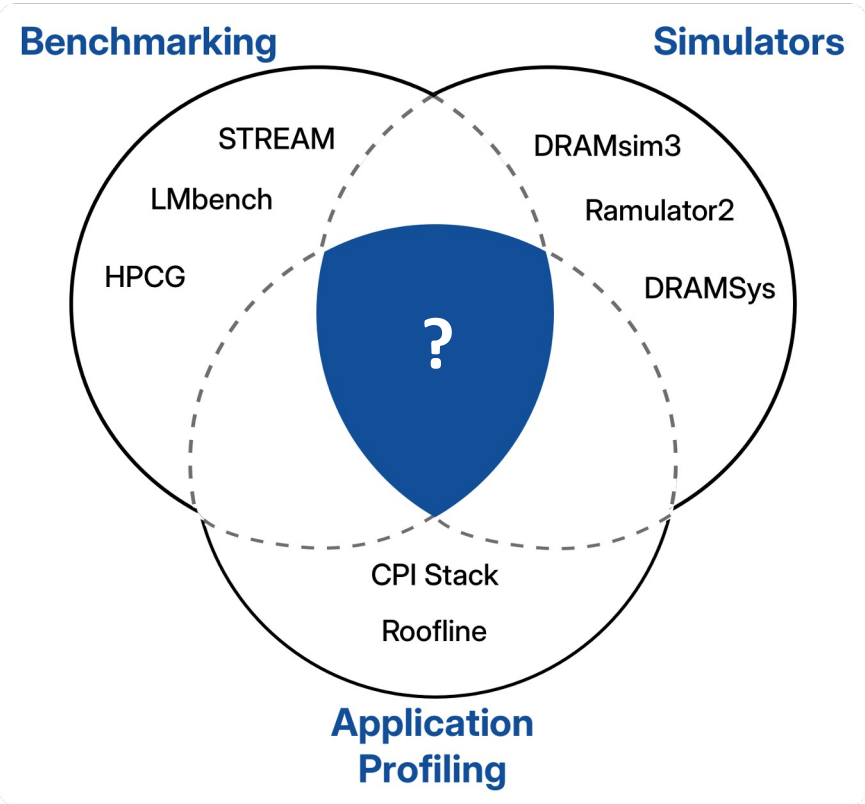
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

It's the Memory, Stupid!

Memory Stress (Mess) Study

- Opened a lot of discussion in the community
  - Best paper runner-up award at MICRO, 2024 🇺🇸
  - Award for most controversial truth-telling panel discussion at MEMSYS, 2024 🇺🇸
  - Invited talk at EPFL university, 2025 🇨🇭
  - International Workshop on Characterization and Modeling of Memory devices (ICWM2), 2025 🇮🇹
  - Workshop on Computer Architecture Modeling and Simulation (CAMS), 2025 🇰🇷
  - Mess tutorial at MICRO, 2025 🇰🇷
  - Now we are back 🇪🇸

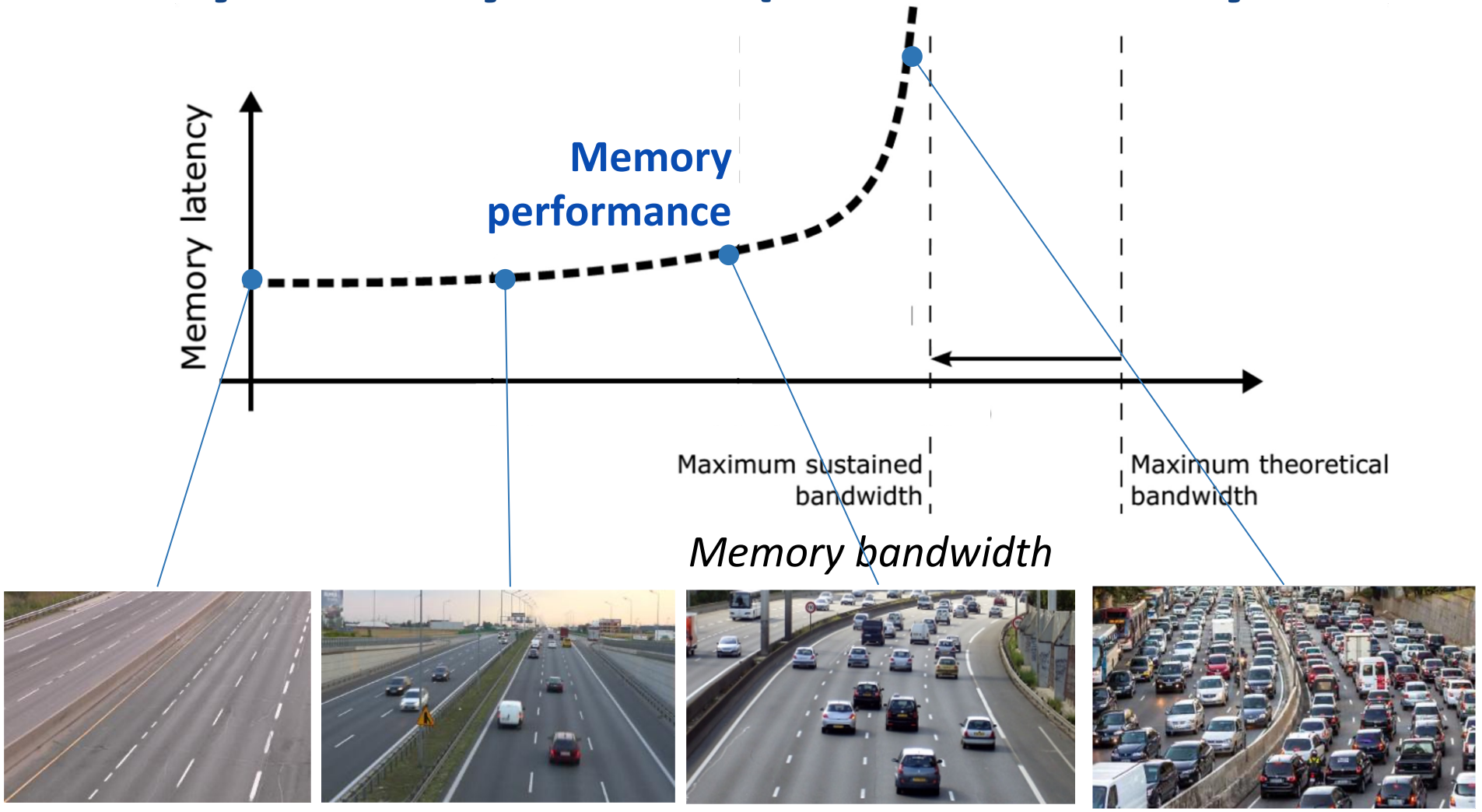
Decoupled memory system benchmarking, simulation and application profiling



Question:
Is there any *"common ground"*?

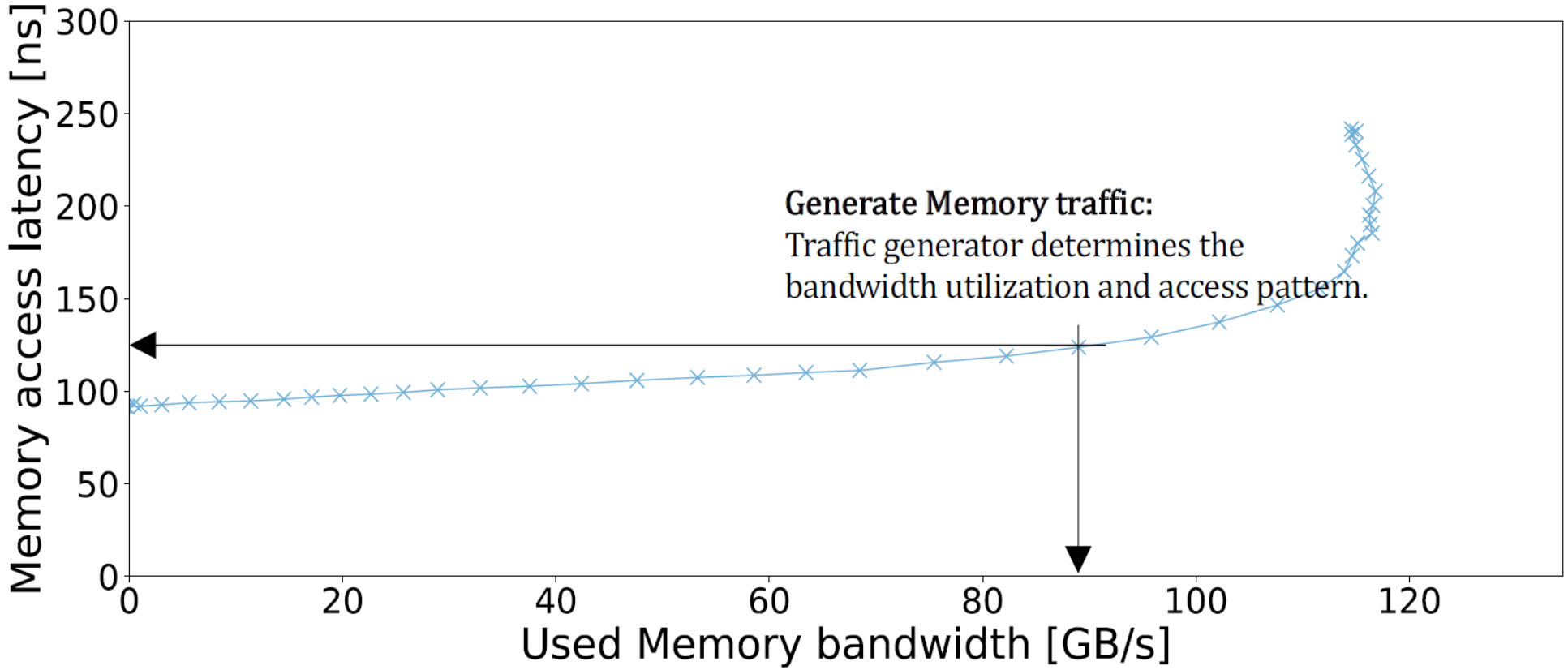
Is there any
"fundamental memory system performance"
that all each of these tools see through
a different lens/angle?

Memory latency = func(Used memory bandwidth)



B. L. Jacob. The memory system: *You can't avoid it, you can't ignore it, you can't fake it.* Synthesis Lectures on Computer Architecture, 2009.

Let’s make a benchmark that will “plot” the curve



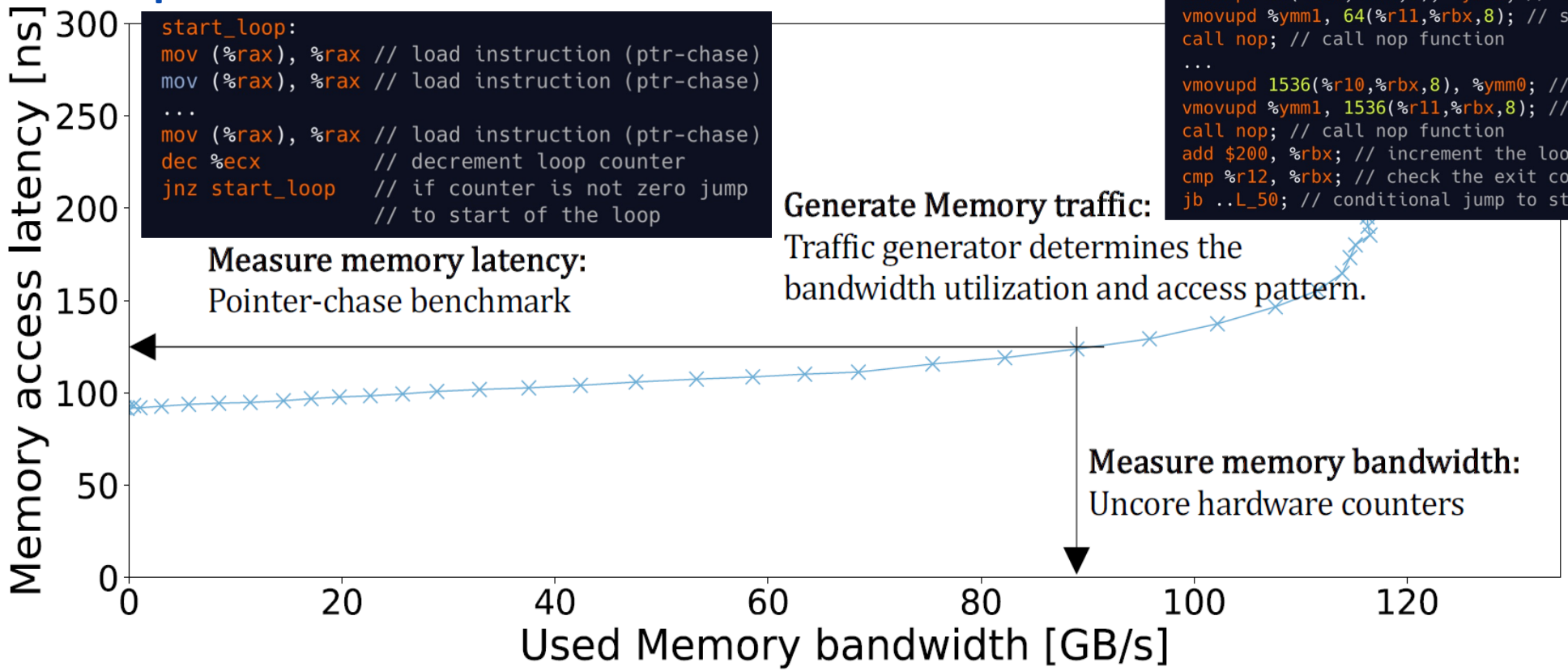
Mess benchmark code “snippet”

Traffic generator:

```
..L_50:
vmovupd (%r10,%rbx,8), %ymm0; // load instruction
vmovupd %ymm1, (%r11,%rbx,8); // store instruction
vmovupd 32(%r10,%rbx,8), %ymm0; // load instruction
vmovupd %ymm1, 32(%r11,%rbx,8); // store instruction
vmovupd 64(%r10,%rbx,8), %ymm0; // load instruction
vmovupd %ymm1, 64(%r11,%rbx,8); // store instruction
call nop; // call nop function
...
vmovupd 1536(%r10,%rbx,8), %ymm0; // load instruction
vmovupd %ymm1, 1536(%r11,%rbx,8); // store instruction
call nop; // call nop function
add $200, %rbx; // increment the loop counter
cmp %r12, %rbx; // check the exit condition
jb ..L_50; // conditional jump to start of the loop
```

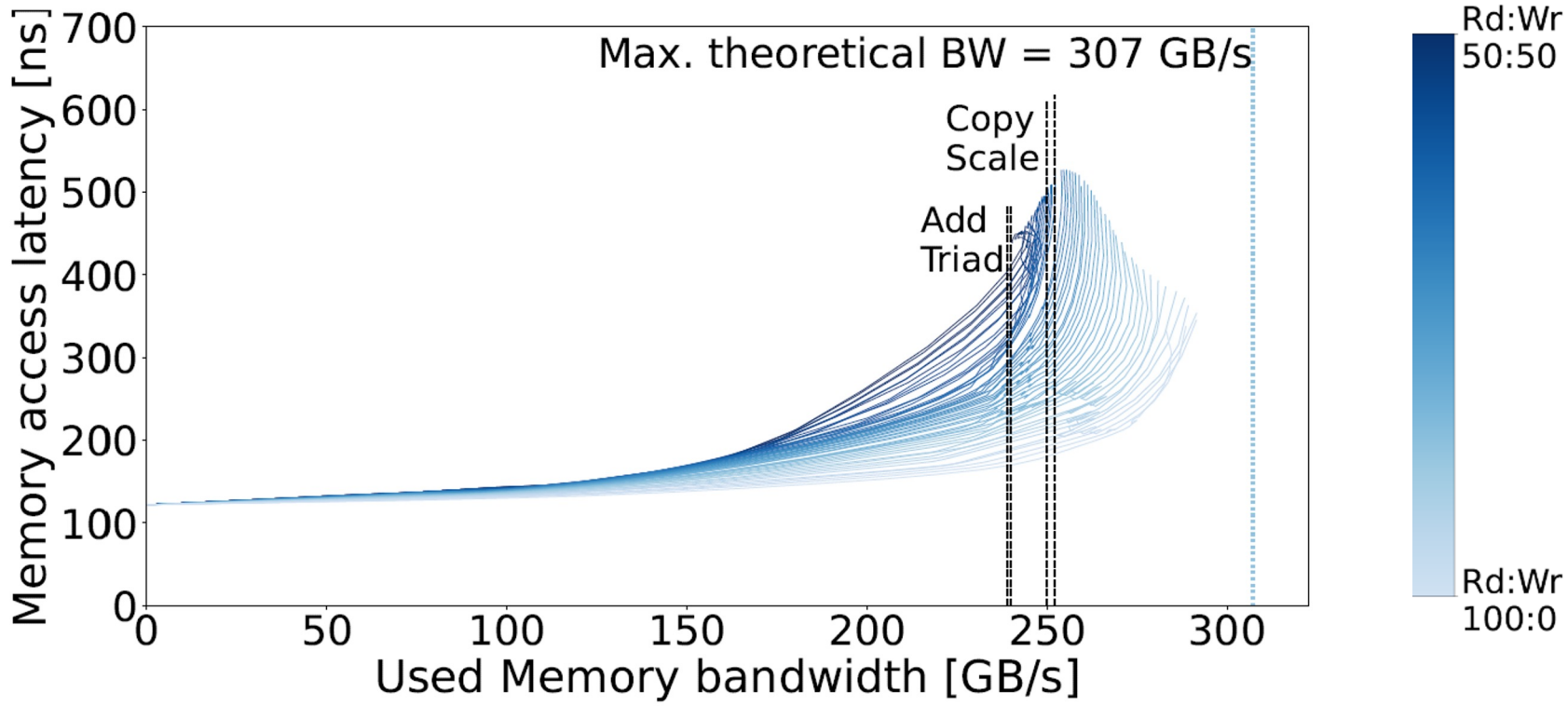
ptr-chase:

```
start_loop:
mov (%rax), %rax // load instruction (ptr-chase)
mov (%rax), %rax // load instruction (ptr-chase)
...
mov (%rax), %rax // load instruction (ptr-chase)
dec %ecx         // decrement loop counter
jnz start_loop   // if counter is not zero jump
                 // to start of the loop
```



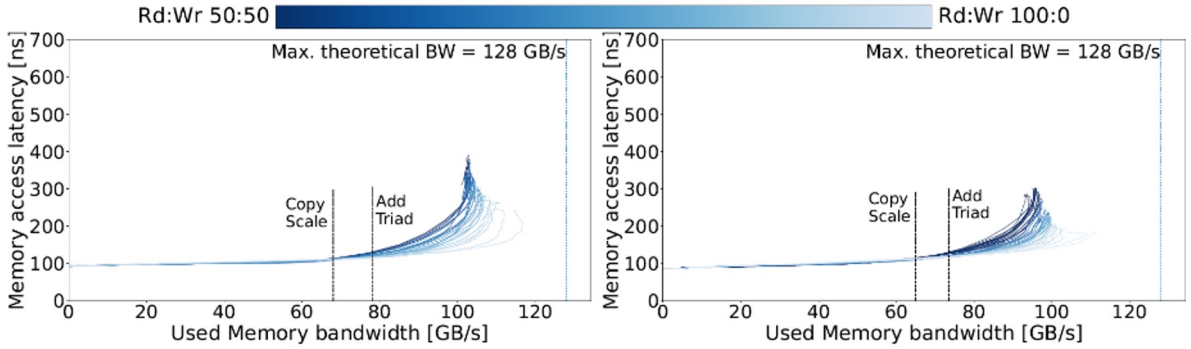
Memory stress (Mess) benchmark

- Connecting the dots



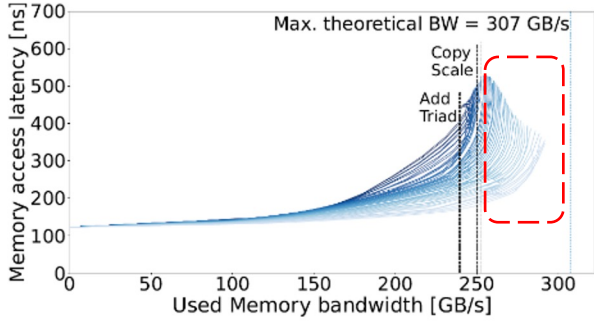
Amazon Graviton3 with 8× DDR5-4800.

Mess benchmark: Actual platforms

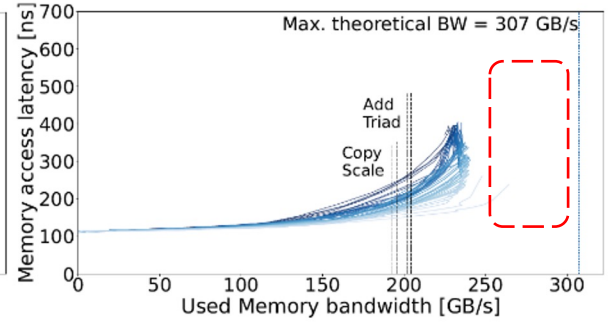


Intel Skylake with 6× DDR4-2666.

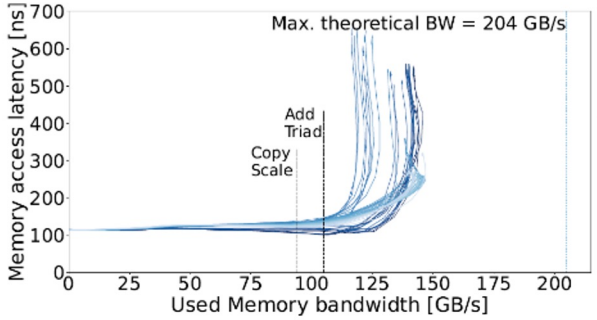
Intel Cascade Lake with 6× DDR4-2666.



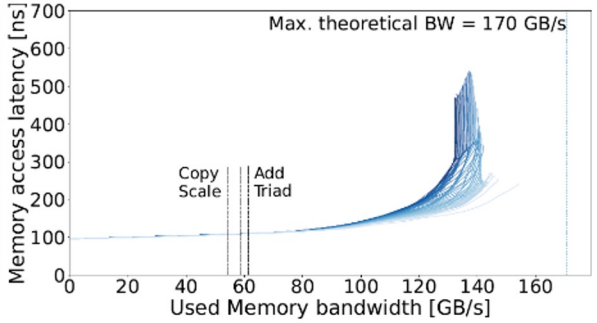
Amazon Graviton3 with 8× DDR5-4800.



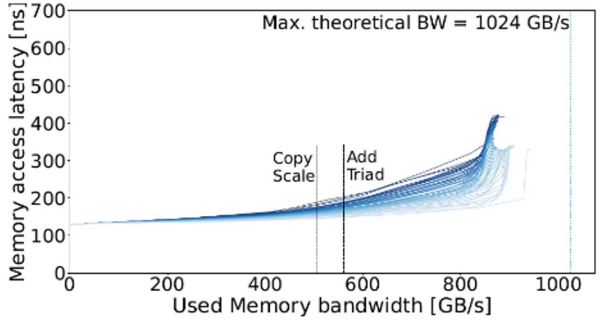
Intel Sapphire Rapids with 8× DDR5-4800.



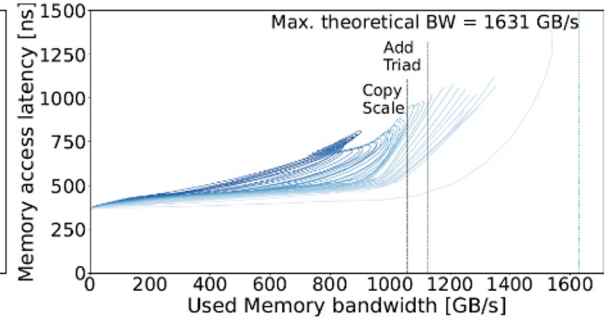
AMD Zen2 with 8× DDR4-3200



IBM Power 9 with 8× DDR4-2666.



Fujitsu A64FX with 4× HBM2.



NVIDIA H100 with 4× HBM2E.

- Memory system performance is very, very different

It's the Memory, Stupid!

Intel Xeon 6980P
12 x MRDIMM 8800 MT/s

- **CPU**
  - Intel Xeon 6980P (Granite Rapids)
  - 128 cores @3.2GHz
- **Memory (per socket)**
  - 12 x DDR5 MRDIMMs 8800 MT/s
  - 64 GB per DIMM
- **System configuration**
  - Red Hat Enterprise Linux 9.5
  - Hyperthreading: Off
  - Latency Optimized Mode: Enabled
  - Energy Performance Preference (EPP): 0
  - Energy Performance Bias (EPB): 0
- **Thanks to Intel for granting access to the system**

Downloa

Essentials

Product Collection	Intel® Xeon® 6 processors
Code Name	Products formerly Granite Rapids

CPU Specifications

Total Cores	128
Total Threads	256
Max Turbo Frequency	3.9 GHz
All Core Turbo Frequency	3.2 GHz
Processor Base Frequency	2 GHz
Cache	504 MB

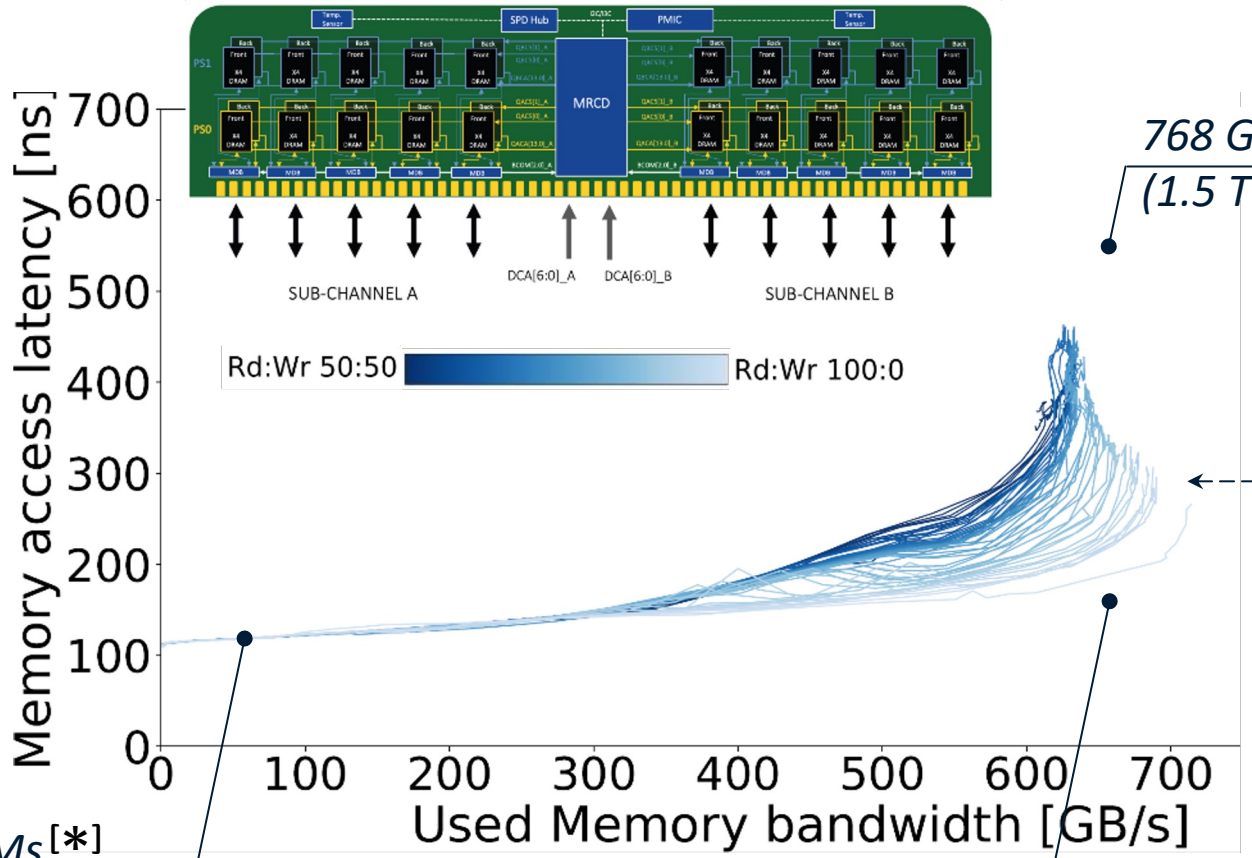
Memory Specifications

Maximum Memory Speed	8800 MT/s
Max # of Memory Channels	12

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Single CPU socket: 12 x MRDIMM 8800 MT/s

Unprecedented memory system performance



768 GB capacity per socket (1.5 TB per server)

+75% bandwidth utilization: Sustained/Max theoretical

Comparable to or better than servers with DDR4/5 RDIMMs [*]

Similar latency to DDR4/5 RDIMMs [*]

2.5x sustained bandwidth over 8xDDR5-4800 RDIMMs on Amazon Graviton 3 and 4th Gen Intel Xeon Intel [*]

Disclaimer: Results may vary

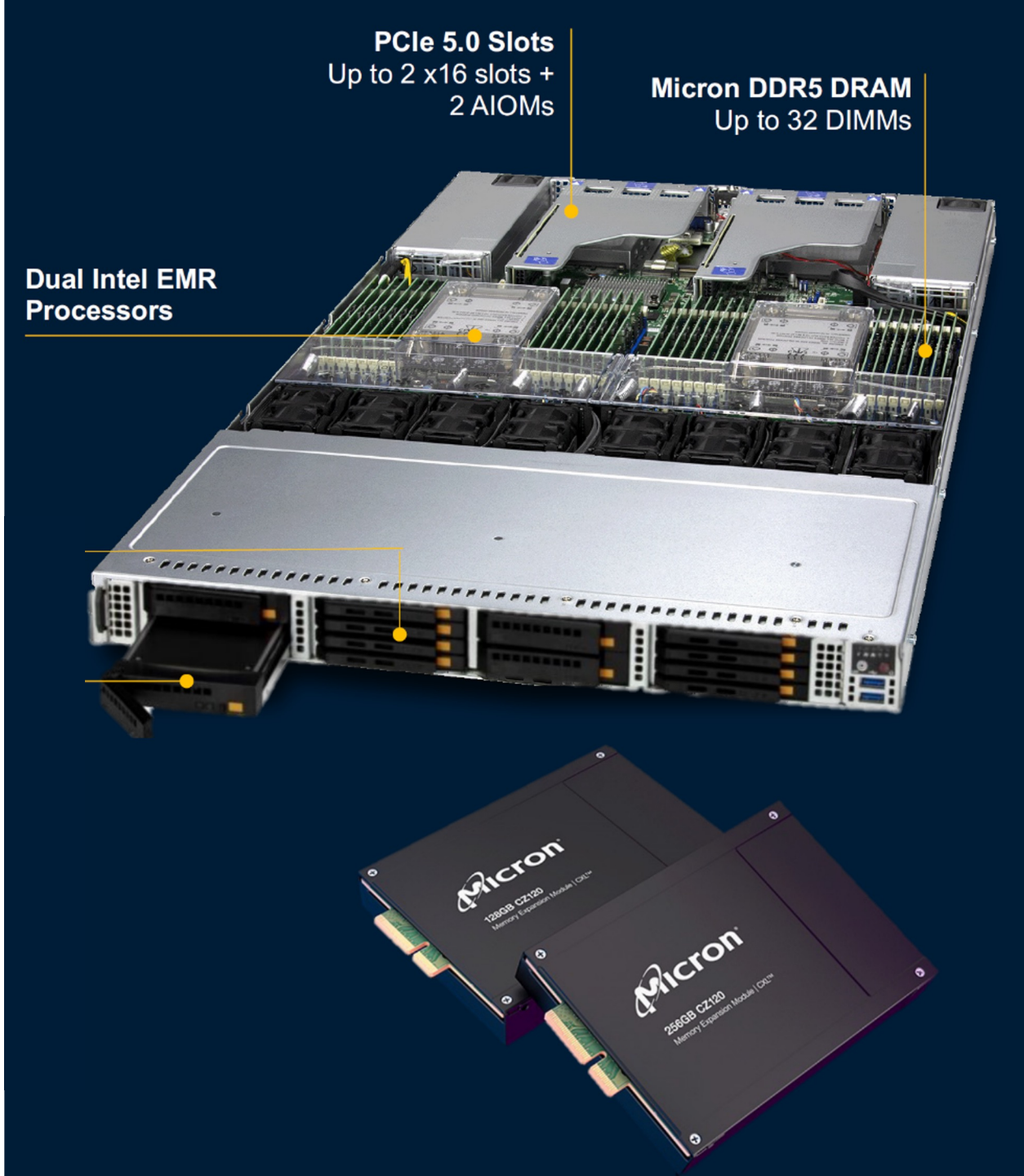
[*] Esmaili-Dokht et al. *A Mess of Memory System Benchmarking, Simulation and Application Profiling.* MICRO-2024.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CXL memory expanders

Supermicro server:

- **CPU**
  - 2x Intel Xeon Platinum (Emerald Rapids)
  - 48 cores @2.3GHz
- **Memory (per socket)**
  - 8 x DDR5-4800 RDIMMs (16 GB per DIMM)
  - 2 x Micron CZ120 CXL memory expanders 256 GB per device



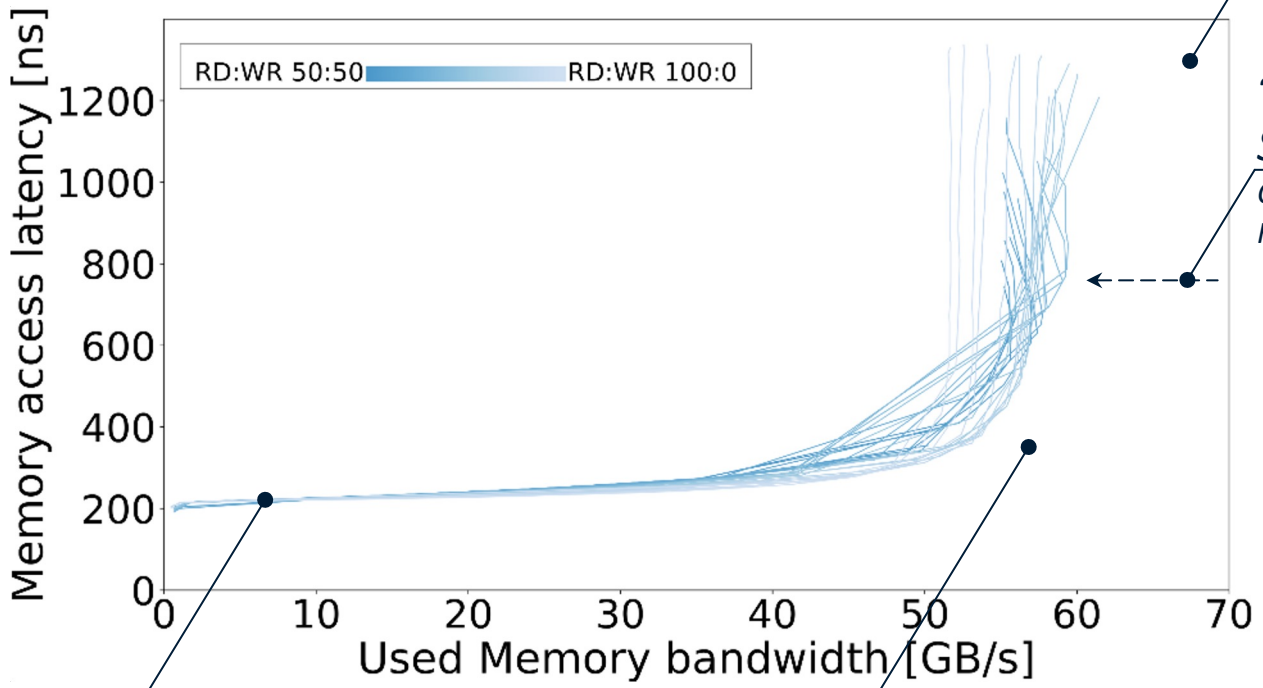
Thanks to
Micron Technology Enablement Program
https://www.micron.com/partners/technology-enablement-programs/cxl-tep-resources

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CXL memory expanders

Expand your directly-attached DIMMs

2 x Micron CZ120 (Gen 1)
CXL memory expanders
Interleaved



+0.5TB capacity

~65% bandwidth utilization: Sustained/Max theoretical

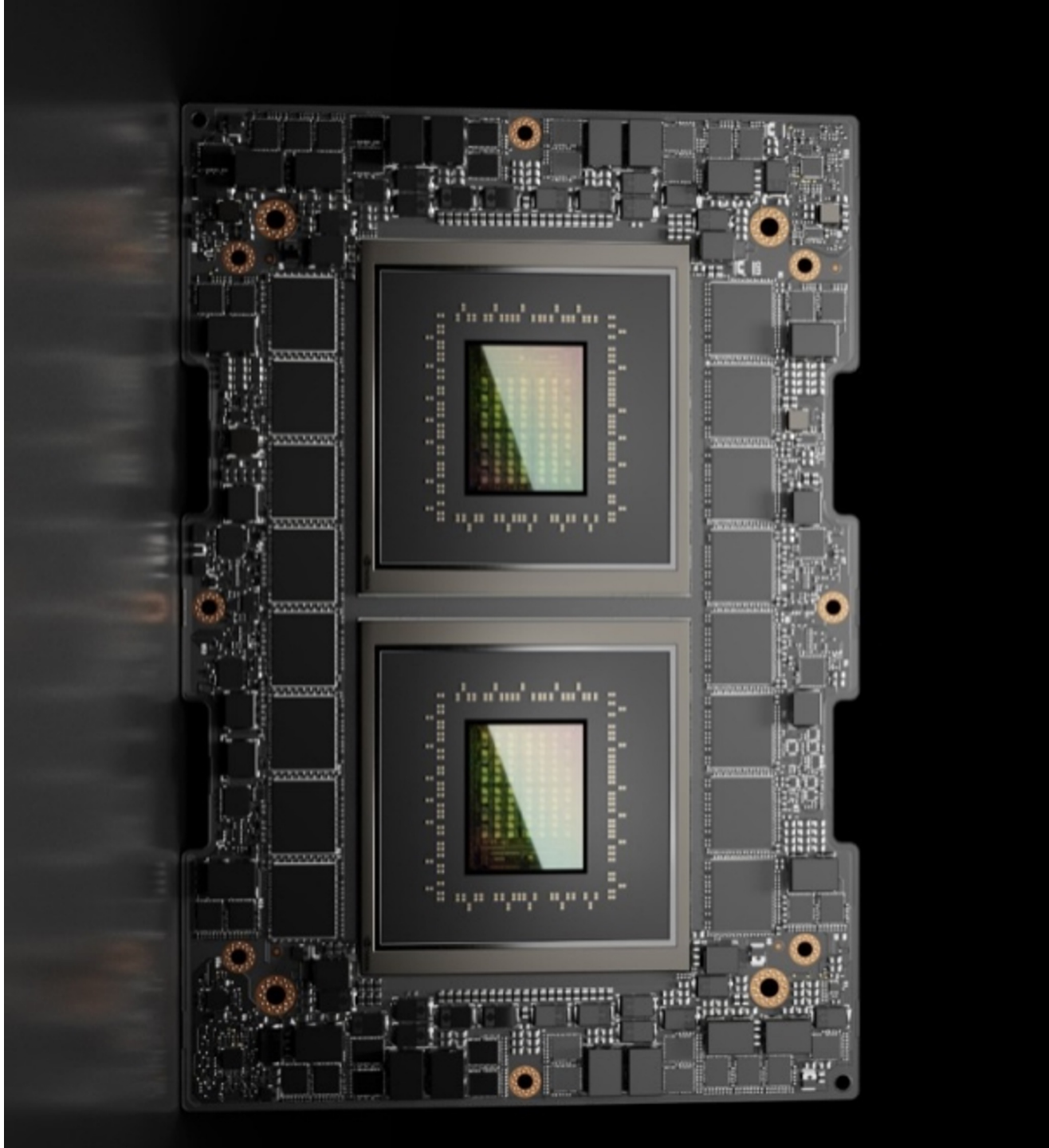
Comparable to directly attached DIMMs in HPC servers

Latency comparable to the remote-socket memory access

+30% of the sustained memory bandwidth (on the top of 8xDDR5-4800)

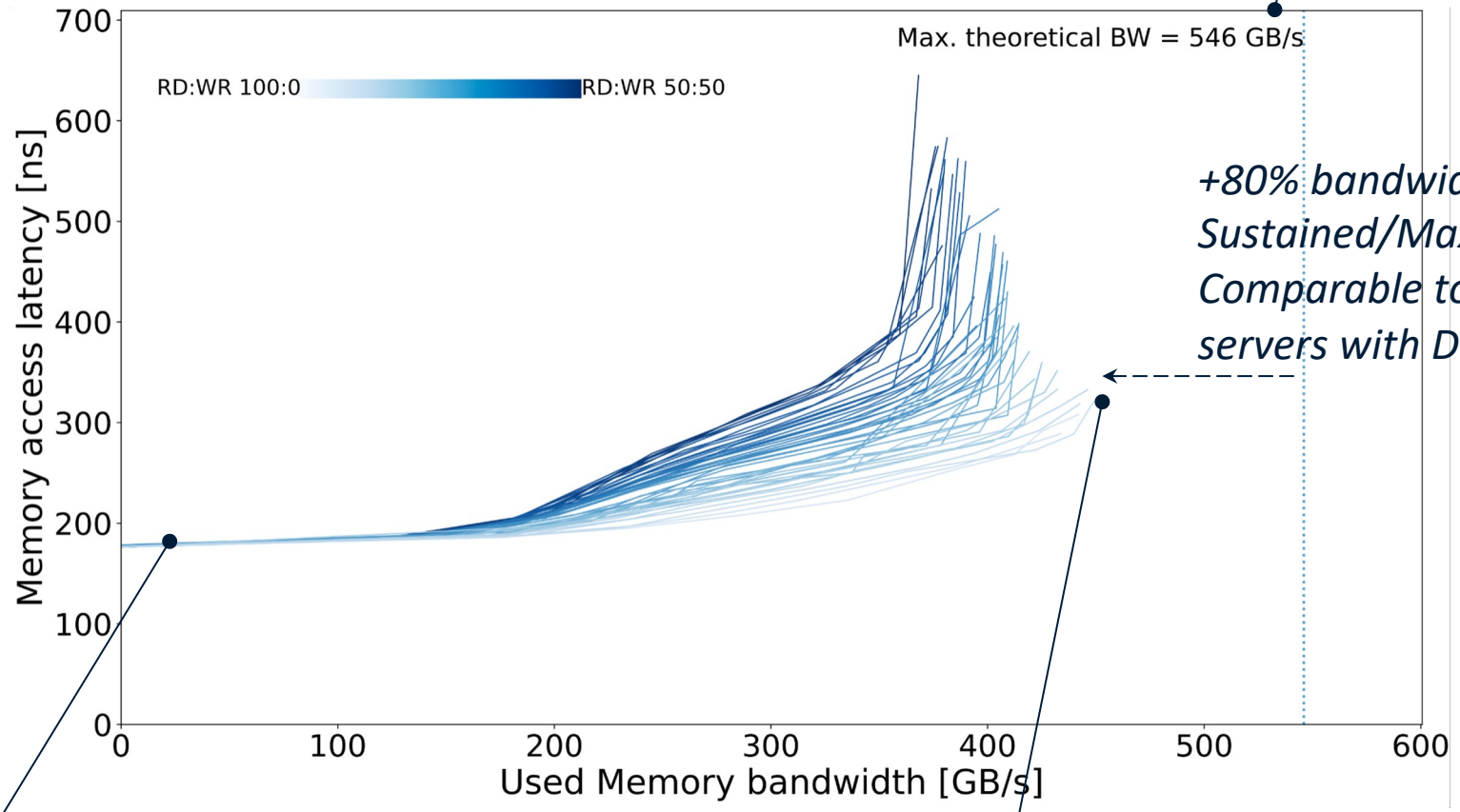
NVIDIA Grace CPU LPDDR5X

- **CPU**
  - 2x NVIDIA Grace 72C 3.1GHz
  - 72 cores @3.3GHz
- **Memory (per socket)**
  - 2x Die 4266MHz LPDDR5
  - 120 GB per Die



NVIDIA Grace CPU LPDDR5X

120 GB capacity per socket (240 GB per server)



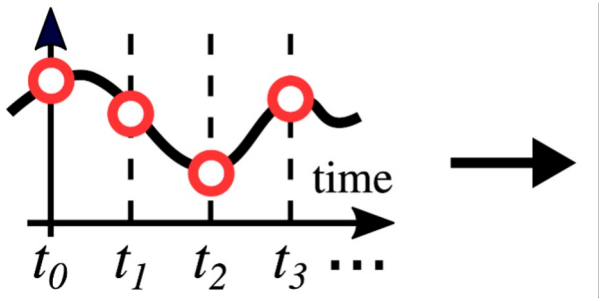
+80% bandwidth utilization: Sustained/Max theoretical Comparable to or better than servers with DDR4/5.

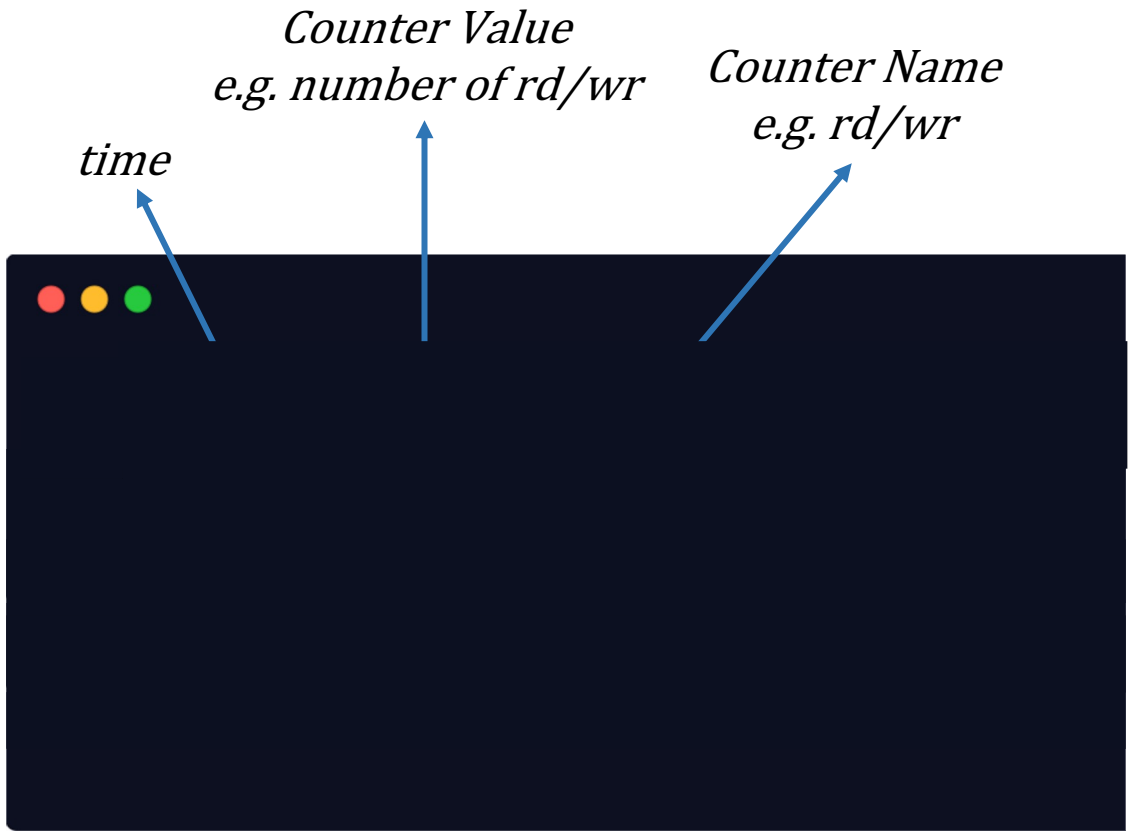
Latency is higher than intel platform

~75% higher sustained bandwidth compared to Amazon Graviton 3 and 4th Gen Intel Xeon Intel with DDR5

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

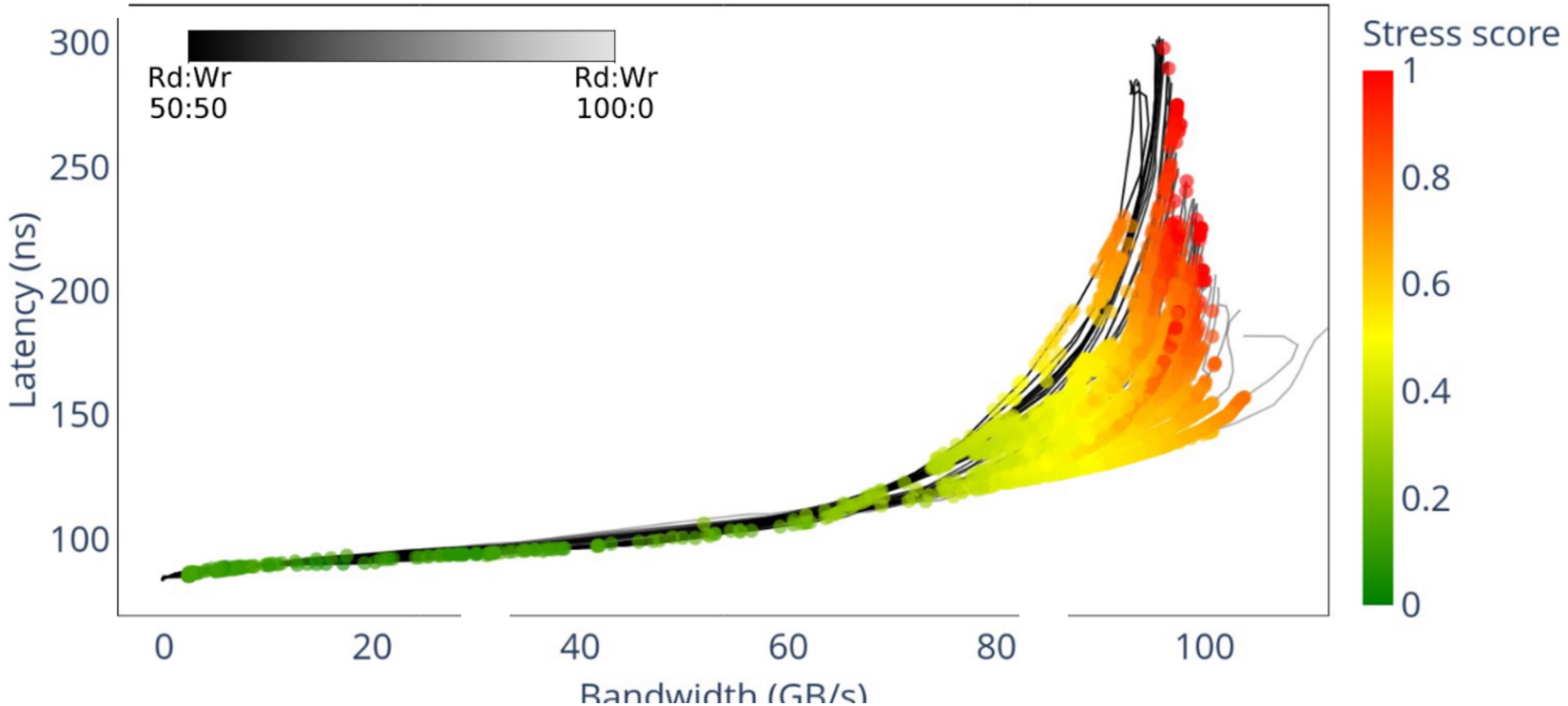
Application profiling





Putting it all together: System & Application profiling

- Application performance in the context of the system performance



24-core Intel Skylake CPU with 6x DDR4-2666.
HPCG. Sampling frequency: 10 ms.

As simple as…

- **Get memory curves** of your favorite server – Mess Benchmark
  - https://github.com/bsc-mem/Mess-2.0
  - Or use the available already collected curves in https://github.com/bsc-mem/Mess-Results
  - For this course we'll also provide some in https://memory.bsc.es/itms/materials
- **Profile your app** (memory read and write traffic)
  - Use whatever tool you like: PAPI, perf, Extrae, LIKWID…
  - For this demo we'll use perf
- **Visualize** your results
  - Use the Plotter in the Benchmark: https://github.com/bsc-mem/Mess-2.0/wiki/Plotter-Parser
  - Or use this course's visualizers: https://memory.bsc.es/itms/materials#visualizers
- You know where your application located on the memory curves

Demo

Follow along!

rb.gy/s6hfzr

- We are going to be showing the following apps
  - STREAM
    - https://www.cs.virginia.edu/stream/FTP/Code/
  - Google Multichase
    - https://github.com/google/multichase

- On the following Systems
  - **MN5-GPP**
    - Intel Xeon Platinum 8480+ 56C 2GHz (SPR) w/ 8x **16GB** DDR5-4800
  - **MN5 GPP-HighMem**
    - Intel Xeon Platinum 8480+ 56C 2GHz (SPR) w/ 8x **64GB** DDR5-4800
  - **MN5 GPP-HBM**
    - Intel Xeon CPU Max 9480 56C 1.9GHz w/ 4x 16 GB HBM2-3200
  - **MN5 NGrace**
    - NVIDIA Grace 72C 3.1GHz w/ 1x 120GB LPDDR5-4266
  - **MN5 ACC**
    - NVIDIA Hopper H100 w/ 64GB HBM2

- Wide range of systems, same approach

Follow along!

rb.gy/s6hfzr

Let’s see it…

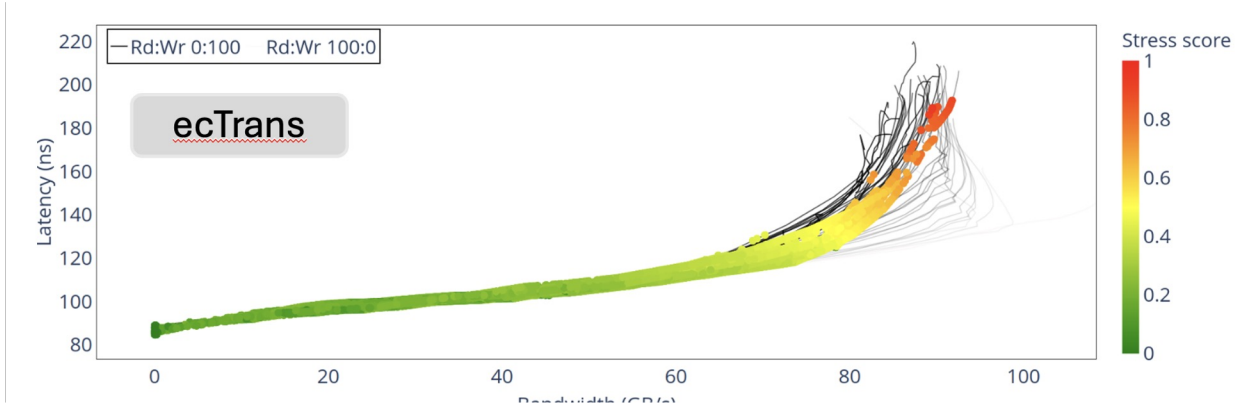
Do you start to “have a feeling”?

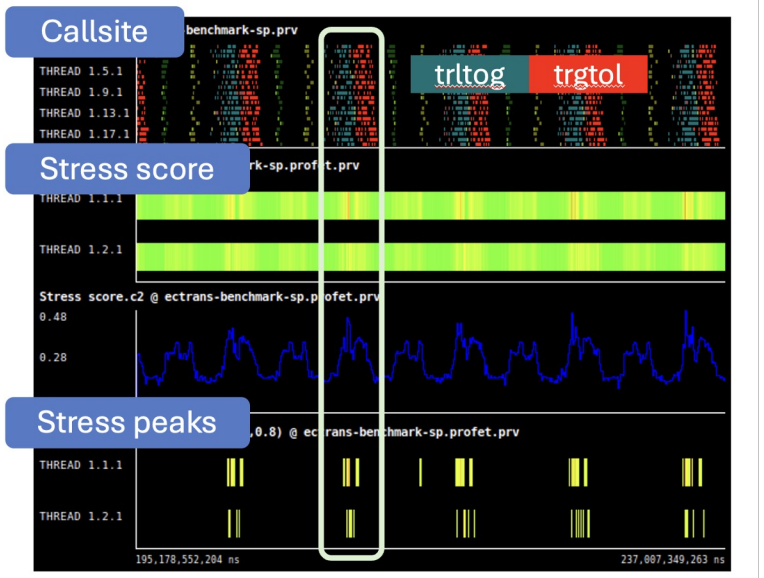
- If you do …
  - This course is already a success

The same hammer to a different nail

- The same principle applies to
  - DDR5 + HBM
  - DDR5 + CXL expanders
  - Past:
    - Intel KNL: DDR4 + MCDRAM
    - Intel Cascade Lake: DDR4 + Optane
  - UnifiedBus [by Huawei]

When combined with Extrae + Paraver…





Memory stress @ ectrans-benchmark-sp.profet.prv

	[0.00..0.10)	[0.10..0.20)	[0.20..0.30)	[0.30..0.40)	[0.40..0.50)	[0.50..0.60)	[0.60..0.70)	[0.70..0.80)	[0.80..0.90)	[0.90..1.00]
THREAD 1.1.1	0.46 %	22.92 %	45.57 %	28.06 %	2.50 %	0.34 %	0.03 %	0.01 %	0.11 %	0.00 %
THREAD 1.2.1	0.38 %	20.48 %	55.90 %	22.57 %	0.53 %	0.03 %	0.00 %	0.02 %	0.09 %	0.00 %
Average	0.42 %	21.70 %	50.73 %	25.31 %	1.52 %	0.18 %	0.02 %	0.01 %	0.10 %	0.00 %
Avg/Max	0.92	0.95	0.91	0.90	0.61	0.54	0.51	0.77	0.94	0.60

When combined with Extrae + Paraver…

- Easily identify where memory stress is highest and correlate with sources

- For details on how to get started
  - Installation guide: https://memory.bsc.es/tools/mess-paraver/installation-guide
  - https://github.com/bsc-mem/Mess-Paraver
  - https://tools.bsc.es/
  - Check an example of this on the paper: https://arxiv.org/pdf/2405.10170

- The BSC Tools team has a lot of training material and participates in many performance analysis and tuning workshops.
  - If you are interested in this work, please contact them: tools@bsc.es

https://www.bsc.es/es/education/training/other-training/vi-hps-tuning-workshop

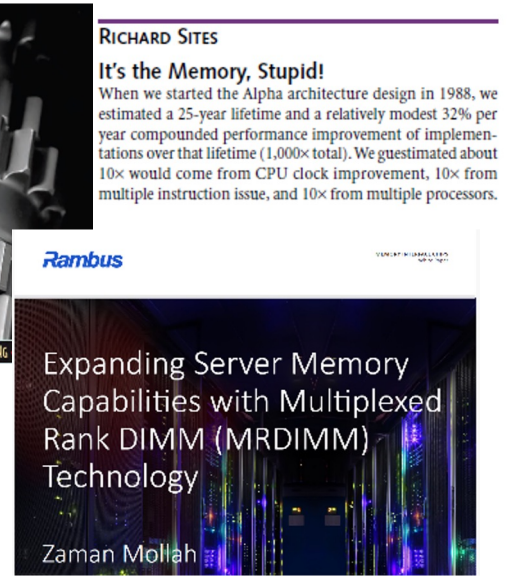
Explore more

1. Mess
   a. Mess Paper [link]
   b. Mess 2.0 [link]
   c. Mess Website [link]
   d. Contact: Pouya, Victor, mess@bsc.es

2. Tools and benchmarks
   a. Intel MLC [link] and STREAM [link]
   b. Roofline [link]
   c. perf [link], LIKWID [link], Intel Vtune [link] and PCM [link].
   d. Light-weight visualizer [link]
   e. Extrae and Paraver [link]

3. Reading materials
   a. Hitting the Memory Wall…, 1995. [link]
   b. Another Trip to the Wall…, 2015. [link]
      i. Contact: Petar
   c. B. L. Jacob. The memory system: *You can't avoid it, you can't ignore it, you can't fake it.* [link]
   d. PROFET paper [link]
      i. Contact: Mariana
   e. RAMBUS ISCA tutorials [link]
   f. RAMBUS white paper on MRDIMM [link]

4. Materials [link]



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

It's the Memory, Stupid!

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you!

pouya.esmaili@bsc.es
victor.xirau@bsc.es
mess@bsc.es