Improving Reliability and Performance of Datacenter Systems via Coherence

Adarsh Patil

3rd Year PhD student

THE UNIVERSITY *of* EDINBURGH
informatics

arm

Outline

Improving Reliability and Performance

Dvé: Coherent Replication for DRAMs *(ISCA 2021)*

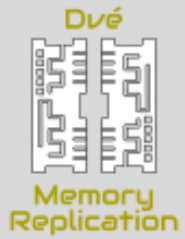
Coherent Disaggregated Shared Memory for FaaS *(WIP)*

Outline

Improving Reliability and Performance

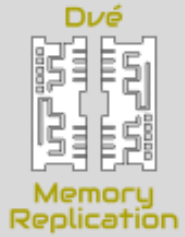
Dvé: Coherent Replication for DRAMs *(ISCA 2021)*

Coherent Disaggregated Shared Memory for FaaS *(WIP)*



Dvé: Outline

- Reliability benefits
- Performance gains
- On-demand Reliability



Outline

- Reliability benefits
- Performance gains
- On-demand Reliability

Dvé
Memory Replication

Increasing DRAM Faults

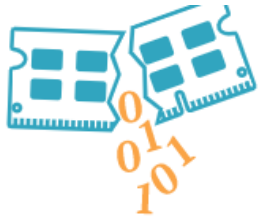
Bloomberg

Markets

How One Piece of Hardware Took Down a $6 Trillion Stock Market

By Gearoid Reidy, Shoko Oda, Min Jeong Lee, and Toshiro Hasegawa
2 October 2020, 10:47 BST *Updated on 5 October 2020, 01:48 BST*

That all changed on Thursday, when a piece of hardware called the No. 1 shared disk device, one of two square-shaped data-storage boxes, detected a memory error. These devices store management data used across the servers, and distribute information such as commands and ID and password combinations for terminals that monitor trades.



RAMBleed

Reading Bits in Memory Without Accessing Them

RAMBleed is a side-channel attack that enables an attacker to read out physical memory belonging to other processes. The implications of violating

VUSec

ECCPLOIT: ECC MEMORY VULNERABLE TO ROWHAMMER ATTACKS AFTER ALL

Where many people thought that high-end servers were safe from the (unpatchable) Rowhammer bitflip

IEEE SPECTRUM

Hardware

DRAM's Damning Defects—and How They Cripple Computers

By **Ioan Stefanovici, Andy Hwang and Bianca Schroeder**
Posted 23 Nov 2015 | 16:00 GMT

COMPUTERWORLD UNITED KINGDOM

NEWS

Google: DRAM error rates vastly higher than previously thought

PCs will likely require error correction code in the future due to DRAM issues

By Lucas Mearian
Senior Reporter, Computerworld | 8 OCTOBER 2009 23:51 GMT

ZDNet MENU UK

DRAM error rates: Nightmare on DIMM street

A two-and-a-half year study of DRAM on 10s of thousands Google servers found DIMM error rates are hundreds to thousands of times higher than thought -- a mean of 3,751 correctable errors per DIMM per year.This is the world's first large-scale study of RAM errors in the field.

By Robin Harris for Storage Bits | October 4, 2009 -- 22:04 GMT (23:04 BST) | Topic: Hardware

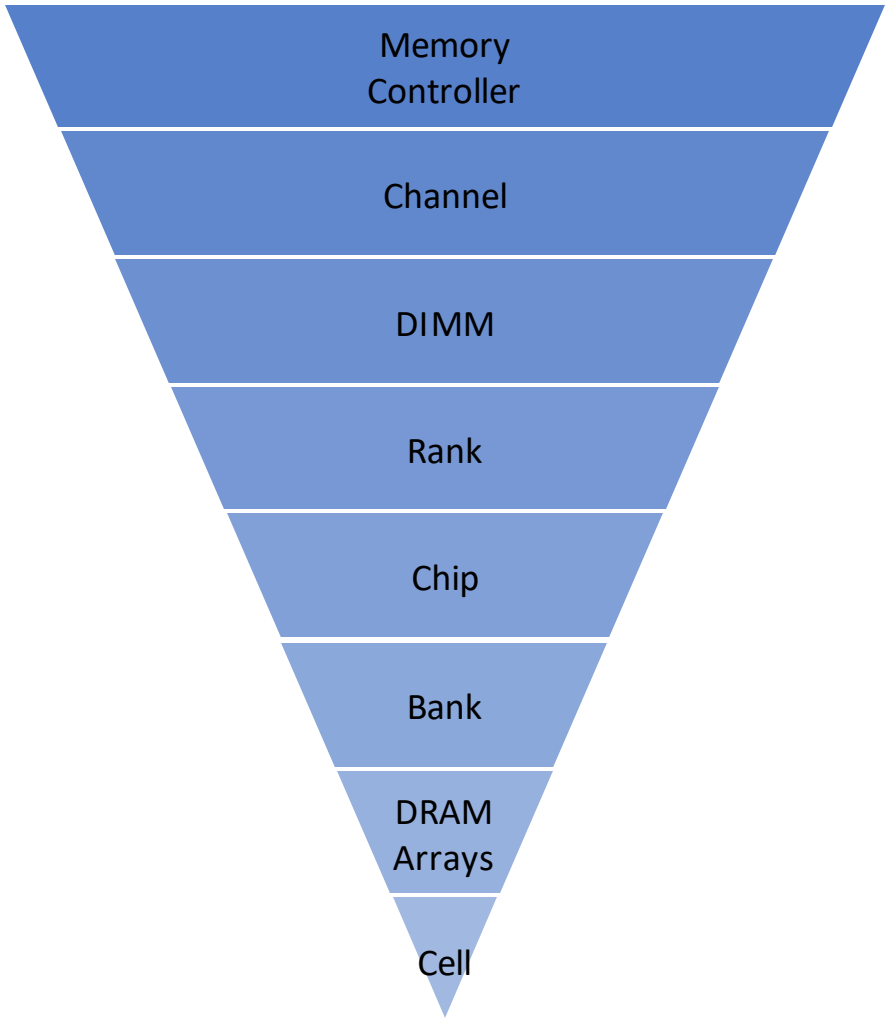
ZDNet MENU UK

DRAM errors: from soft to hard

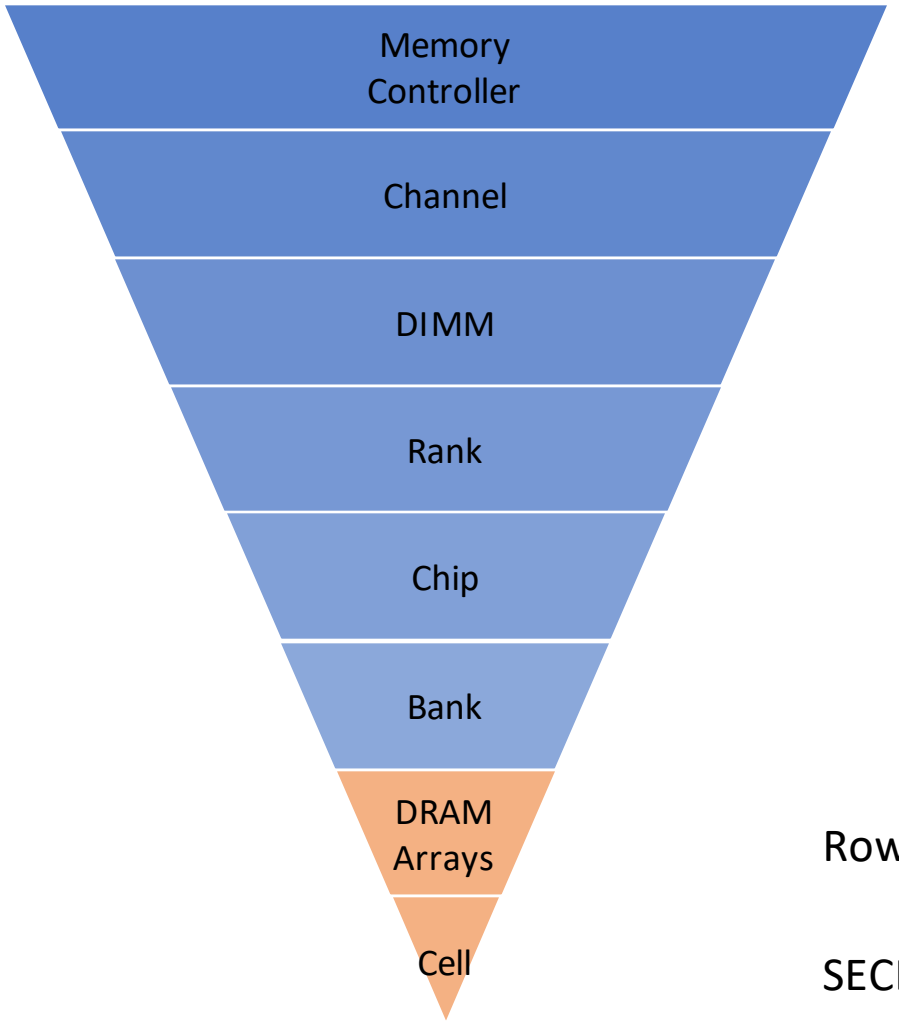
Every system uses dynamic random access memory (DRAM), but how good is it? Bad news: not nearly as good as vendors would like us to think. Good news: we're learning.

By Robin Harris for Storage Bits | October 24, 2012 -- 16:26 GMT (17:26 BST) | Topic: Storage

Progression of Reliability Mechanisms



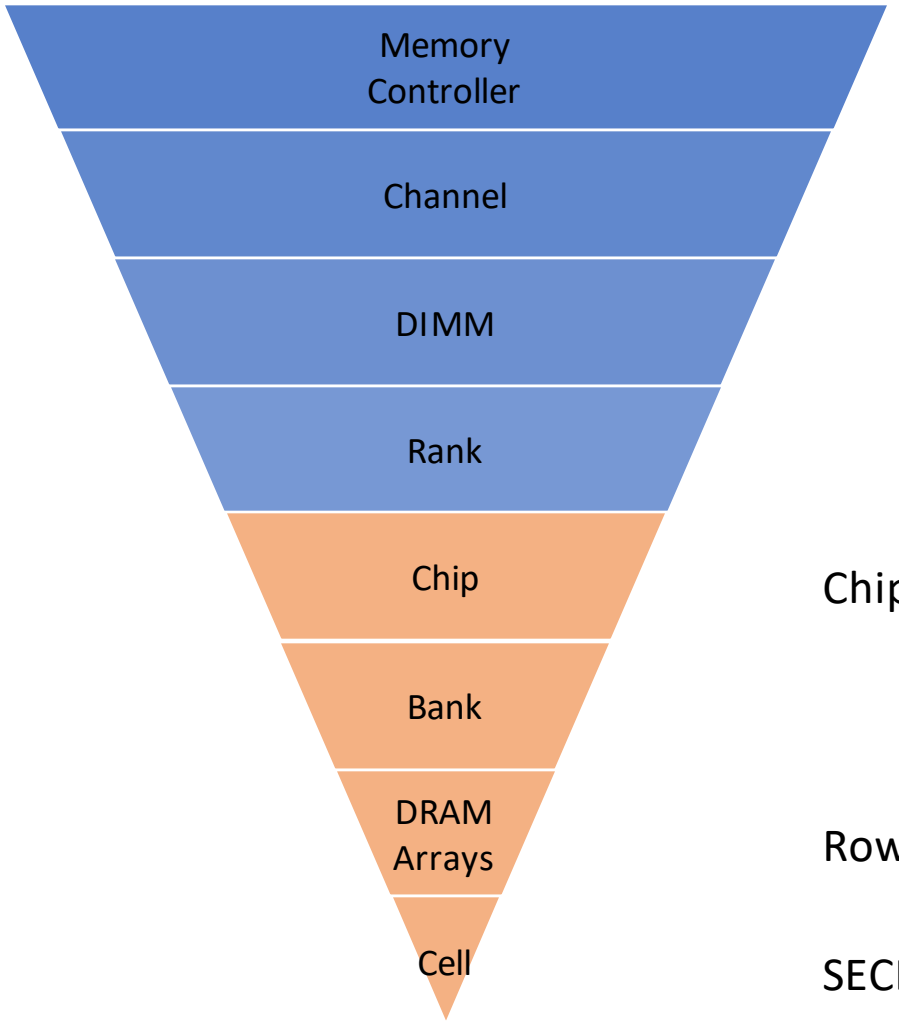
Progression of Reliability Mechanisms



Row/column sparing

SECDED ECC, in-DRAM ECC

Progression of Reliability Mechanisms

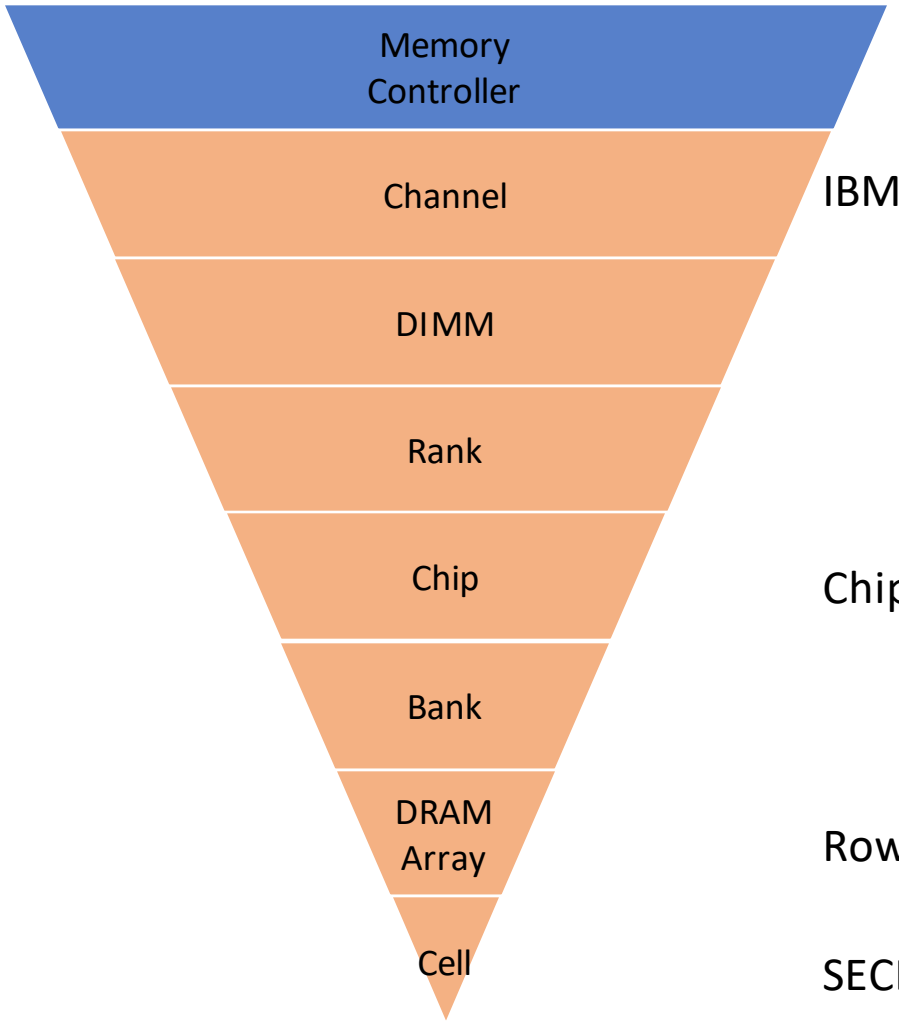


Chipkill ECC / Multi-tier ECC

Row/column sparing

SECDED ECC, in-DRAM ECC

Progression of Reliability Mechanisms



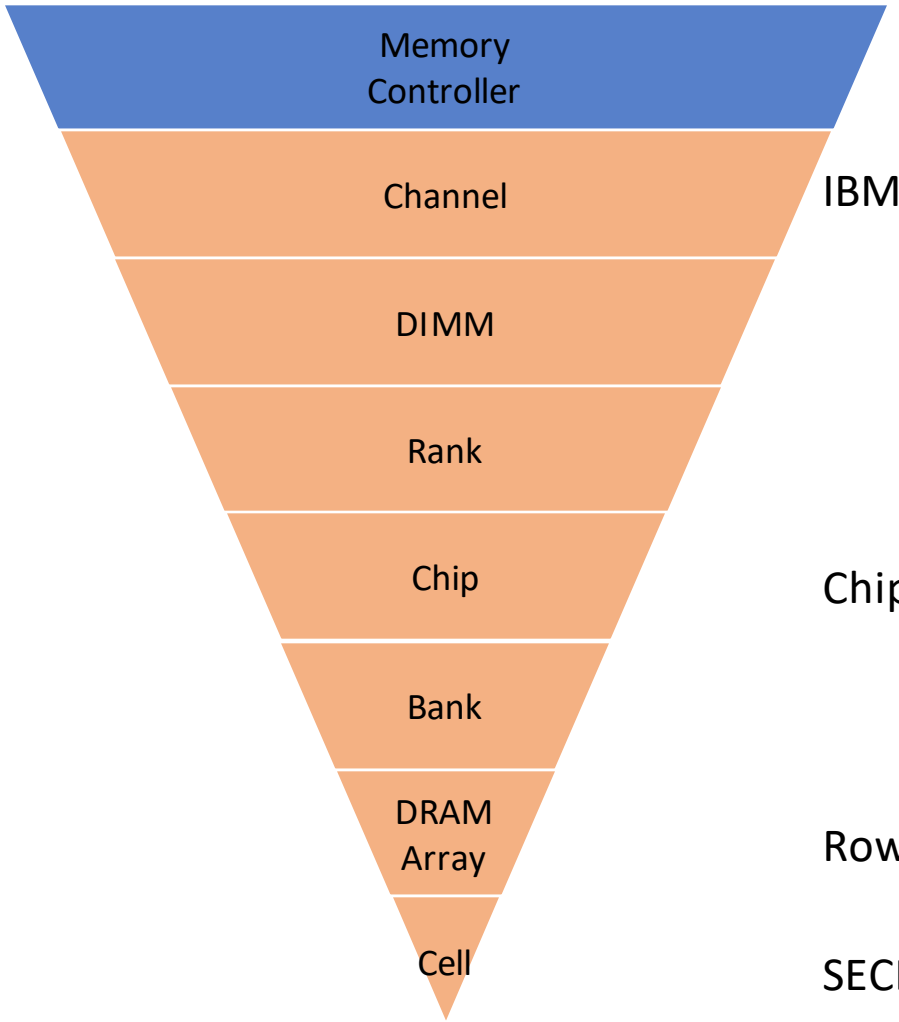
IBM RAIM, Intel Memory Mirroring

Chipkill ECC / Multi-tier ECC

Row/column sparing

SECDED ECC, in-DRAM ECC

Progression of Reliability Mechanisms



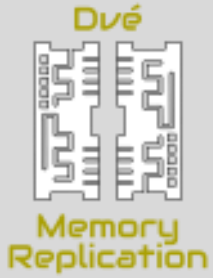
IBM RAIM, Intel Memory Mirroring

Chipkill ECC / Multi-tier ECC

Row/column sparing

SECDED ECC, in-DRAM ECC

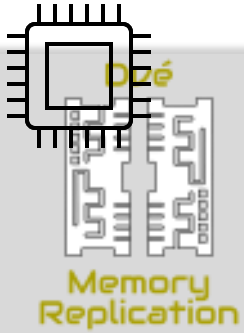
Performance overheads



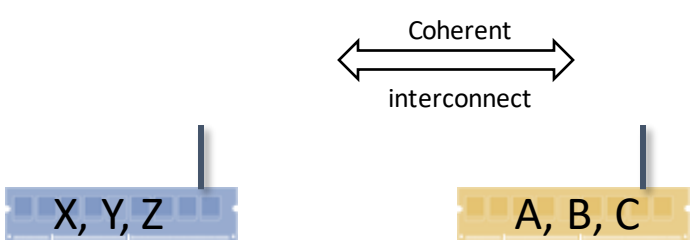
Replication for Reliability

Dvé insights

- ❑ Full data replica (not ECC code)
- ❑ Keep Replicas as far apart and disjoint as possible
- ❑ Tolerate errors arising from anywhere in the memory path

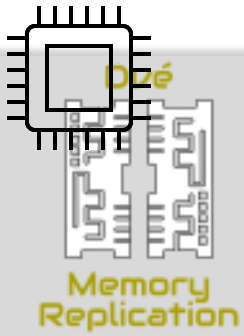


Replication for Reliability

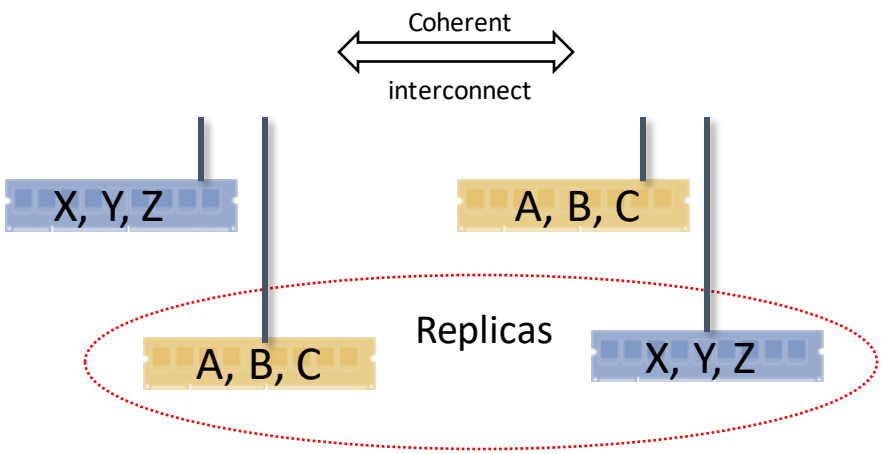


Dvé insights

- ❑ Full data replica (not ECC code)
- ❑ Keep Replicas as far apart and disjoint as possible
- ❑ Tolerate errors arising from anywhere in the memory path



Replication for Reliability

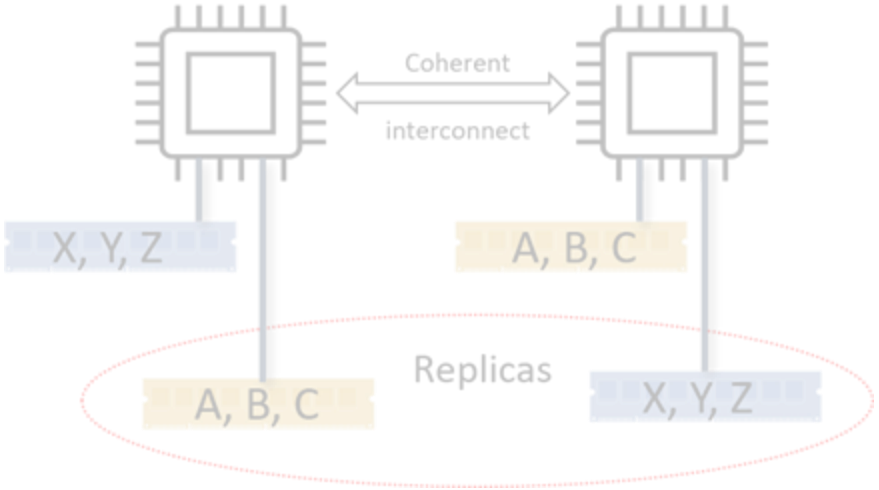


Dvé insights

- ❑ Full data replica (not ECC code)
- ❑ Keep Replicas as far apart and disjoint as possible
- ❑ Tolerate errors arising from anywhere in the memory path

Dvé
Memory Replication

Replication for Reliability



Dvé insights

- ❑ Full data replica (not ECC code)
- ❑ Keep Replicas as far apart and disjoint as possible
- ❑ Tolerate errors arising from anywhere in the memory path

For Detection

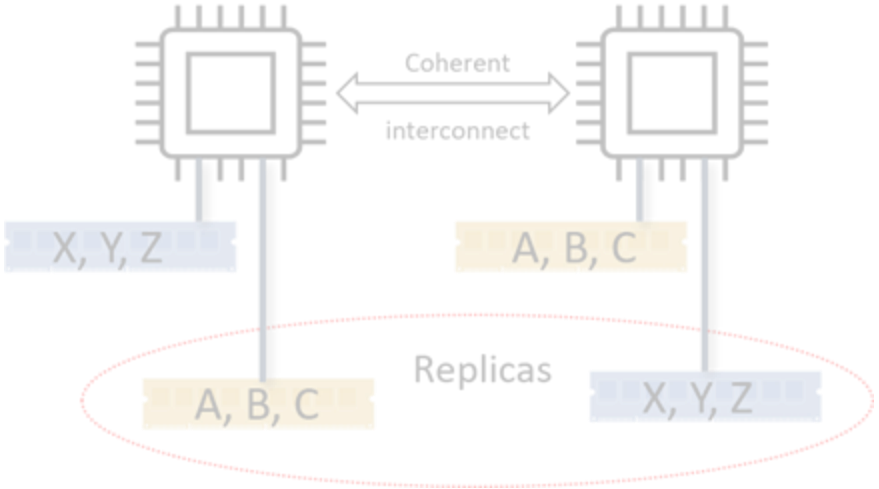
- ❑ Existing ECC, CRC, Parity
- ❑ Strong detection-only code
- ❑ Other diagnostic capabilities

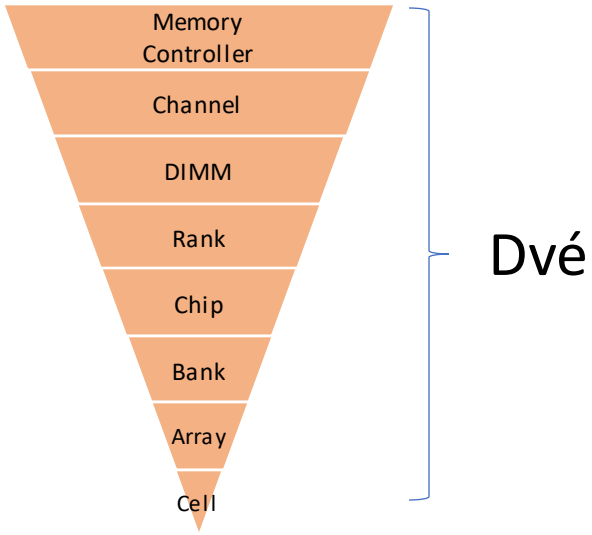
For Correction

- ❑ Rely on replica

Dvé
Memory Replication

Replication for Reliability





Dvé

Dvé insights

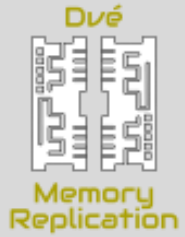
- ❑ Full data replica (not ECC code)
- ❑ Keep Replicas as far apart and disjoint as possible
- ❑ Tolerate errors arising from anywhere in the memory path

For Detection

- ❑ Existing ECC, CRC, Parity
- ❑ Strong detection-only code
- ❑ Other diagnostic capabilities

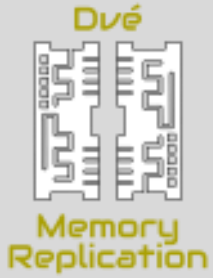
For Correction

- ❑ Rely on replica



Outline

- Reliability benefits
- Performance gains
- On-demand Reliability

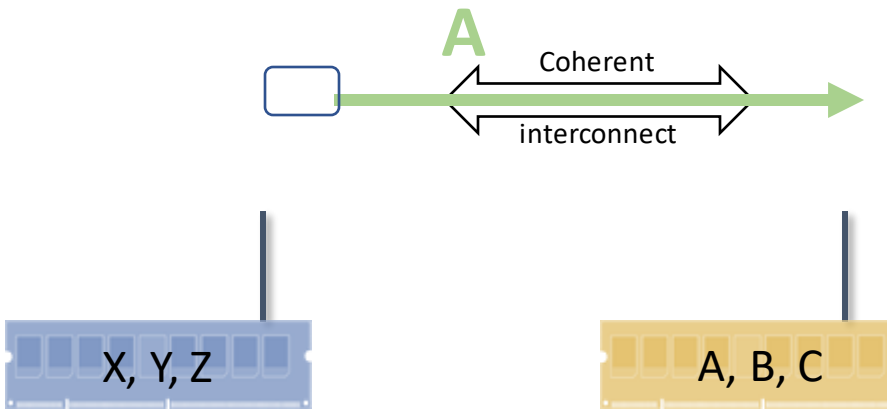


Coherent Replication

Dvé insights

- ❑ Use replica to improve performance

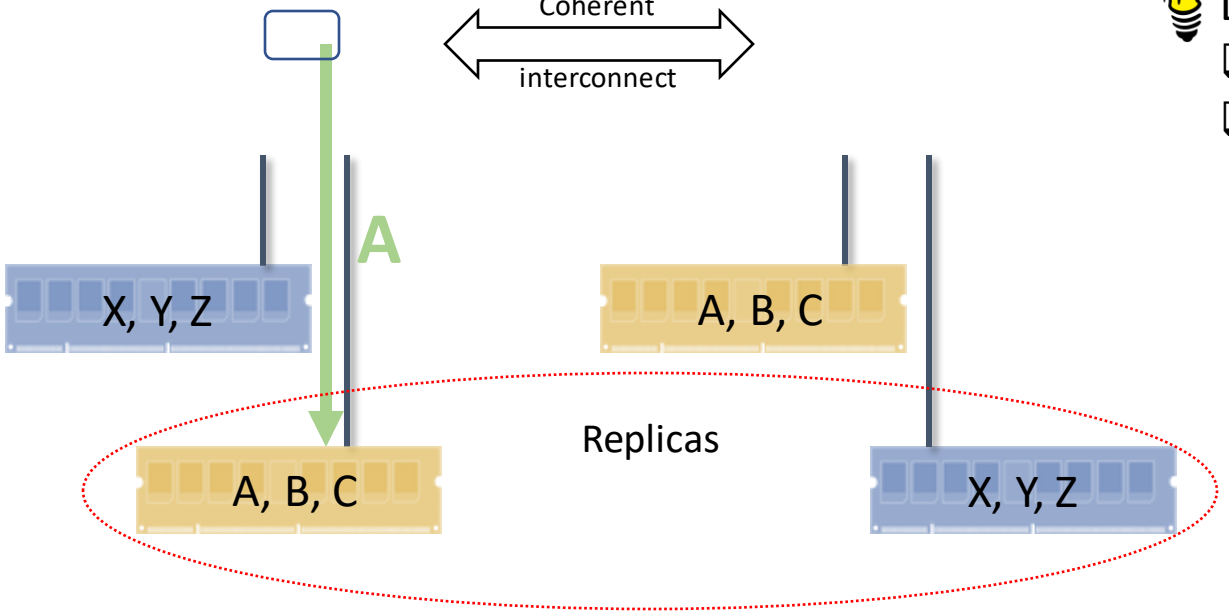
Coherent Replication



Dvé insights

- ❑ Use replica to improve performance

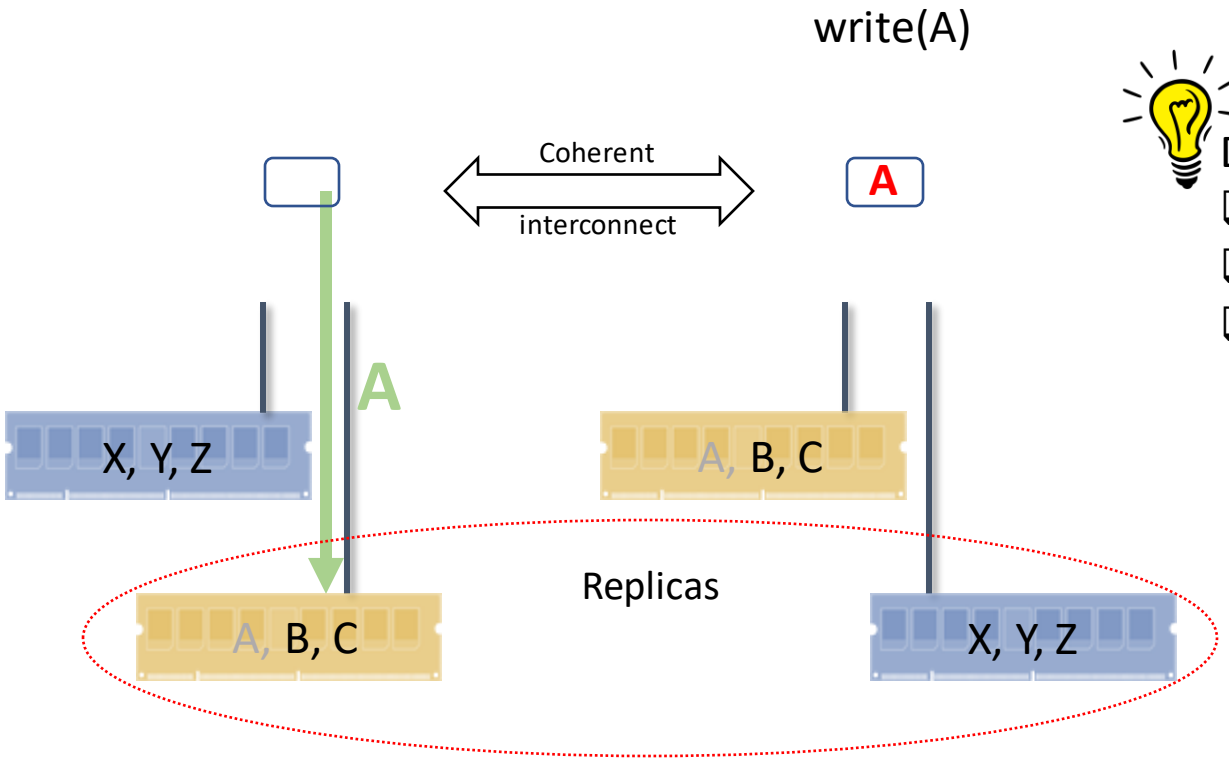
Coherent Replication



Dvé insights

- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica

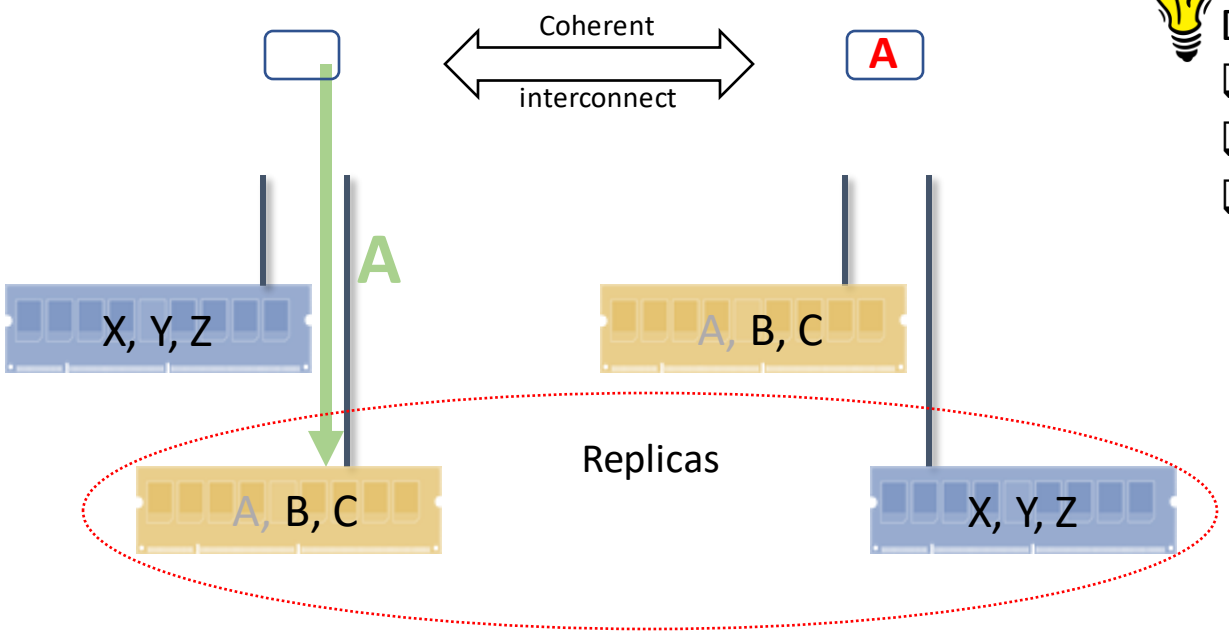
Coherent Replication



Dvé insights

- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica
- ❑ Ensure safe access to replica

Coherent Replication

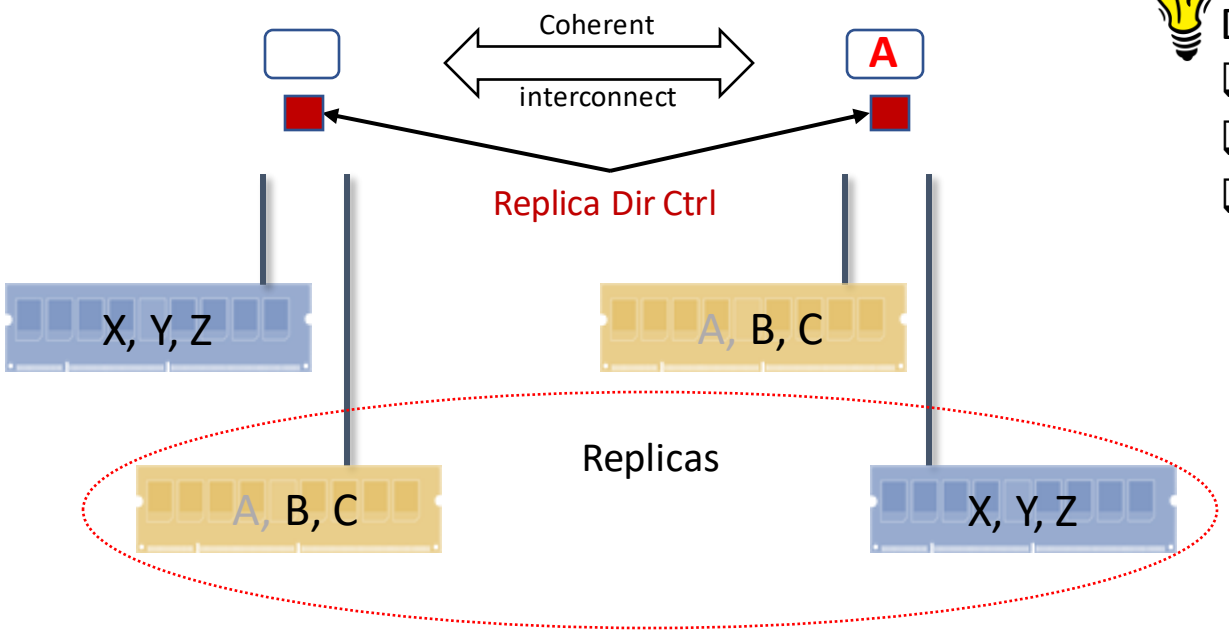


Dvé insights

- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica
- ❑ Ensure safe access to replica

Dvé
Memory Replication

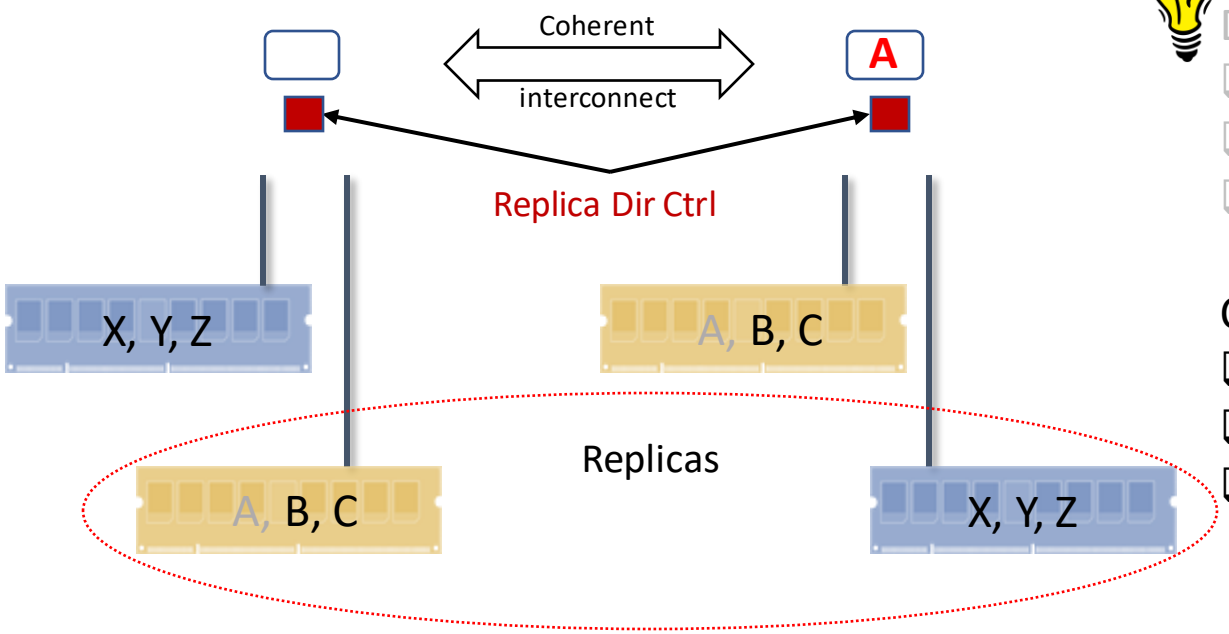
Coherent Replication



Dvé insights

- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica
- ❑ Ensure safe access to replica

Coherent Replication



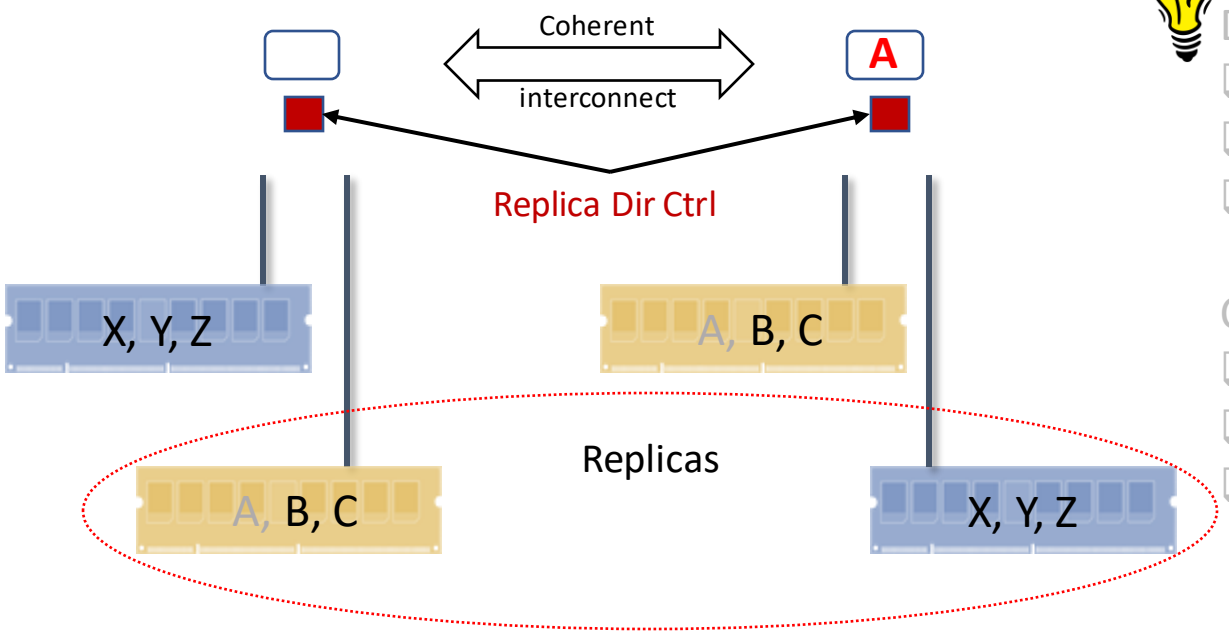
Dvé insights

- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica
- ❑ Ensure safe access to replica

Coherent Replication

- ❑ Builds on existing cache coherence protocols
- ❑ maintain the replicas in sync (for reliability)
- ❑ provide coherent access to both replicas during fault-free operation (for performance)

Coherent Replication



Dvé insights

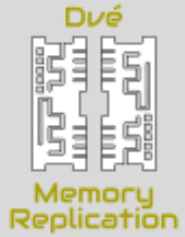
- ❑ Use replica to improve performance
- ❑ Route memory requests to nearest replica
- ❑ Ensure safe access to replica

Coherent Replication

- ❑ Builds on existing cache coherence protocols
- ❑ maintain the replicas in sync (for reliability)
- ❑ provide coherent access to both replicas during fault-free operation (for performance)

Mechanisms

- ❑ Allow-based
- ❑ Deny-based



Outline

- Reliability benefits
- Performance gains
- On-demand Reliability

Dvé
Memory Replication

Capacity overheads?

Dvé
Memory Replication

Capacity overheads?

Dvé insights

- ❑ Utilize idle memory

Skewed memory utilization

- ❑ 50% of the memory is idle in 90% of the servers
- ❑ Provisioning for peak

Dvé
Memory Replication

Capacity overheads?

Dvé insights

- ❑ Utilize idle memory
- ❑ Overheads applicable only as and when demanded by the application

Skewed memory utilization

- ❑ 50% of the memory is idle 90% of the servers
- ❑ Provisioning for peak

Interface to allocate high-reliability memory

- ❑ Hardware-software co-design
- ❑ OS support

Dvé
Memory Replication

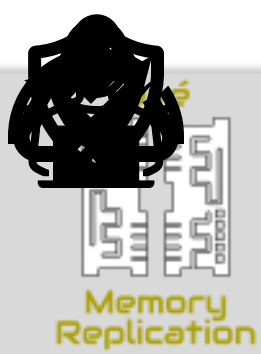
Capacity overheads?

Dvé insights
- ❑ Utilize idle memory
- ❑ Overheads applicable only as and when demanded by the application

Skewed memory utilization
- ❑ 50% of the memory is idle 90% of the time
- ❑ Provisioning for peak

Interface to allocate high-reliability memory
- ❑ Hardware-software co-design
- ❑ OS support

Flexible trade-off between capacity and reliability



Summary

Replication for Reliability

Lowers DUE by

4x over Chipkill
172x over IBM RAIM
11% over Intel Memory Mirroring

Coherent Replication for Performance

Improves performance by

5% - 117% over baseline NUMA
3% - 107% over an improved Intel mirroring scheme

On-demand Replication

hardware-software co-design
using OS/compiler support

Paper in ISCA '21 Artifacts available

https://github.com/adarshpatil/dve
https://adar.sh/dve

Outline

Improving Reliability and Performance

Dvé: Coherent Replication for DRAMs *(ISCA 2021)*

Coherent Disaggregated Shared Memory for FaaS *(WIP)*

Function-as-a-Service characteristics

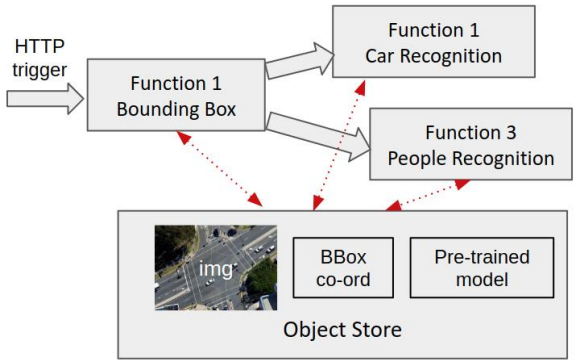


Fig: Sample FaaS app (AMBER alert pipeline)

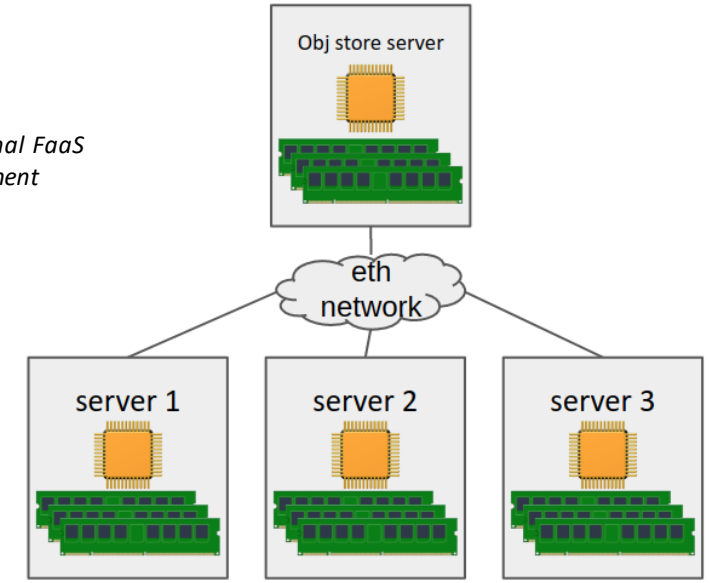


Fig: Traditional FaaS deployment

FaaS Applications

- Composition of stand-alone functions
- DAG invocation sequence

FaaS Functions

- Stateless: No access to state created by previous invocations
- object store backend with a get/put API

FaaS execution Infrastructure

- managed by cloud provider
- Scales by adding/removing function instances
- Runtime orchestrates and load balances

Our proposal: FaaS-DM

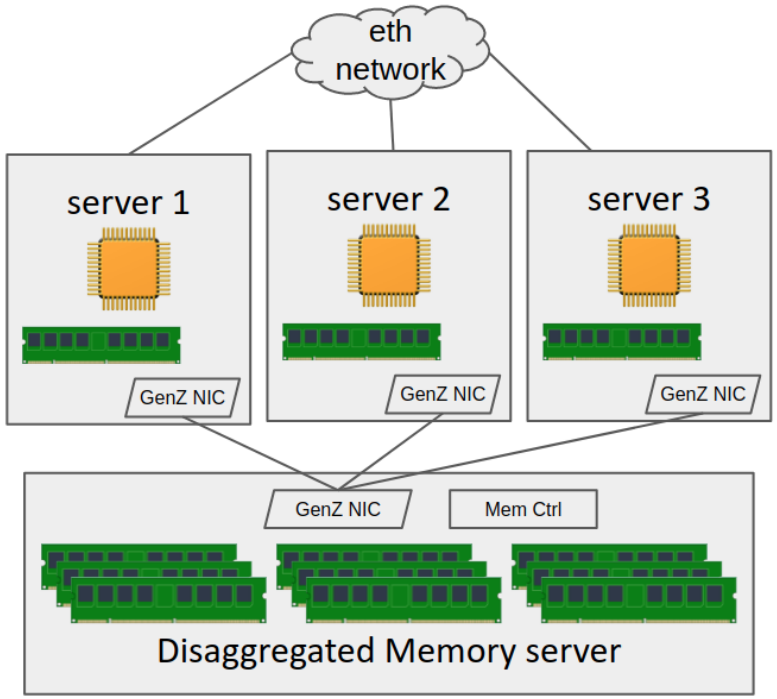


Fig: FaaS with Disaggregated Shared Memory

FaaS with Disaggregated Shared Memory

- ✓Hardware caching
- ✓Benefit from hardware prefetching and intra-node coherence
- ✓Application transparent caching
- ✓Implicit data movement using inter-node hardware cache coherence
- ✓Use existing shared memory synchronization techniques
- ✓Avoids overfetch and critical path writeback

Open Problems in FaaS-DM

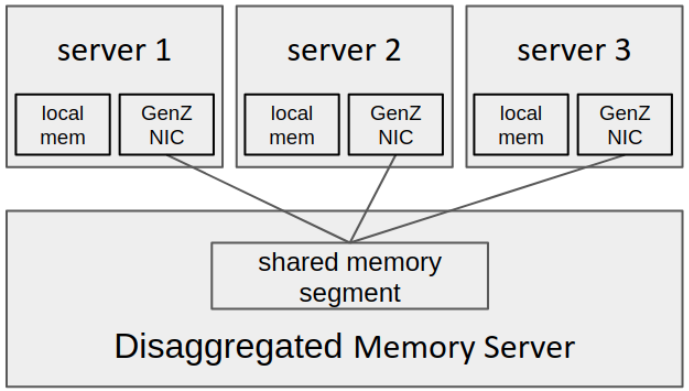


Fig: FaaS-DM logical view

Address mapping and translation

- Sharing the same physical memory region between independent servers
- Communication/co-ordination between executing functions

Performance

- Optimizing inter-node coherence protocol

Availability

- Partial system failure (non-fate sharing)

Our approach

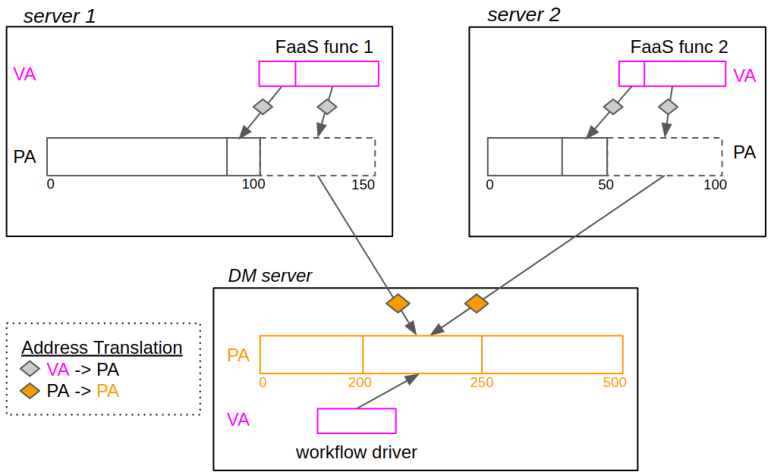


Fig: FaaS-DM addressing schematic

Address mapping and translation

- Extends shared memory inter-process communication (IPC)
  [POSIX API: shm_open, shm_close, mmap]
- Extends annexation process to allow mapping of FaaS-DM memory segment into FaaS function VA
- OS's exchange messages via RPC for naming and identification of FaaS-DM memory segments [LegoOS, OSDI '18]
- Address translation similar to 2-level page table [DeACT, HPCA '21]

Our approach

Performance

- Sharing characteristics of FaaS workloads
- Imposed function execution limits

Availability

- Non-blocking coherence protocol to guarantee forward progress
- Atomic durability and Memory consistency guarantees

Summary

Improving Reliability and Performance

Dvé: Coherent Replication for DRAMs *(ISCA 2021)*

Coherent Disaggregated Shared Memory for FaaS *(WIP)*

me@adarshpatil.in

adarsh_patil