Ελλάδα 2.0
ΕΘΝΙΚΟ ΣΧΕΔΙΟ ΑΝΑΚΑΜΨΗΣ ΚΑΙ ΑΝΘΕΚΤΙΚΟΤΗΤΑΣ

Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης
NextGenerationEU

SUB1.1 Συμπράξεις Ερευνητικής Αριστείας – ΣΕΑ

HAR.S.H: Hardware-Aware extReme-scale Similarity search
Αναζήτηση Ομοιότητας σε Μεγάλες Συλλογές Σειρών Δεδομένων συνυπολογίζοντας το Υλικό

ΥΠ3ΤΑ-0560901

Παραδοτέο Π6.4

Αναρτημένα Πόστερ Έργου

Ιούνιος 2026

Π6.4.1 Εισαγωγή

Το παρόν παραδοτέο καταγράφει τα αναρτημένα πόστερ και τις προσκεκλημένες ομιλίες που πραγματοποιήθηκαν στο πλαίσιο του έργου HAR.S.H. κατά την περίοδο υλοποίησής του. Οι δράσεις αυτές συνέβαλαν στη διάχυση των ερευνητικών αποτελεσμάτων του έργου προς τη διεθνή επιστημονική κοινότητα, την ακαδημαϊκή κοινότητα, καθώς και προς φορείς και οργανισμούς που δραστηριοποιούνται σε συναφείς τεχνολογικούς τομείς.

Συνολικά, στο πλαίσιο του έργου παρουσιάστηκαν δέκα (10) πόστερ σε διεθνείς και εθνικές επιστημονικές διοργανώσεις, καθώς και δεκαεπτά (17) προσκεκλημένες ομιλίες σε πανεπιστήμια, ερευνητικά ινστιτούτα, διεθνή workshops, θερινά σχολεία και βιομηχανικούς φορείς. Μέσω αυτών των δράσεων, η ομάδα του HAR.S.H. ανέδειξε τα αποτελέσματα του έργου, σε περιοχές όπως η αναζήτηση ομοιότητας, οι παράλληλες και ταυτόχρονα προσπελάσιμες δομές δεδομένων, η πολυτροπική μάθηση και η ανάλυση διανυσματικών δεδομένων μεγάλης κλίμακας.

Π6.4.2 Κατάλογος πόστερ

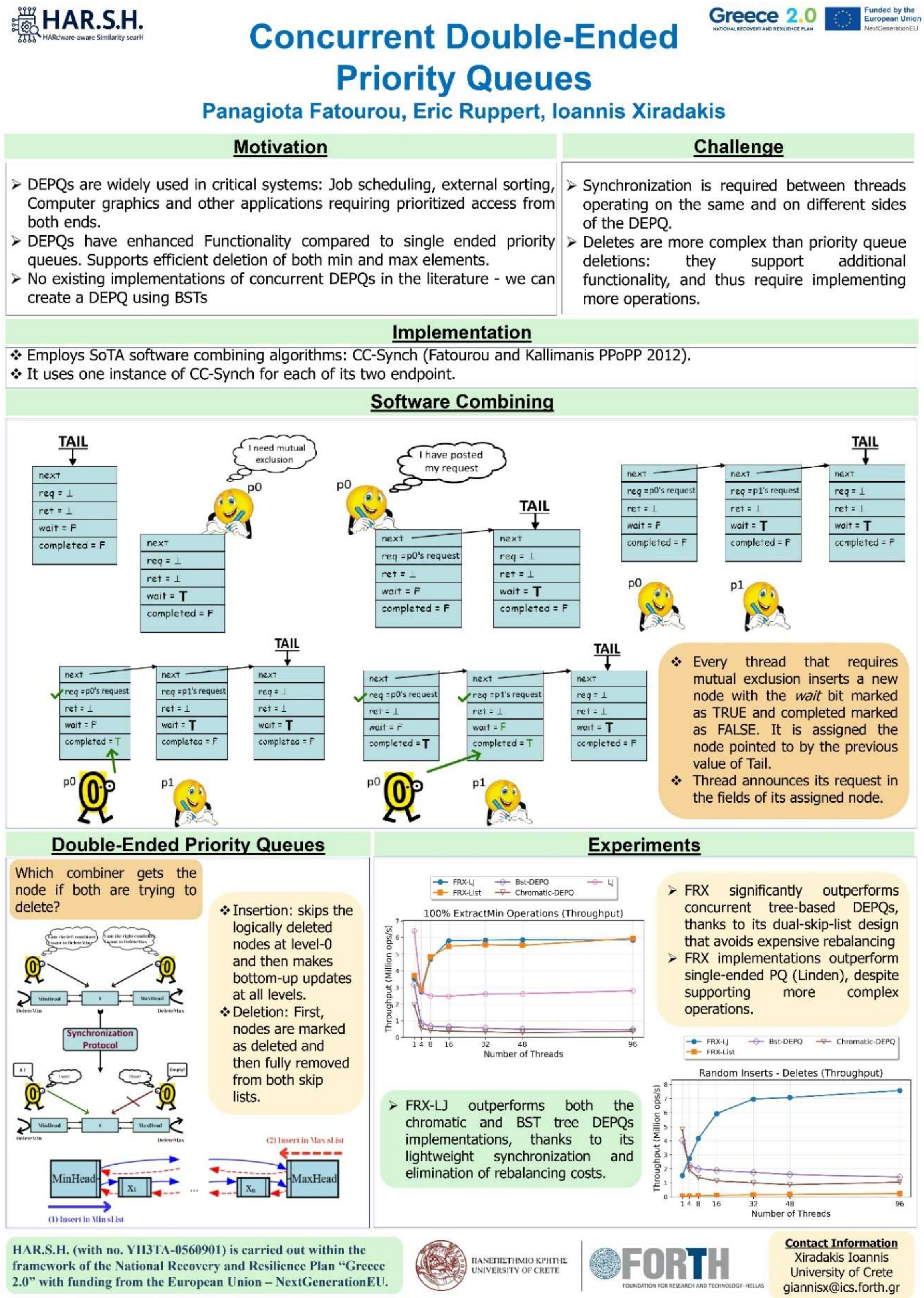
1. Ioannis Xiradakis, "Concurrent Double-Ended Priority Queues", Αφίσα που παρουσιάστηκε στο School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, 2026.
2. Myron Tsatsarakis, "PPMR: Persistent Polymorphic Memory Resources", Αφίσα που παρουσιάστηκε στο School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, 2026.
3. Myron Tsatsarakis, "ATROPOS: Benchmarking Concurrent Persistent Algorithms under Failures", Αφίσα που παρουσιάστηκε στο School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, 2026.
4. Georgios Paterakis, "PFRESH: A persistent Lock Free Data Series Index", Αφίσα που παρουσιάστηκε στο School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, 2026.
5. Michaela Areti Zervou, "The More, the Merrier: Contrastive Fusion for Higher-Order Multimodal Alignment", Αφίσα που παρουσιάστηκε στο School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, 2026.
6. Antonios Katsarakis, Vasilis Gavrielatos, Emmanouil Giortamis, Pramod Bhatotia, Aleksandar Dragojevic, Boris Grot, Vijay Nagarajan, Panagiota Fatourou, "Beyond reCAP: Local Reads and Linearizable Asynchronous Replication", Αφίσα που παρουσιάστηκε στο 51st International Conference on Very Large Data Bases (VLDB 2025), London, UK, 2025. (**Προηγούμενη έκδοση της αφίσας προτάθηκε για το Βραβείο Καλύτερης Αφίσας στο EuroSys.**).
7. Nikolaos Grigoroudis, "Concurrent Binary Search Trees Supporting Split and Join", Αφίσα που παρουσιάστηκε στο ACM-W Celebration of Women in Computing: 7th Summit on Gender Equality in Computing (GEC 2025), Piraeus, Greece, 2025.
8. Ioannis Xiradakis, "Concurrent Double-Ended Priority Queues", Αφίσα που παρουσιάστηκε στο ACM-W Celebration of Women in Computing: 7th Summit on Gender Equality in Computing (GEC 2025), Piraeus, Greece, 2025.
9. HAR.S.H. team, “HARSH: Hardware-Aware extReme-scale Similarity search”, Αφίσα που παρουσιάστηκε στο 40 Years CSD Event, Department of Computer Science – University of Crete, Greece, 2025.
10. HAR.S.H. team, “HARSH: Hardware-Aware extReme-scale Similarity search”, Αφίσα που παρουσιάστηκε στο Computer Science Graduates Reunion (Classes of '85, '95, '05'), Department of Computer Science – University of Crete, Greece, 2025.

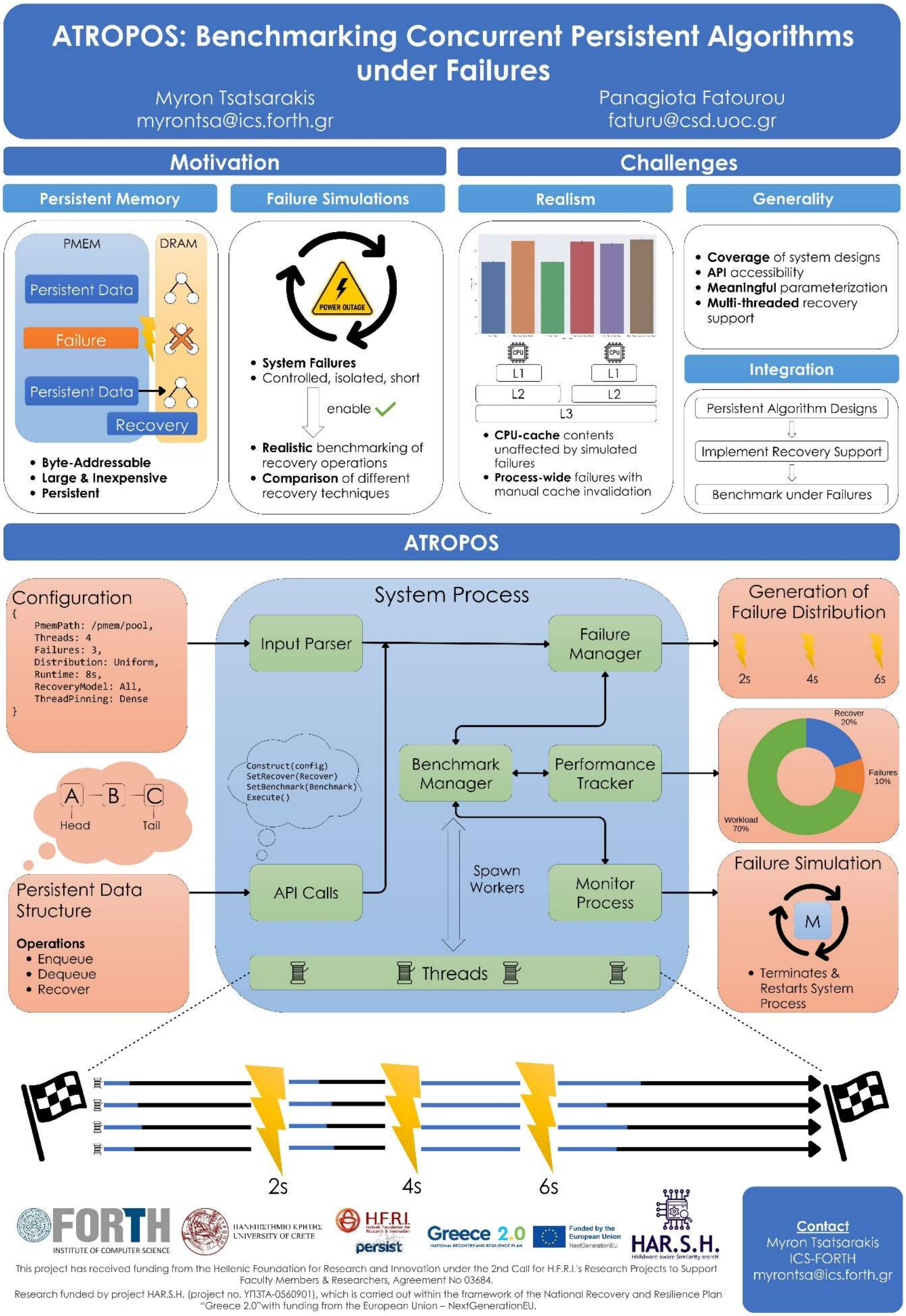
Π6.4.3 Προσκεκλημένες ομιλίες

1. Panagiota Fatourou, “Concurrent Trees Supporting Complex Queries”, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), November 2025.

2. Panagiota Fatourou, “Concurrent Trees Supporting Complex Queries”, Khoury College of Computer Science, Northeastern University, November 2025.
3. Panagiota Fatourou, “Parallel Data Series Indexing and Similarity Search on Modern Hardware”, Invited Talk at the Managing Specialized and Heterogeneous Architectures Workshop, Melvin Calvin Laboratory, Simons Institute, USA, November 2025. (video: https://simons.berkeley.edu/talks/panagiota-fatourou-university-crete-forth-2025-11-07)
4. Panagiota Fatourou, “Combining and Aggregating for Fast Synchronization”, Invited Talk at the Simons Institute for the Theory of Computing, University of California at Berkeley, USA, October 2025. *(video: https://www.youtube.com/watch?v=8-eZIZfoEzQ)*
5. Panagiota Fatourou, “Distributed Computing Mexico Summer School”, Invited Talk at the Distributed Computing Mexico Summer School 2025, Huatulco, Mexico, June 2025.
6. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at Nanyang Technological University (NTU), Singapore, May 2026.
7. Grigorios Tsagkatakis, “Parsimony in Distributed Learning”, Invited Talk at the School on the Practice and Theory of Distributed Computing (SPTDC 2026), FORTH, Greece, April 2026.
8. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at East China Normal University, China, December 2025.
9. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at FORTH Cert, Greece, October 2025.
10. Themis Palpanas, "Scalable Vector Analytics: A Story of Twists and Turns", Plenary Keynote at Guide-AI Workshop, in conjunction with VLDB 2025, London, UK, September 2025.
11. Themis Palpanas, "Scalable Vector Analytics: A Story of Twists and Turns", Invited Presentation at Technical University Berlin, Germany, June 2025.
12. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at Google New York, USA, November 2025.
13. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at Google Sunnyvale, USA, November 2025.
14. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at AT&T Labs Bedminster, USA, November 2025.
15. Themis Palpanas, "Vector Similarity Search: Past, Present, and Future", Invited Presentation at Amazon Palo Alto, USA, November 2025.
16. Themis Palpanas, "Scalable Vector Analytics: A Story of Twists and Turns", Invited Presentation at Snowflake Computing GmbH, June 2025.
17. Themis Palpanas, "High-Dimensional Vector Similarity Search at the AI Era", Plenary Keynote at Huawei Vision Forum, Paris, France, May 2025.

Π6.4.4 Παράρτημα: Πόστερ έργου



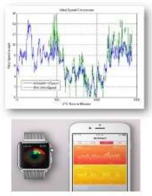

ATROPOS: Benchmarking Concurrent Persistent Algorithms under Failures
Myron Tsatsarakis
myrontsa@ics.forth.gr
Panagiota Fatourou
faturu@csd.uoc.gr
Motivation
Challenges
Persistent Memory
Failure Simulations
Realism
Generality
PMEM
DRAM
Persistent Data
Failure
Persistent Data
Recovery
• Byte-Addressable
• Large & Inexpensive
• Persistent
POWER OUTAGE
• System Failures
• Controlled, isolated, short
enable
• Realistic benchmarking of recovery operations
• Comparison of different recovery techniques
CPU
L1
L2
L3
• CPU-cache contents unaffected by simulated failures
• Process-wide failures with manual cache invalidation
• Coverage of system designs
• API accessibility
• Meaningful parameterization
• Multi-threaded recovery support
Integration
Persistent Algorithm Designs
Implement Recovery Support
Benchmark under Failures
ATROPOS
Configuration
{
PmemPath: /pmem/pool,
Threads: 4
Failures: 3,
Distribution: Uniform,
Runtime: 8s,
RecoveryModel: All,
ThreadPinning: Dense
}
A B C
Head
Tail
Persistent Data Structure
Operations
• Enqueue
• Dequeue
• Recover
System Process
Input Parser
Failure Manager
Construct(config)
SetRecover(Recover)
SetBenchmark(Benchmark)
Execute()
Benchmark Manager
Performance Tracker
API Calls
Spawn Workers
Monitor Process
Threads
Generation of Failure Distribution
2s
4s
6s
Recover 20%
Failures 10%
Workload 70%
Failure Simulation
M
• Terminates & Restarts System Process
2s
4s
6s
FORTH
INSTITUTE OF COMPUTER SCIENCE
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE
H.F.R.I.
persist
Greece 2.0
NATIONAL RECOVERY AND RESILIENCE PLAN
Funded by the European Union
NextGenerationEU
HAR.S.H.
HARdware aware Similarity searcH
This project has received funding from the Hellenic Foundation for Research and Innovation under the 2nd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers, Agreement No 03684.
Research funded by project HAR.S.H. (project no. ΥΠ3ΤΑ-0560901), which is carried out within the framework of the National Recovery and Resilience Plan "Greece 2.0" with funding from the European Union – NextGenerationEU.
Contact
Myron Tsatsarakis
ICS-FORTH
myrontsa@ics.forth.gr

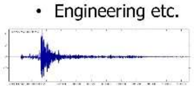
PFRESH: A persistent Lock Free Data Series Index

Georgios Paterakis, Panagiota Fatourou, Eleftherios Kosmas, Themis Palpanas

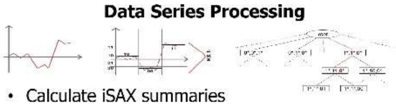
Motivation and Challenges

Data Series
- Many Applications
  - Seismology
  - Astrophysics
  - Neuroscience
  - Engineering etc.





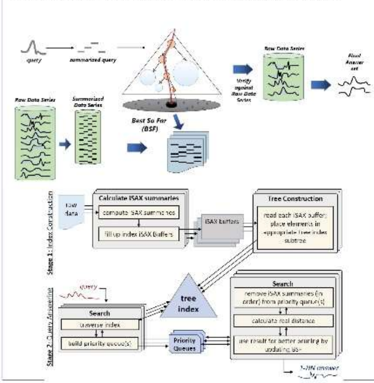
Data Series Processing



- Calculate iSAX summaries

Similarity Search
Find the most similar series of a collection to a query series

How an iSAX data series index works



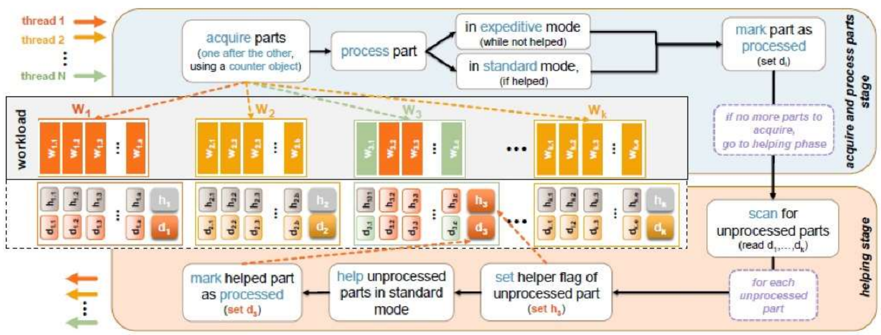
Blocking Implementations

- Use of locks
  - Convoying
  - Priority Inversion
  - Restricted Parallelism
  - Thread holding the lock fails

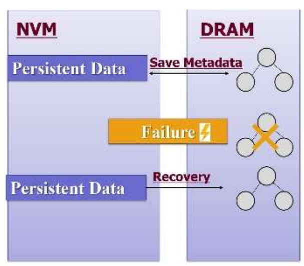
Transform to non-Blocking Implementations
- Some sort of helping is needed.
  - Helping is costly
  - Adds complexity
  - It's not always easy to implement.

Refresh

Generic approach
- Applicable to any blocking locality-aware data series index
- Ensures Lock-Freedom



Non-Volatile Memory



NVM Properties
- Byte Addressable
- Large & Inexpensive
- Persistent

Data Series Indexes
- Data series indexes grow with dataset size
- Larger datasets slow down construction time
- Index failure requires full reconstruction

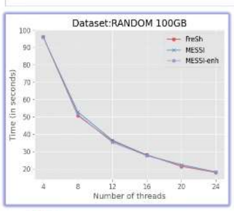
Evaluation

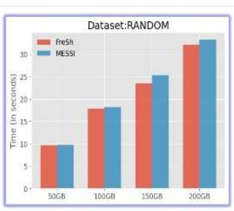
Results

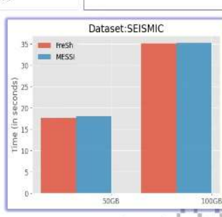
DRAM Configurations and Datasets
- Compared with optimized SOTA MESSI index.
- Use both synthetic and real datasets.

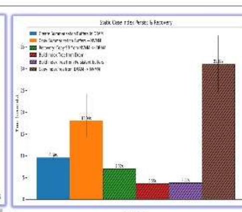
DRAM Experimental Findings
- FreSh performs as good as the SOTA blocking index.
- No penalty for providing lock-freedom.
- FreSh outperforms by far several lock-free baselines we have designed.
- In case of delays achieves better performance.

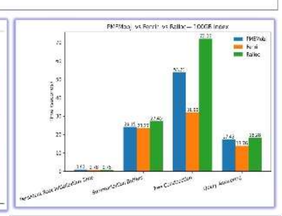
Non-Volatile Configurations
- Measure the persistent and recovery cost of each stage of PFreSh.
- Compared the implementation using three Persistent Allocators.











Contact Information
Paterakis Georgios
University of Crete, ICS-FORTH
Email: geopat@ics.forth.gr

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ UNIVERSITY OF CRETE · FORTH FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS · H.F.R.I. Hellenic Foundation for Research & Innovation · persist · HAR.S.H. · Université Paris Cité · LIPADE Laboratoire d'Informatique PAris DEscartes · diNo · diip data intelligence institute of Paris · Ελλάδα 2.0 · Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης NextGenerationEU

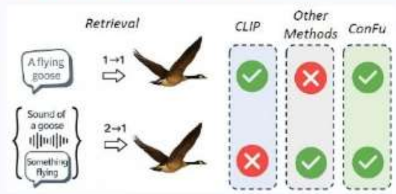
This project has received funding from the Hellenic Foundation for Research and Innovation under the 2nd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers, Agreement No 03684. Research funded by project HAR.S.H. (project no. ΥΠ3ΤΑ-0560901), which is carried out within the framework of the National Recovery and Resilience Plan "Greece 2.0" with funding from the European Union – NextGenerationEU.

THE MORE, THE MERRIER: CONTRASTIVE FUSION FOR HIGHER-ORDER MULTIMODAL ALIGNMENT

Stefanos Koutoupis, Michaela Areti Zervou, Konstantinos Kontras, Maarten De Vos, Panagiotis Tsakalides, Grigorios Tsagkatakis

HAR.S.H. HARdware-aware Similarity searcH

Motivation & Key Idea



Prevailing methods align only pairs of modalities, missing synergistic higher-order interactions.

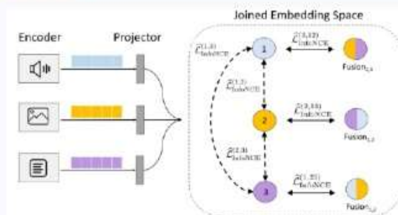
> Can a contrastive learning framework capture not only pairwise alignments but also higher-order, synergistic dependencies among modalities?

- **ConFu**: unifies pairwise (1→1) and higher-order (2→1) contrastive learning in one objective.
- **Theory:** $\mathcal{L}$ maximizes a lower bound on Total Correlation, capturing pairwise + synergistic interactions.

Key Contributions

- **1→1 & 2→1 retrieval:** only method supporting both modes without requiring all modalities at inference.
- **Bird-MML dataset:** 149K image–audio–text triplets across 150 bird species for multimodal complementarity research.
- **Robustness:** shows greater stability when input involves corrupted or non-informative inputs such as missing modalities or noise.

Method

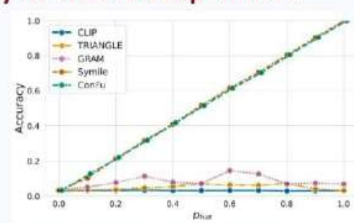


$$\mathcal{L}_{\text{fused}} = \widehat{\mathcal{L}}^{(3,12)}_{\text{InfoNCE}} + \widehat{\mathcal{L}}^{(2,13)}_{\text{InfoNCE}} + \widehat{\mathcal{L}}^{(1,23)}_{\text{InfoNCE}}$$

$$\mathcal{L}_{\text{pair}} = \widehat{\mathcal{L}}^{(1,2)}_{\text{InfoNCE}} + \widehat{\mathcal{L}}^{(1,3)}_{\text{InfoNCE}} + \widehat{\mathcal{L}}^{(2,3)}_{\text{InfoNCE}}$$

$$\mathcal{L} = (1-\lambda)\,\mathcal{L}_{\text{pair}} + \lambda\,\mathcal{L}_{\text{fused}}$$

Synthetic XOR Experiment



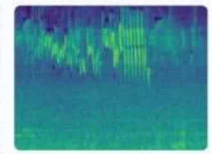
Predict z_2 from (z_1, z_3) where pairwise MI=0.
Only synergy can solve it.

- **ConFu** ✓ strong positive trend with $\hat{p}$
- **Symile** ✓ also captures XOR synergy
- **GRAM/TRIANGLE** ✗ <15% even with 1024-dim embeddings
- **CLIP** ✗ ~3%, random chance level

Bird-MML dataset

Image

Mel-Spectrogram



Description

This red male house finch, with a distinctive black and white back stripe, enjoys a meal from a wooden bird feeder.

Results

Retrieval@10 for Target

Dataset	Target	M1			M2			M3			Mean	
	Query	M2	M3	M23	M1	M3	M13	M1	M2	M12	1→1	2→1
MOSI	Trimodal CLIP	**22.9** ± 1.9	**20.5** ± 1.6	–	**23.5** ± 3.6	**23.3** ± 1.8	–	**20.9** ± 2.0	**24.3** ± 2.1	–	**22.6**	–
	Symile [27]	-	-	16.3 ± 0.9	-	-	18.1 ± 2.4	-	-	17.1 ± 1.7	–	17.2
	GRAM [5]	6.0 ± 3.1	10.3 ± 2.0	16.3 ± 2.5	7.7 ± 2.3	0.3 ± 0.6	7.9 ± 3.4	12.2 ± 4.1	0.2 ± 0.2	12.0 ± 4.2	6.1	12.1
	TRIANGLE [6]	-	-	8.3 ± 1.3	-	-	4.9 ± 0.7	-	-	8.8 ± 1.4	–	7.3
	ConFu	21.0 ± 1.7	16.4 ± 2.3	**16.7** ± 1.8	19.2 ± 1.1	21.0 ± 3.1	**21.6** ± 1.9	16.1 ± 1.7	23.5 ± 2.7	**20.5** ± 2.9	19.5	**19.6**
UR-FUNNY	Trimodal CLIP	**3.7** ± 0.6	**4.0** ± 0.6	–	**3.8** ± 0.5	**16.2** ± 1.6	–	**4.0** ± 0.4	**16.8** ± 1.2	–	**8.1**	–
	Symile [27]	-	-	3.7 ± 0.5	-	-	15.4 ± 0.9	-	-	16.5 ± 0.6	–	11.9
	GRAM [5]	3.2 ± 0.5	3.0 ± 0.9	3.9 ± 0.8	3.2 ± 0.9	0.1 ± 0.2	3.1 ± 0.4	3.9 ± 0.4	0.1 ± 0.0	3.3 ± 0.2	2.3	3.4
	TRIANGLE [6]	-	-	**4.2** ± 0.4	-	-	3.3 ± 0.7	-	-	3.6 ± 0.9	–	3.7
	ConFu	3.2 ± 0.3	3.5 ± 0.3	3.6 ± 0.3	3.5 ± 0.2	15.1 ± 0.8	**16.9** ± 0.5	3.5 ± 0.4	15.6 ± 0.8	**20.3** ± 1.1	7.4	**13.6**
MUSTARD	Trimodal CLIP	70.7 ± 4.1	29.1 ± 5.6	–	70.5 ± 4.3	26.5 ± 3.2	–	30.4 ± 3.4	24.6 ± 1.8	–	42.0	–
	Symile [27]	-	-	61.5 ± 6.3	-	-	57.0 ± 3.0	-	-	21.3 ± 4.4	–	46.6
	GRAM [5]	61.7 ± 8.0	24.8 ± 5.3	**81.0** ± 4.1	67.2 ± 5.4	24.8 ± 3.3	67.8 ± 5.4	27.9 ± 8.7	5.1 ± 2.1	31.2 ± 9.2	35.2	60.0
	TRIANGLE [6]	-	-	68.7 ± 4.2	-	-	59.9 ± 3.7	-	-	21.9 ± 3.0	–	50.2
	ConFu	**73.8** ± 3.2	**33.6** ± 3.1	79.6 ± 3.6	**74.4** ± 2.7	**28.1** ± 3.2	**74.6** ± 4.1	**33.9** ± 4.2	**28.0** ± 2.9	**33.5** ± 3.2	**45.3**	**62.6**

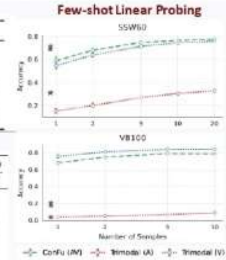
Zero-shot Classification on AV-MNIST

Method	A	V	A+V
CLIP	**41.1** ± 0.3	63.0 ± 0.4	-
Tri-CLIP	40.0 ± 0.4	62.5 ± 0.4	-
Symile [27]	-	-	70.9 ± 0.3
GRAM [5]	9.8 ± 0.4	63.9 ± 0.8	64.4 ± 0.5
TRIANGLE [6]	-	-	64.9 ± 0.3
ConFu	39.7 ± 0.1	**64.6** ± 0.3	**71.2** ± 0.2

Zero-shot Brid Classification

Method	SSW60 Acc. (%)			VB100 Acc. (%)		
	A	V	A+V	A	V	A+V
CLIP	29.9	70.1	–	4.2	20.6	–
Tri-CLIP	31.1	69.0	–	3.9	**20.7**	–
Symile [27]	-	-	60.2	-	-	13.4
TRIANGLE [6]	-	-	64.1	-	-	12.1
GRAM [5]	0.7	66.6	56.9	1.3	13.7	8.0
ConFu	30.3	69.4	**71.4**	3.4	19.3	18.1

Few-shot Linear Probing



Accuracy under Gaussian Noise

Method	10dB SNR		15dB SNR		20dB SNR	
	A deg.	V deg.	A deg.	V deg.	A deg.	V deg.
Tri-CLIP (V)	60.0	5.5	60.0	13.2	60.0	25.4
Tri-CLIP (A)	29.0	31.1	29.1	31.1	30.3	31.1
Symile [27]	58.4	21.2	58.5	27.9	60.0	40.9
GRAM [5]	58.9	4.0	58.3	8.98	58.1	25.9
TRIANGLE [6]	63.9	3.4	64.0	7.24	64.0	26.5
ConFu	71.2	30.2	71.4	33.1	71.5	45.4

Multimodal Competition

Research funded by project HAR.S.H. (project no. ΥΠ3ΤΑ-0560901), which is carried out within the framework of the National Recovery and Resilience Plan "Greece 2.0" with funding from the European Union – NextGenerationEU.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ UNIVERSITY OF CRETE · FORTH FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS · KU LEUVEN · Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN · Funded by the European Union NextGenerationEU

HAR.S.H.
HARdware-aware Similarity searH

Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN
Funded by the European Union NextGenerationEU

Concurrent Binary Search Trees Supporting Split and Join

Nikolaos Grigoroudis*, Panagiota Fatourou, Eric Ruppert

Motivation, Contribution

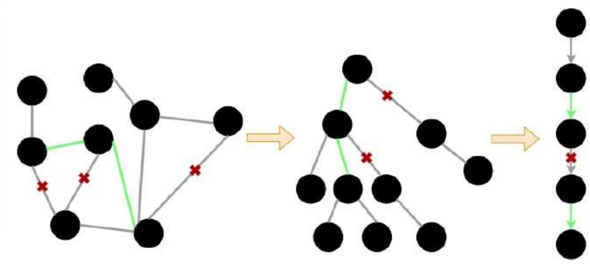
Need for BSTs that support Split and Join Operations

1.Dynamic Graph problems

- **Maintain a graph under edge insertions/deletions**
- **Answer queries efficiently**

2.Sequences that support Split and Join

- **Essential for some Dynamic Tree data structure (e.g. Link/Cut Trees, Euler Tour Trees)**
- **Represented using BSTs**



BST Operations

❑ Split(v): return two tree roots (v: node)

- R_1: keys < Key(v)
- R_2 keys ≥ Key(v)

❑ Join(R_1,R_2): join two tree roots

- Resulting tree has keys of both
- Precondition: the keys of Tree(R_1) are less than the keys of Tree(R_2)

Contribution

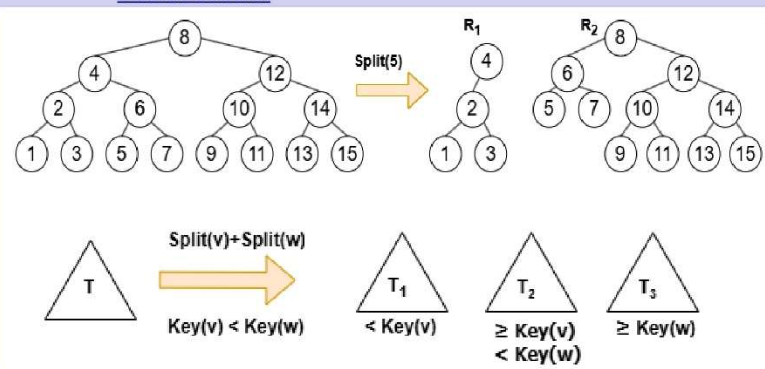
Introducing a lock-based approach for Concurrent Binary Search Trees

- **Maintained under Split/Join Operations**
- **Efficient design: Split operations can combine as they traverse up the Tree**
- **Intermediate step for solving Dynamic Graph problems in a concurrent setting**

Serial Model

Basic Idea: Split(v)

1. v: leaf node (leaf-oriented tree)
2. Keep track of two roots (R_1, R_2) for the constructed trees
3. Initially R_1 := Null, R_2 := v
4. Traverse path from leaf v to root
5. If current node is a left child of its parent p:
   i. Connect R_2 as left child of p
   ii. R_2 := p
6. Else:
   i. Connect R_1 as right child of p
   ii. R_1 := p



Concurrent Model

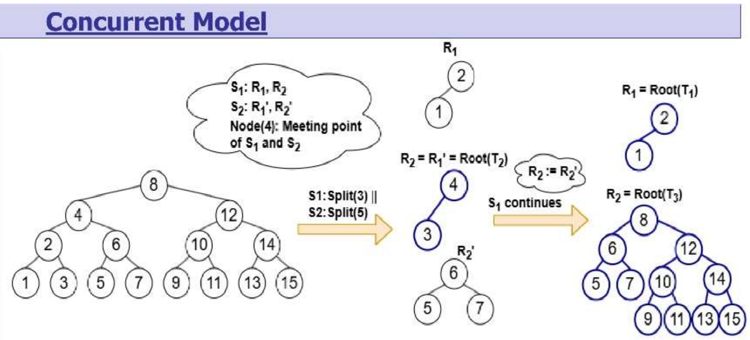
Basic Idea: Split(v)

❑ Similar algorithm to Serial Model

❑ Locking full path from v to root (synchronization)

❑ Combining: 2 Splits meet at node m

- m is the root of T_2
- One Split continues:
  - replaces its R_1 or R_2, with one of the roots from the other Split
  - completes the other Split operation along with its own



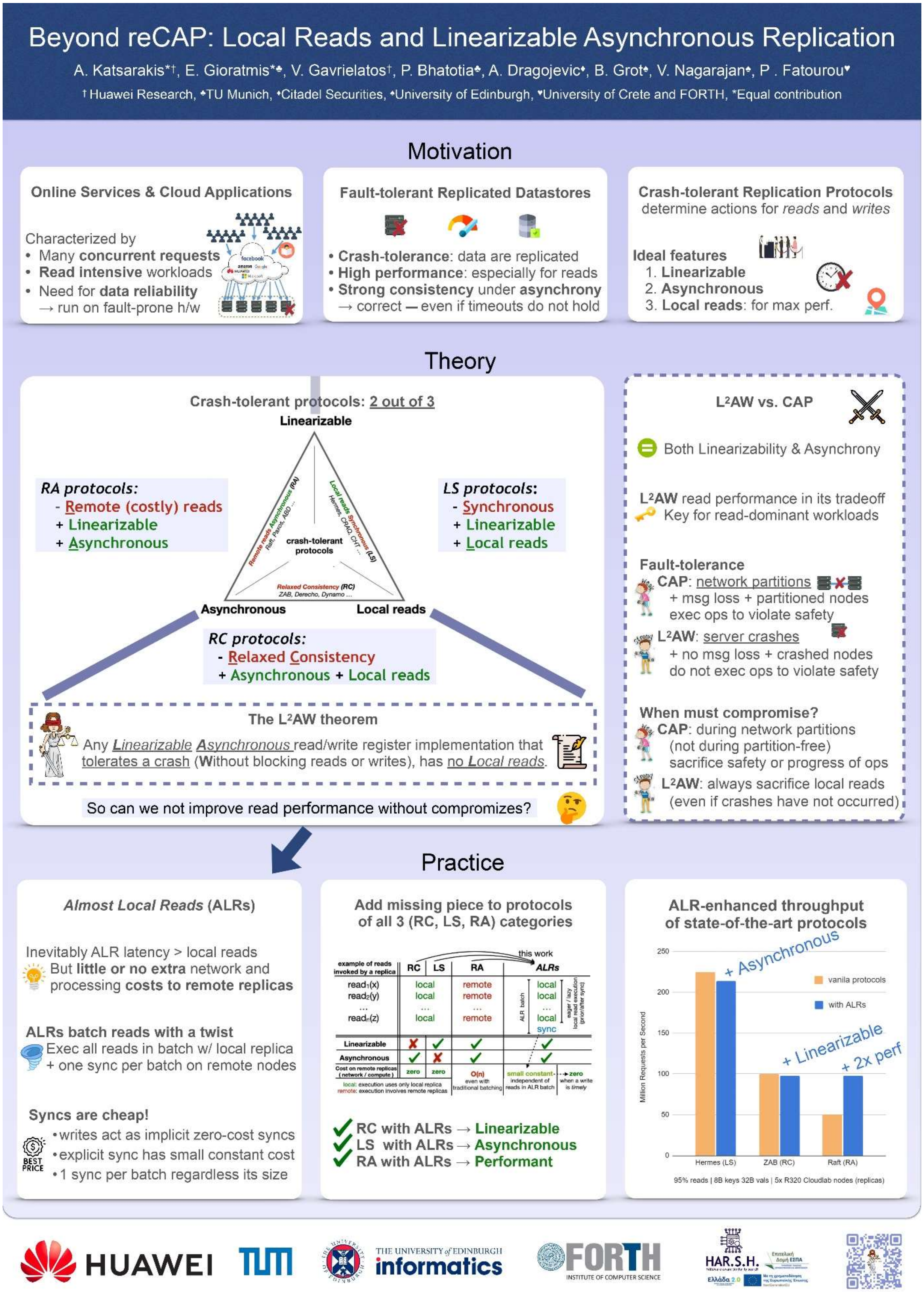
HAR.S.H. (with no. ΥΠ3ΤΑ-0560901) is carried out within the framework of the National Recovery and Resilience Plan "Greece 2.0" with funding from the European Union – NextGenerationEU.

* Thanks to FORTH ICS for the provided support.

HAR.S.H.Partners

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ UNIVERSITY OF CRETE
NEURO LINGO
AMS
mcbs

Contact Information
Nikolaos Grigoroudis
University of Crete, ICS-FORTH
ngrigor@ics.forth.gr


Beyond reCAP: Local Reads and Linearizable Asynchronous Replication
A. Katsarakis*†, E. Gioratmis*♣, V. Gavrielatos†, P. Bhatotia♣, A. Dragojevic•, B. Grot♠, V. Nagarajan♠, P . Fatourou♥
† Huawei Research, ♣TU Munich, •Citadel Securities, ♠University of Edinburgh, ♥University of Crete and FORTH, *Equal contribution
Motivation
Online Services & Cloud Applications
Characterized by
• Many concurrent requests
• Read intensive workloads
• Need for data reliability → run on fault-prone h/w
Fault-tolerant Replicated Datastores
• Crash-tolerance: data are replicated
• High performance: especially for reads
• Strong consistency under asynchrony → correct — even if timeouts do not hold
Crash-tolerant Replication Protocols
determine actions for reads and writes
Ideal features
1. Linearizable
2. Asynchronous
3. Local reads: for max perf.
Theory
Crash-tolerant protocols: 2 out of 3
Linearizable
Asynchronous
Local reads
crash-tolerant protocols
Remote reads Asynchronous (RA) Raft, Paxos, ABD ...
Local reads Synchronous (LS) Hermes, CRAQ, CHT ...
Relaxed Consistency (RC) ZAB, Derecho, Dynamo ...
RA protocols:
- Remote (costly) reads
+ Linearizable
+ Asynchronous
LS protocols:
- Synchronous
+ Linearizable
+ Local reads
RC protocols:
- Relaxed Consistency
+ Asynchronous + Local reads
The L²AW theorem
Any Linearizable Asynchronous read/write register implementation that tolerates a crash (Without blocking reads or writes), has no Local reads.
So can we not improve read performance without compromizes?
L²AW vs. CAP
Both Linearizability & Asynchrony
L²AW read performance in its tradeoff
Key for read-dominant workloads
Fault-tolerance
CAP: network partitions
+ msg loss + partitioned nodes exec ops to violate safety
L²AW: server crashes
+ no msg loss + crashed nodes do not exec ops to violate safety
When must compromise?
CAP: during network partitions (not during partition-free) sacrifice safety or progress of ops
L²AW: always sacrifice local reads (even if crashes have not occurred)
Practice
Almost Local Reads (ALRs)
Inevitably ALR latency > local reads
But little or no extra network and processing costs to remote replicas
ALRs batch reads with a twist
Exec all reads in batch w/ local replica
+ one sync per batch on remote nodes
Syncs are cheap!
• writes act as implicit zero-cost syncs
• explicit sync has small constant cost
• 1 sync per batch regardless its size
BEST PRICE
Add missing piece to protocols of all 3 (RC, LS, RA) categories
this work
example of reads invoked by a replica
RC
LS
RA
ALRs
read₁(x) local remote local
read₂(y) local remote local
readₙ(z) local remote local
sync
ALR batch
eager / lazy local read execution (prior/after sync)
Linearizable ✗ ✓ ✓ ✓
Asynchronous ✓ ✗ ✓ ✓
Cost on remote replicas (network / compute) zero zero O(n) even with traditional batching | small constant independent of reads in ALR batch | zero when a write is timely
local: execution uses only local replica
remote: execution involves remote replicas
RC with ALRs → Linearizable
LS with ALRs → Asynchronous
RA with ALRs → Performant
ALR-enhanced throughput of state-of-the-art protocols
+ Asynchronous
+ Linearizable
+ 2x perf
vanila protocols
with ALRs
Million Requests per Second
250
200
150
100
50
0
Hermes (LS)
ZAB (RC)
Raft (RA)
95% reads | 8B keys 32B vals | 5x R320 Cloudlab nodes (replicas)
HUAWEI
TUM
THE UNIVERSITY of EDINBURGH informatics
FORTH INSTITUTE OF COMPUTER SCIENCE
HAR.S.H.
Ελλάδα 2.0
Επιχειρησιακή Δομή ΕΣΠΑ
Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης NextGenerationEU

HARSH: Hardware-Aware extReme-scale Similarity search

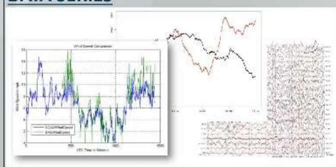
CSD UOC

Coordinator & Scientific Director: Prof. Panagiota Fatourou

Tel.: +30 6973 991277, Email: faturu@csd.uoc.gr

HARSH

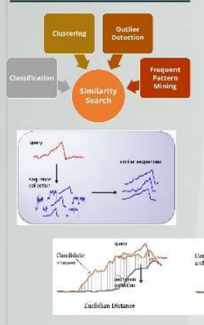
DATA SERIES



Appear in many scientific fields:

- seismology and earth sciences,
- finance
- astrophysics, neuroscience,
- engineering, etc.

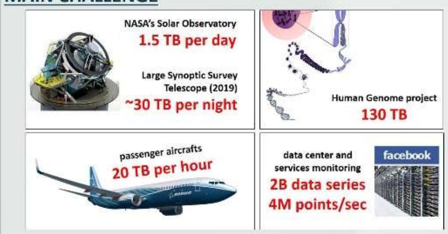
COMPLEX ANALYTICS



Similarity Search

Identify the series from the collection with the smallest distance to the query series

MAIN CHALLENGE



- **Massive data series collections**
- **Data in multiple formats (audio, video, text)**
- **High-dimensional embeddings**

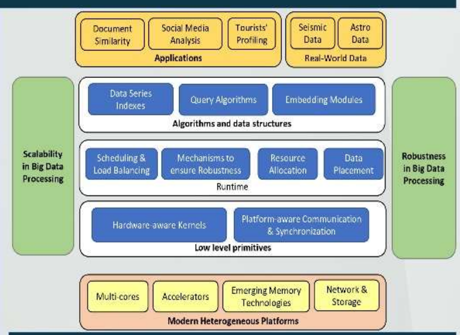
HARSH Objectives and Methodology

OBJECTIVES

1. **Multimodal Learning**
   1. Advanced learning-based models for encoding text and images
   2. Kernels for image to keyword decoding
   3. Hardware-aware methods to accelerate the data processing pipeline
2. **Hardware-Aware/Hardware-Supported Data Series Processing utilizing the full computational power of modern computing platforms in an agnostic way**
   1. Communication and synchronization primitives for hardware-aware computation in heterogeneous platforms
   2. Hardware-Aware Algorithms and Data Structures
   3. Adaptive and agnostic data series processing
3. **Dynamic Data Series Processing**

USE CASES

- **Use case 1 – Similar Document/File Finding Application.**
- **Use Case 2 – Photo Analysis for Travel Profile Enhancement Application.**
- **Use case 3 – Public Opinion Analysis Application.**

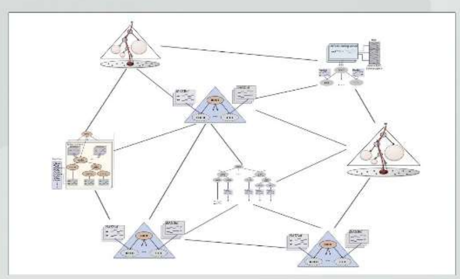


HARSH Innovation & Impact

- Novel deep machine learning algorithms for the automated extraction of key features through low dimensional embeddings
- Robust Hardware-aware Data Processing Systems
- Novel data structures and algorithms for large-scale data-series processing.

Economic Impact

- **Domains that are aggressive data series producers**
  - Construction of EU E-ELT telescope3: 1 billion €
  - EU Human Brain Project5: estimated 1.19 billion €
  - US Human Genome Project has a $796 billion impact and created 310,000 jobs)
- **Relevant Markets**
  - Tourism Market (one of the fastest rising markets in Greece)
  - Natural Language Processing Market
  - Social Media Platforms market



AMS mcbs NEURO LINGO

Ελλάδα 2.0 ΕΘΝΙΚΟ ΣΧΕΔΙΟ ΑΝΑΚΑΜΨΗΣ ΚΑΙ ΑΝΘΕΚΤΙΚΟΤΗΤΑΣ

Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης NextGenerationEU

40 years Computer Science Department University of Crete