

THE UNIVERSITY OF TEXAS AT AUSTIN
Cockrell School of Engineering
Curriculum Vitae

EDUCATION:

Technion, Israel Institute of Tech.	Electrical Engineering	B.Sc. <i>summa cum laude</i>	December 1999
Technion, Israel Institute of Tech.	Physics	B.Sc. <i>summa cum laude</i>	December 1999
Stanford University	Electrical Engineering	MS	June 2002
Stanford University	Electrical Engineering	Ph.D.	January 2007

ACADEMIC POSITIONS:

The University of Texas at Austin	Professor	September 2018 - Present
The University of Texas at Austin	Associate Professor	September 2012 - 2018
The University of Texas at Austin	Assistant Professor	January 2007 - August 2012
Stanford University	Research Assistant	September 1999 - September 2006
Stanford University	Teaching Assistant	March 2000 - June 2002

OTHER PROFESSIONAL EXPERIENCE:

Intel Corporation, Haifa, Israel	Computer Architect	August 1997 - September 1999
Israel Defense Force	Non-commisioned officer	March 1993 - June 1996

HONORS AND AWARDS:

Cullen Trust for Higher Education Endowed Professorship in Engineering	2021 –
Presidential Early Career Award for Scientists and Engineers (PECASE)	2012 (awarded 2014)
DOE Early Career Research Award	June 2012
IEEE Micro TopPicks, January 2012	
IEEE Micro TopPicks, January 2011	
NSF CAREER	March 2010
NVIDIA Corp. Faculty Partnership	October 2009

PUBLICATIONS:

A. Refereed Archival Journal Publications

1. Majid Jalili and Mattan Erez. Managing prefetchers with deep reinforcement learning. *IEEE Computer Architecture Letters*, 21(2):105–108, 2022.
[doi:10.1109/LCA.2022.3210397](https://doi.org/10.1109/LCA.2022.3210397).
2. Esha Choukse, Mattan Erez, and Alaa Alameldeen. CompressPoints: An Evaluation Methodology for Compressed Memory Systems. *IEEE Computer Architecture Letters*, 17(2):126–129, July 2018.
[doi:10.1109/LCA.2018.2821163](https://doi.org/10.1109/LCA.2018.2821163).

3. Zhihao Jia, Yongkee Kwon, Galen Shipman, Pat McCormick, Mattan Erez, and Alex Aiken. A Distributed Multi-GPU System for Fast Graph Processing. *Proceedings of the VLDB Endowment*, 11:297–310, November 2017.
<http://www.vldb.org/pvldb/vol11/p297-jia.pdf>,
[doi:10.14778/3157794.3157799](https://doi.org/10.14778/3157794.3157799).
4. Tomer Morad, Gil Shomron, Mattan Erez, Avinoam Kolodny, and Uri Weiser. Optimizing Read-Once Data Flow in Big-Data Applications. *IEEE Computer Architecture Letters*, 16(1):68–71, January 2017.
[doi:10.1109/LCA.2016.2520927](https://doi.org/10.1109/LCA.2016.2520927).
5. Jaeyoung Park, Tianhao Zheng, Mattan Erez, and Michael Orshansky. Variation-Tolerant Write Completion Circuit for Variable-Energy Write STT-RAM Architecture. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 24(4):1351–1360, April 2016.
[doi:10.1109/TVLSI.2015.2449739](https://doi.org/10.1109/TVLSI.2015.2449739).
6. Marc Snir, Robert W Wisniewski, Jacob A Abraham, Sarita V Adve, Saurabh Bagchi, Pavan Balaji, Jim Belak, Pradip Bose, Franck Cappello, Bill Carlson, Andrew A Chien, Paul Coteus, Nathan A DeBardeleben, Pedro C Diniz, Christian Engelmann, Mattan Erez, Saverio Fazzari, Al Geist, Rinku Gupta, Fred Johnson, Sriram Krishnamoorthy, Sven Leyffer, Dean Liberty, Subhasish Mitra, Todd Munson, Rob Schreiber, Jon Stearley, and Eric Van Hensbergen. Addressing Failures in Exascale Computing. *International Journal of High Performance Computing Applications*, 28(2):129–173, May 2014.
[doi:10.1177/1094342014522573](https://doi.org/10.1177/1094342014522573).
7. Jinsuk Chung, Ikhwan Lee, Michael Sullivan, Jee Ho Ryoo, Dong Wan Kim, Doe Hyun Yoon, Larry Kaplan, and Mattan Erez. Containment Domains: A Scalable, Efficient, and Flexible Resilience Scheme for Exascale Systems. *Scientific Programming*, 21(3):197–212, January 2013.
[doi:10.3233/SPR-130374](https://doi.org/10.3233/SPR-130374).
8. Evgeni Krimer and Mattan Erez. The power of $1 + \alpha$; for memory-efficient bloom filters. *Internet Mathematics*, 7(1):28–44, March 2011.
[doi:10.1080/15427951.2011.560785](https://doi.org/10.1080/15427951.2011.560785).
9. Evgeni Krimer, Isaac Keslassy, Avinoam Kolodny, Isask'har Walter, and Mattan Erez. Static timing analysis for modeling QoS in networks on chip. *Journal of Parallel and Distributed Computing*, 71(5):687–699, May 2011.
[doi:10.1016/j.jpdc.2010.10.003](https://doi.org/10.1016/j.jpdc.2010.10.003).
10. Evgeni Krimer, Robert Pawlowski, Mattan Erez, and Patrick Chiang. Synctium: a Near-Threshold Stream Processor for Energy-Constrained Parallel Applications. *IEEE Computer Architecture Letters*, 9(1):21–24, January 2010.
[doi:10.1109/L-CA.2010.5](https://doi.org/10.1109/L-CA.2010.5).

B. Refereed Conference Proceedings

1. Anyesha Ghosh, Neeraja J. Yadwadkar, and Mattan Erez. Fast and efficient scaling for microservices with SurgeGuard. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC24)*, November 2024.
<https://dl.acm.org/doi/pdf/10.1109/SC41406.2024.00103>,
[doi:https://doi.org/10.1109/SC41406.2024.00103](https://doi.org/10.1109/SC41406.2024.00103).
2. William Ruys, Hochan Lee, Bozhi You, Shreya Talati, Jaeyoung Park, James Almgren-Bell, Yineng Yan, Milinda Fernando, George Biros, Mattan Erez and Martin Burtcher, Christopher J. Rossbach, Keshav Pingali, and Milos Gligoric. Performance characterization of python runtimes for multi-device task parallel programming. In *17th International Symposium on High-Level Parallel Processing (HLPP)*, July 2024.
https://hlpp2024.di.unipi.it/wp-content/uploads/2024/07/HLPP2024_Proceedings.pdf.

3. Zihao Deng, Benjamin Ghaemmaghami, Ashish Kumar Singh, Benjamin Cho, Leo Orshansky, Mattan Erez, and Michael Orshansky. Enhancing cross-category learning in recommendation systems with multi-layer embedding training. In Berrin Yanikoğlu and Wray Buntine, editors, *Proceedings of the 15th Asian Conference on Machine Learning*, volume 222 of *Proceedings of Machine Learning Research*, pages 263–278. PMLR, 11–14 Nov 2024.
<https://proceedings.mlr.press/v222/deng24a.html>.
4. William Ruys, Hochan Lee, Bozhi You, Shreya Talati, Jaeyoung Park, James Almgren-Bell, Yineng Yan, Milinda Fernando, George Biros, Mattan Erez, Martin Burtscher, Christopher J. Rossbach, Keshav Pingali, and Milos Gligoric. A deep dive into task-based parallelism in python. In *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 1147–1149, 2024.
[doi:10.1109/IPDPSW63119.2024.00187](https://doi.org/10.1109/IPDPSW63119.2024.00187).
5. Jeageun Jung and Mattan Erez. Predicting Future-System Reliability with a Component-Level DRAM Fault Model. In *to appear in the Proceedings of the ACM/IEEE International Symposium on Microarchitecture (MICRO)*, 2023.
[doi:10.1145/3613424.3614294](https://doi.org/10.1145/3613424.3614294).
6. Ali Fakhrzadehgan, Prakash Ramrakhiani, Moinuddin Qureshi, and Mattan Erez. SecDDR: Enabling Low-Cost Secure Memories by Protecting the DDR Interface. In *Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2023.
[doi:10.1109/DSN58367.2023.00016](https://doi.org/10.1109/DSN58367.2023.00016).
7. Hochan Lee, William Ruys, Ian Henriksen, Arthur Peters, Yineng Yan, Sean Stephens, Bozhi You, Henrique Fingler, Martin Burtscher, Milos Gligoric, et al. Parla: a Python orchestration system for heterogeneous architectures. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC22)*, 2022.
[doi:10.1109/SC41404.2022.00056](https://doi.org/10.1109/SC41404.2022.00056).
8. Majid Jalili and Mattan Erez. Reducing Load Latency with Cache Level Prediction. In *in the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2022.
[doi:10.1109/HPCA53966.2022.00054](https://doi.org/10.1109/HPCA53966.2022.00054).
9. Amanda Raybuck, Tim Stamler, Wei Zhang, Mattan Erez, and Simon Peter. HeMem: Scalable Tiered Memory Management for Big Data Applications and Real NVM. In *in the proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2021.
[doi:10.1145/3477132.3483550](https://doi.org/10.1145/3477132.3483550).
10. Steven Zhu, Nader Al Awar, Mattan Erez, and Milos Gligoric. Dynamic Generation of Python Bindings for HPC Kernels. In *in the proceedings of the IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2021.
[doi:10.1109/ASE51524.2021.9678726](https://doi.org/10.1109/ASE51524.2021.9678726).
11. Benjamin Y. Cho, Jeageun Jung, and Mattan Erez. Accelerating Bandwidth-Bound Deep Learning Inference with Main-Memory Accelerators. In *in the proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC21)*, nov 2021.
[doi:10.1145/3458817.3476146](https://doi.org/10.1145/3458817.3476146).
12. L. Jaulmes, M. Moreto, M. Valero, M. Erez, and M. Casas. Runtime-Guided ECC Protection using Online Estimation of Memory Vulnerability. In *the proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC20)*, pages 1–14, nov 2020.
<https://doi.ieeecomputersociety.org/10.1109/SC41405.2020.00080>,
[doi:10.1109/SC41405.2020.00080](https://doi.org/10.1109/SC41405.2020.00080).
13. L. Yavits, L. Orosa, S. Mahar, J. D. Ferreira, M. Erez, R. Ginosar, and O. Mutlu. WoLFRAm: Enhancing Wear-Leveling and Fault Tolerance in Resistive Memories using Programmable Address Decoders. In *in the proceedings of the IEEE 38th International Conference on Computer Design (ICCD)*, pages 187–196, 2020.
[doi:10.1109/ICCD50377.2020.00044](https://doi.org/10.1109/ICCD50377.2020.00044).

14. Benjamin Y. Cho, Yongkee Kwon, Sangkug Lym, and Mattan Erez. Near Data Acceleration with Concurrent Host Access. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, 2020.
[doi:10.1109/ISCA45697.2020.00072](https://doi.org/10.1109/ISCA45697.2020.00072).
15. Esha Choukse, Michael B. Sullivan, Mike O'Connor, Mattan Erez, Jeff Pool, David Nellans, and Stephen W. Keckler. Buddy Compression: Enabling Larger Memory for Deep Learning and HPC Workloads on GPUs. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, 2020.
[doi:10.1109/ISCA45697.2020.00080](https://doi.org/10.1109/ISCA45697.2020.00080).
16. Sangkug Lym, Esha Choukse, Siavash Zangeneh, Wei Wen, Sujay Sanghavi, and Mattan Erez. Prune-Train: Fast Neural Network Training by Dynamic Sparse Model Reconfiguration. In *the Proceedings of the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis (SC19)*, 2019.
17. Chun-Kai Chang and Mattan Erez. Assessing The Impact of Timing Errors on HPC Applications. In *the Proceedings of the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis (SC19)*, 2019.
18. Kyushick Lee, Michael Sullivan, Siva Kumar Sastry Hari, Timothy Tsai, Stephen W. Keckler, and Mattan Erez. GPU Snapshot: Checkpoint Offloading for GPU-Dense Systems. In *the Proceedings of the International Conference on Supercomputing (ICS)*, 2019.
[doi:10.1145/3330345.3330361](https://doi.org/10.1145/3330345.3330361).
19. Sangkug Lym, Donghyuk Lee, Mike O'Connor, Niladrish Chatterjee, and Mattan Erez. DeLTA: GPU Performance Model for Deep Learning. In *the Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2019.
[doi:10.1109/ISPASS.2019.00041](https://doi.org/10.1109/ISPASS.2019.00041).
20. Sangkug Lym, Armand Behroozi, Wei Wen, Ge Li, Yongkee Kwon, and Mattan Erez. Mini-batch Serialization: CNN Training with Inter-layer Data Reuse. In *the Proceedings of SysML 2019*, 2019.
<http://www.sysml.cc/doc/2019/142.pdf>.
21. Haishan Zhu, David Lo, Liqun Cheng, Rama Govindaraju, Parthasarathy Ranganathan, and Mattan Erez. Kelp: QoS for Accelerators in Machine Learning Platforms. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, Washington D.C., February 2019.
[doi:10.1109/HPCA.2019.00036](https://doi.org/10.1109/HPCA.2019.00036).
22. Chun-Kai Chang, Sangkug Lym, Nicholas Kelly, Michael B. Sullivan, and Mattan Erez. Evaluating and Accelerating High-Fidelity Error Injection for HPC. In *the Proceedings of the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis (SC18)*, Dallas, TX, November 2018.
[doi:10.1109/SC.2018.00048](https://doi.org/10.1109/SC.2018.00048).
23. Esha Chouke, Alaa Alameldeen, and Mattan Erez. Compresso: Pragmatic Main Memory Compression. In *the Proceedings of the IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Fukuoka, Japan, October 2018.
[doi:10.1109/MICRO.2018.00051](https://doi.org/10.1109/MICRO.2018.00051).
24. Seong-Lyong Gong, Jung-rae Kim, Sangkug Lym, Michael Sullivan, Howard David, and Mattan Erez. DUO: Exposing On-Chip Redundancy to Rank-Level ECC for High Reliability. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–14, Vienna, Austria, February 2018.
[doi:10.1109/HPCA.2018.00064](https://doi.org/10.1109/HPCA.2018.00064).

25. Sangkug Lym, Heonjae Ha, Yongkee Kwon, Chun-Kai Chang, Jungrae Kim, and Mattan Erez. ERUCA: Efficient DRAM Resource Utilization and Resource Conflict Avoidance for Memory System Parallelism. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–14, Vienna, Austria, February 2018.
[doi:10.1109/HPCA.2018.00063](https://doi.org/10.1109/HPCA.2018.00063).
26. Tianhao Zheng, Haishan Zhu, and Mattan Erez. SIPT: Speculatively Indexed, Physically Tagged Caches. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–14, Vienna, Austria, February 2018.
[doi:10.1109/HPCA.2018.00020](https://doi.org/10.1109/HPCA.2018.00020).
27. Zhihao Jia, Yongkee Kwon, Galen Shipman, Pat McCormick, Mattan Erez, and Alex Aiken. A Distributed Multi-GPU System for Fast Graph Processing. *Proceedings of the VLDB Endowment*, 11:297–310, November 2017.
<http://www.vldb.org/pvldb/vol11/p297-jia.pdf>,
[doi:10.14778/3157794.3157799](https://doi.org/10.14778/3157794.3157799).
28. Hana Alam, Tianhao Zheng, Mattan Erez, and Yoav Etsion. Do It Yourself Virtual Memory Translation. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 1–12, Toronto, Canada, June 2017.
[doi:10.1145/3079856.3080209](https://doi.org/10.1145/3079856.3080209).
29. Dong-Wan Kim and Mattan Erez. RelaxFault Memory Repair. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 645–657, Seoul, South Korea, June 2016.
[doi:10.1109/ISCA.2016.62](https://doi.org/10.1109/ISCA.2016.62).
30. Jungrae Kim, Michael Sullivan, Esha Choukse, and Mattan Erez. Bit-Plane Compression: Transforming Data for Better Compression in Many-core Architectures. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 329–340, Seoul, South Korea, June 2016.
[doi:10.1109/ISCA.2016.37](https://doi.org/10.1109/ISCA.2016.37).
31. Jungrae Kim, Michael Sullivan, Sangkug Lym, and Mattan Erez. All-Inclusive ECC: Thorough End-to-End Protection for Reliable Computer Memory. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 622–633, Seoul, South Korea, June 2016.
[doi:10.1109/ISCA.2016.60](https://doi.org/10.1109/ISCA.2016.60).
32. Haishan Zhu and Mattan Erez. Dirigent: Enforcing QoS for Latency-Critical Tasks on Shared Multicore Systems. In *the Proceedings of the ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 33–47, Atlanta, GA, April 2016.
[doi:10.1145/2872362.2872394](https://doi.org/10.1145/2872362.2872394).
33. Seong-Lyong Gong, Minsoo Rhu, Jungrae Kim, Jinsuk Chung, and Mattan Erez. CLEAN-ECC: High Reliability ECC for Adaptive Granularity Memory System. In *the Proceedings of the IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 611–622, Waikiki, HI, December 2015.
[doi:10.1145/2830772.2830799](https://doi.org/10.1145/2830772.2830799).
34. Jungrae Kim, Michael Sullivan, Seong-Lyong Gong, and Mattan Erez. Frugal ECC: Efficient and Versatile Memory Error Protection through Fine-Grained Compression. In *the Proceedings of SC15: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 12:1–12, Austin, TX, November 2015.
[doi:10.1145/2807591.2807659](https://doi.org/10.1145/2807591.2807659).
35. Jungrae Kim, Michael Sullivan, and Mattan Erez. Bamboo ECC: Strong, Safe, and Flexible Codes for Reliable Computer Memory. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 101–112, Burlingame, CA, February 2015.
[doi:10.1109/HPCA.2015.7056025](https://doi.org/10.1109/HPCA.2015.7056025).

36. Dong Li, Minsoo Rhu, Daniel R. Johnson, Mike O'Connor, Mattan Erez, Doug Burger, Donald S. Fussell, and Stephen W. Keckler. Priority-Based Cache Address in Throughput Processors. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 89–100, Burlingame, CA, February 2015.
[doi:10.1109/HPCA.2015.7056024](https://doi.org/10.1109/HPCA.2015.7056024).
37. Dong Wan Kim and Mattan Erez. Balancing Reliability, Cost, and Performance Tradeoffs with FreeFault. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 439–450, Burlingame, CA, February 2015.
[doi:10.1109/HPCA.2015.7056053](https://doi.org/10.1109/HPCA.2015.7056053).
38. Minsoo Rhu, Michael Sullivan, Jingwen Leng, and Mattan Erez. A Locality-Aware Memory Hierarchy for Energy-Efficient GPU Architectures. In *the Proceedings of the IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 86–98, Davis, CA, December 2013.
[doi:10.1145/2540708.2540717](https://doi.org/10.1145/2540708.2540717).
39. Tianhao Zheng, Jaeyoung Park, Michael Orshansky, and Mattan Erez. Variable-Energy Write STT-RAM Architecture with Bit-Wise Write-Completion Monitoring. In *the Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, pages 229–234, Beijing, China, September 2013.
[doi:10.1109/ISLPED.2013.6629299](https://doi.org/10.1109/ISLPED.2013.6629299).
40. Minsoo Rhu and Mattan Erez. Maximizing SIMD Resource Utilization in GPGPUs with SIMD Lane Permutation. In *the Proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 356–367, Tel Aviv, Israel, June 2013.
[doi:10.1145/2485922.2485953](https://doi.org/10.1145/2485922.2485953).
41. Minsoo Rhu and Mattan Erez. The Dual-Path Execution Model for Efficient GPU Control Flow. In *the Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 561–602, Shenzhen, China, February 2013.
[doi:10.1109/HPCA.2013.6522352](https://doi.org/10.1109/HPCA.2013.6522352).
42. Jinsuk Chung, Ikhwan Lee, Michael Sullivan, Jee Ho Ryoo, Dong Wan Kim, Doe Hyun Yoon, Larry Kaplan, and Mattan Erez. Containment Domains: A Scalable, Efficient, and Flexible Resilience Scheme for Exascale Systems. In *the Proceedings of SC12: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 58:1–11, Salt Lake City, UT, November 2012.
[doi:10.1109/SC.2012.36](https://doi.org/10.1109/SC.2012.36).
43. Minsoo Rhu and Mattan Erez. CAPRI: Prediction of Compaction-Adequacy for Handling Control-Divergence in GPGPU Architectures. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 61–71, Portland, OR, June 2012.
[doi:10.1145/2366231.2337167](https://doi.org/10.1145/2366231.2337167).
44. Doe Hyun Yoon, Min Kyu Jeong, Michael B. Sullivan, and Mattan Erez. The dynamic granularity memory system. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 548–559, Portland, OR, June 2012.
[doi:10.1145/2366231.2337222](https://doi.org/10.1145/2366231.2337222).
45. Evgeni Krimer, Patrick Chiang, and Mattan Erez. Lane Decoupling for Improving the Timing-Error Resiliency of Wide-SIMD Architectures. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 237–248, Portland, OR, June 2012.
[doi:10.1145/2366231.2337187](https://doi.org/10.1145/2366231.2337187).
46. Min Kyu Jeong, Chander Sudanthi, Nigel Paver, and Mattan Erez. A QoS-Aware Memory Controller for Dynamically Balancing GPU and CPU Bandwidth Use in an MPSoC. In *the Proceedings of the IEEE Design Automation Conference (DAC)*, pages 855–860, San Francisco, CA, June 2012.
[doi:10.1145/2228360.2228513](https://doi.org/10.1145/2228360.2228513).

47. Min Kyu Jeong, Doe Hyun Yoon, Dam Sunwoo, Michael Sullivan, Ikhwan Lee, and Mattan Erez. Balancing DRAM Locality and Parallelism in Shared Memory CMP Systems. In *the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–12, New Orleans, LA, February 2012.
[doi:10.1109/HPCA.2012.6168944](https://doi.org/10.1109/HPCA.2012.6168944).
48. Robert Pawlowski, Evgeni Krimer, Joseph Crop, Jacob Postman, Nariman Moezzi-Madani, Mattan Erez, and Patrick Chiang. A 530mV 10-Lane SIMD Processor With Variation Resiliency in 45nm SOI. In *the proceedings of the IEEE International Solid State Circuits Conference (ISSCC)*, pages 492–494, San Francisco, CA, February 2012.
[doi:10.1109/ISSCC.2012.6177105](https://doi.org/10.1109/ISSCC.2012.6177105).
49. Doe Hyun Yoon, Min Kyu Jeong, and Mattan Erez. Adaptive Granularity Memory Systems: A Tradeoff between Storage Efficiency and Throughput. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 295–306, San Jose, CA, June 2011.
[doi:10.1145/2000064.2000100](https://doi.org/10.1145/2000064.2000100).
50. Doe Hyun Yoon, Naveen Muralimanohar, Jichuan Chang, Parthasarathy Ranganathan, Norman P. Jouppi, and Mattan Erez. FREE-p: Protecting non-volatile memory against both hard and soft errors. In *the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 466–477, San Antonio, TX, February 2011.
[doi:10.1109/HPCA.2011.5749752](https://doi.org/10.1109/HPCA.2011.5749752).
51. Mehmet Basoglu, Michael Orshansky, and Mattan Erez. NBTI-Aware DVFS: a New Approach To Saving Energy And Increasing Processor Lifetime. In *the proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, pages 253–258, Austin, TX, August 2010.
[doi:10.1145/1840845.1840898](https://doi.org/10.1145/1840845.1840898).
52. Doe Hyun Yoon and Mattan Erez. Virtualized and Flexible ECC for Main Memory. In *the proceedings of the ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 397–408, Pittsburgh, PA, March 2010.
[doi:10.1145/1736020.1736064](https://doi.org/10.1145/1736020.1736064).
53. Doe Hyun Yoon and Mattan Erez. Flexible cache error protection using an ECC FIFO. In *the proceedings of SC09: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 49:1–12, Portland, OR, November 2009.
[doi:10.1145/1654059.1654109](https://doi.org/10.1145/1654059.1654109).
54. Doe Hyun Yoon and Mattan Erez. Memory Mapped ECC: Low-Cost Error Protection for Last Level Caches. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 116–127, Austin, TX, June 2009.
[doi:10.1145/1555754.1555771](https://doi.org/10.1145/1555754.1555771).
55. Tushar Krishna, Amit Kumar, Patrick Chiang, Mattan Erez, and Li-Shiuan Peh. NoC with Near-Ideal Express Virtual Channels Using Global-Line Communication. In *the proceedings of the IEEE Symposium on High-Performance Interconnects (HotI)*, pages 11–20, Stanford, CA, August 2008.
[doi:10.1109/HOTI.2008.22](https://doi.org/10.1109/HOTI.2008.22).
56. Jayanth Gummaraju, Mattan Erez, Joel Coburn, Mendel Rosenblum, and William J. Dally. Architectural Support for the Stream Execution Model on General-Purpose Processors. In *the proceedings of the ACM/IEEE International Conference on Parallel Architectures and Compilation Techniques (PACT)*, pages 3–12, Brasov, Romania, September 2007.
<http://dl.acm.org/citation.cfm?id=1299106>,
[doi:10.1109/PACT.2007.15](https://doi.org/10.1109/PACT.2007.15)
Order of authors on the DOI link is wrong, refer to official publication using link to ACM DL.

57. Mattan Erez, Jung Ho Ahn, Jayanth Gummaraju, Mendel Rosenblum, and William J. Dally. Executing Irregular Scientific Applications on Stream Architectures. In *the proceedings of the ACM International Conference on Supercomputing (ICS)*, pages 93–104, Seattle, WA, June 2007.
[doi:10.1145/1274971.1274987](https://doi.org/10.1145/1274971.1274987).
58. Jung Ho Ahn, William J. Dally, and Mattan Erez. Tradeoff between Data-, Instruction-, and Thread-level Parallelism in Stream Processors. In *the proceedings of the ACM International Conference on Supercomputing (ICS)*, pages 126–137, Seattle, WA, June 2007.
[doi:10.1145/1274971.1274991](https://doi.org/10.1145/1274971.1274991).
59. Timothy Knight, Ji Young Park, Manman Ren, Mike Houston, Mattan Erez, Kayvon Fatahalian, Alex Aiken, William Dally, and Pat Hanrahan. Compilation for Explicitly Managed Memory Hierarchies. In *the proceedings of the ACM International Symposium on the Principles and Practice of Parallel Programming (PPOPP)*, pages 226–236, San Jose, CA, March 2007.
[doi:10.1145/1229428.1229477](https://doi.org/10.1145/1229428.1229477).
60. Kayvon Fatahalian, Timothy J. Knight, Mike Houston, Mattan Erez, Daniel Reiter Horn, Larkhoon Leem, Ji Young Park, Manman Ren, Alex Aiken, William J. Dally, and Pat Hanrahan. Sequoia: programming the memory hierarchy. In *the proceedings of SC06: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 83:1–13, Tampa, FL, November 2006. ACM.
[doi:10.1145/1188455.1188543](https://doi.org/10.1145/1188455.1188543)
The author list was improperly imported by the ACM Digital Library; the author list above is ordered correctly.
61. Jung Ho Ahn, Mattan Erez, and William J. Dally. The design space of data-parallel memory systems. In *the proceedings of SC06: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 80:1–12, Tampa, FL, November 2006. ACM.
[doi:10.1145/1188455.1188540](https://doi.org/10.1145/1188455.1188540).
62. Ulrich Barnhoefer, Moon-Jung Kim, and Mattan Erez. A low power, passively cooled 2000cd/m² hybrid led-lcd display. In *the proceedings of IEEE International Symposium on Consumer Electronics*, pages 1–4, St. Petersburg, Russia, June 2006.
[doi:10.1109/ISCE.2006.1689480](https://doi.org/10.1109/ISCE.2006.1689480).
63. Mattan Erez, Nuwan Jayasena, Timothy J. Knight, and William J. Dally. Fault Tolerance Techniques for the Merrimac Streaming Supercomputer. In *the proceedings of SC05: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 29:1–11, Seattle, WA, November 2005.
[doi:10.1109/SC.2005.26](https://doi.org/10.1109/SC.2005.26).
64. Jung Ho Ahn, Mattan Erez, and William J. Dally. Scatter-Add in Data Parallel Architectures. In *the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 132–142, San Francisco, CA, February 2005.
[doi:10.1109/HPCA.2005.30](https://doi.org/10.1109/HPCA.2005.30).
65. Mattan Erez, Jung Ho Ahn, Ankit Garg, William J. Dally, and Eric Darve. Analysis and Performance Results of a Molecular Modeling Application on Merrimac. In *the proceedings of SC04: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 42:1–10, Pittsburgh, PA, November 2004.
[doi:10.1109/SC.2004.69](https://doi.org/10.1109/SC.2004.69).
66. Nuwan Jayasena, Mattan Erez, Jung Ho Ahn, and William J. Dally. Stream Register Files with Indexed Access. In *the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 60–72, Madrid, Spain, February 2004.
[doi:10.1109/HPCA.2004.10007](https://doi.org/10.1109/HPCA.2004.10007).

67. William J. Dally, Patrick Hanrahan, Mattan Erez, Timothy J. Knight, Francois Labonte, Jung-Ho Ahn, Nuwan Jayasena, Ujval J. Kapasi, Abhishek Das, Jayanth Gummaraju, and Ian Buck. Merrimac: Supercomputing with Streams. In *the proceedings of SC03: the ACM/IEEE International Conference on High-Performance Computing, Networking, Storage, and Analysis*, pages 35:1–8, Phoenix, AZ, November 2003.
[doi:10.1145/1048935.1050187](https://doi.org/10.1145/1048935.1050187)
The author list was improperly imported by the ACM Digital Library; the author list above is ordered correctly.
68. Stephan Jourdan, Lihu Rappoport, Yoav Almog, Mattan Erez, Adi Yoaz, and Ronny Ronen. eXtendedBlock Cache. In *the proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 61–70, Toulouse, France, January 2000.
[doi:10.1109/HPCA.2000.824339](https://doi.org/10.1109/HPCA.2000.824339).
69. Adi Yoaz, Mattan Erez, Ronny Ronen, and Stephan Jourdan. Speculation Techniques for Improving Load Related Instruction Scheduling. In *the proceedings of the IEEE/ACM International Symposium on Computer Architecture (ISCA)*, pages 42–53, Atlanta, GA, May 1999.
[doi:10.1109/ISCA.1999.765938](https://doi.org/10.1109/ISCA.1999.765938).

C. Other Major Publications

1. Doe Hyun Yoon, Naveen Muralimanohar, Jichuan Chang, Parthasarthy Ranganathan, Norman P. Jouppi, and Mattan Erez. FREE-p: A Practical End-to-End Nonvolatile Memory Protection Mechanism. *IEEE Micro TopPicks*, 32(3):79–87, May 2012.
[doi:10.1109/MM.2012.15](https://doi.org/10.1109/MM.2012.15).
2. Doe Hyun Yoon and Mattan Erez. Virtualized ecc: Flexible reliability in main memory. *IEEE Micro*, 31(1):11–19, 2011.
[doi:10.1109/MM.2010.103](https://doi.org/10.1109/MM.2010.103).
3. Tushar Krishna, Amit Kumar, Li-Shiuan Peh, Jacob Postman, Patrick P. Chiang, and Mattan Erez. Express Virtual Channels with Capacitively Driven Global Links. *IEEE Micro*, 29:48–61, August 2009.
[doi:10.1109/MM.2009.64](https://doi.org/10.1109/MM.2009.64).

D. Books, Chapters of Books; Editor of Books

1. Mattan Erez and William J. Dally. Stream processors. In Stephen W. Keckler, Kunle Olukotun, and H. Peter Hofstee, editors, *Multicore Processors and Systems*, chapter 8, pages 231–270. Springer, 2009.
<http://www.springer.com/us/book/9781441902627>.
2. Alan Gatherer, Haishan Zhu, and Mattan Erez. Baseband architectures to support wireless cellular infrastructure: History and future evolution. In Sarah Kate Wilson, Stephen G. Wilson, and Ezio Biglieri, editors, *Academic Press Library in Mobile and Wireless Communications – Transmission Techniques for Digital Communications*, chapter 18, pages 689–705. Elsevier, 2016.
<http://www.sciencedirect.com/science/book/9780123982810>.

E. Reviews

1. Joseph Crop, Evgeni Krimer, Nariman Moezzi-Madani, Robert Pawlowski, Thomas Ruggeri, Patrick Chiang, and Mattan Erez. Error Detection and Recovery Techniques for Variation-Aware CMOS Computing: A Comprehensive Review. *Journal of Low Power Electronics and Applications*, 1(3):334–356, October 2011.
[doi:10.3390/jlpea1030334](https://doi.org/10.3390/jlpea1030334).

F. Technical Reports

1. Michael Sullivan, Ikhwan Lee, Jinsuk Chung, Song Zhang, Seong-Lyong Gong, Derong Liu, Michael LeBeane, Kyushick Lee, and Mattan Erez. Containment Domains Semantics version 0.2. Technical Report Tr-LPH-2014-001, LPH Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, February 2014.

2. Michael Sullivan, Ikhwan Lee, Jinsuk Chung, Song Zhang, Seong-Lyong Gong, Derong Liu, Michael LeBeane, and Mattan Erez. "containment domains semantics version 0.1". Technical Report "TR-LPH-2013-001", "LPH Group, Department of Electrical and Computer Engineering, The University of Texas at Austin", October 2013.
3. Ikhwan Lee, Michael Sullivan, Evgeni Krimer, Dong Wan Kim, Mehmet Basoglu, Doe Hyun Yoon, Larry Kaplan, and Mattan Erez. Survey of error and fault detection mechanisms v2. Technical Report "TR-LPH-2012-001", "LPH Group, Department of Electrical and Computer Engineering, The University of Texas at Austin", December 2012.
4. Ikhwan Lee, Mehmet Basoglu, Michael Sullivan, Doe Hyun Yoon, Larry Kaplan, and Mattan Erez. Survey of error and fault detection mechanisms. Technical Report "TR-LPH-2011-002", "LPH Group, Department of Electrical and Computer Engineering, The University of Texas at Austin", April 2011.
5. Michael Sullivan, Doe Hyun Yoon, and Mattan Erez. Containment domains: A full-system approach to computational resiliency. Technical Report "TR-LPH-2011-001", "LPH Group, Department of Electrical and Computer Engineering, The University of Texas at Austin", January 2011.
6. Mehmet Basoglu and Mattan Erez. Improving multi-core processor energy efficiency and lifetime by embracing variability and wearout. In *the proceedings of the Austin Conference on Integrated Systems and Circuits (ACISC)*, pages 1–5, Austin, TX, October 2009.
7. Evgeni Krimer, Isaac Keslassy, Avinoam Kolodny, Isask'har Walter, and Mattan Erez. Packet-level static timing analysis for nocs. Technical Report CCIT #737, Department of Electrical Engineering, Technion, July 2009.
8. Mattan Erez. *Merrimac – High-Performance, Highly-Efficient Scientific Computing with Streams*. PhD thesis, Stanford University, Stanford, CA, November 2006.
9. Mattan Erez, Brian Towles, and William J. Dally. Spills, Fills, and Kills - An Architecture for Reducing Register-Memory Traffic. Technical Report Concurrent VLSI Architecture (TR-23), Stanford University, July 2000.

G. Workshops

1. William Ruys, Hochan Lee, Bozhi You, Shreya Talati, Jaeyoung Park, James Almgren-Bell, Yineng Yan, Milinda Fernando, George Biros, Mattan Erez, Martin Burtcher, Christopher J. Rossbach, Keshav Pingali, and Milos Gligoric. A deep dive into task-based parallelism in python. In *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 1147–1149, 2024.
[doi:10.1109/IPDPSW63119.2024.00187](https://doi.org/10.1109/IPDPSW63119.2024.00187).
2. Chun-Kai Chang, Guanpeng Li, and Mattan Erez. Evaluating Compiler IR-Level Selective Instruction Duplication with Realistic Hardware Errors. In *the Proceedings of the IEEE/ACM Workshop on Fault Tolerance for HPC at eXtreme Scale (FTXS)*, 2020.
[doi:10.1109/FTXS49593.2019.00010](https://doi.org/10.1109/FTXS49593.2019.00010).

H. Preprints

1. Benjamin Ghaemmaghami, Zihao Deng, Benjamin Cho, Leo Orshansky, Ashish Kumar Singh, Mattan Erez, and Michael Orshansky. Training with multi-layer embeddings for model reduction, 2020.
2. Sangkug Lym and Mattan Erez. Flexsa: Flexible systolic array architecture for efficient pruned dnn model training, 2020.

PATENTS:

1. Adi Yoaz, Ronny Ronen, Lihu Rappoport, Mattan Erez, Stephan Jourdan, Robert Valentine, Cache Memory Bank Access Prediction, US Patent #6,694,421, Issued February 17, 2004.
2. Adi Yoaz, Mattan Erez, Ronny Ronen, System and Method for Early Resolution of Low Confidence Branches and Safe Data Cache Accesses, US Patent #6,697,932, Issued February 24, 2004.
3. Adi Yoaz, Gregory Pribush, Freddy Gabbay, Mattan Erez, Ronny Ronen, Fast Branch Misprediction Recovery Method and System, US Patent #6,757,816, Issued June 29, 2004.
4. Adi Yoaz, Ronny Ronen, Lihu Rappoport, Mattan Erez, Stephan Jourdan, Robert Valentine, Memory Cache Bank Prediction, US Patent #6,880,063, Granted April 12, 2005.
5. Adi Yoaz, Ronny Ronen, Lihu Rappoport, Mattan Erez, Stephan Jourdan, and Robert Valentine, Memory Cache Bank Prediction, US Patent #7,644,236, Granted January 5, 2010.
6. Stephan Jourdan, Adi Yoaz, Mattan Erez, and Ronny Ronen, Method and Apparatus for Predicting Branches Using a Meta Predictor, US Patent #8,285,976, Granted October 9, 2012.
7. Stephan Jourdan, Adi Yoaz, Mattan Erez, and Ronny Ronen, Meta predictor restoration upon detecting misprediction, US Patent #8,572,358, Granted October 29, 2013.
8. Stephan Jourdan, Adi Yoaz, Mattan Erez, and Ronny Ronen, Meta predictor restoration upon detecting misprediction, US Patent #8,943,298, Granted October 29, 2013.
9. Jung Ho Ahn, Mattan Erez, and William J. Dally, Atomic Memory Access Hardware Implementations, US Patent #8,959,292, Granted February 17, 2015.

PH.D. SUPERVISIONS COMPLETED:

Yoon, Doe Hyun	May 2011	ECE	UT Austin
Krimer, Evgeni	May 2012	ECE	UT Austin
Jeong, Min Kyu	December 2012	ECE	UT Austin
Rhu, Minsoo	May 2014	ECE	UT Austin
Sullivan, Michael (co-advised with Earl Swartzlander)	August 2015	ECE	UT Austin
Lee, Ikhwan	August 2015	ECE	UT Austin
Kim, Jungrae	October 2016	ECE	UT Austin
Kim, Dong Wan	September 2017	ECE	UT Austin
Zhu, Haishan	March 2018	ECE	UT Austin
Zheng, Tianhao	May 2018	ECE	UT Austin
Gong, Seong-Lyong	August 2018	ECE	UT Asutin
Choukse, Esha	May 2019	ECE	UT Austin
Lee, Kyushick	August 2019	ECE	UT Austin
Lym, Sangkug	December 2019	ECE	UT Austin
Chang, Chun-Kai	May 2020	ECE	UT Austin
Cho, Benjamin	March 2021	ECE	UT Austin
Kwon, Yongkee	August 2022	ECE	UT Austin
Jalili, Majid	May 2023	ECE	UT Austin

Fakhrzadehgan, Ali	May 2023	ECE	UT Austin
Jung, Jeageun	December 2024	ECE	UT Austin
Ghaemmagemi, Benjamin (<i>co-advised with Michael Orshansky</i>)	May 2025	ECE	UT Austin

BIO AND VITA:

Mattan Erez is a Professor at the Department of Electrical and Computer Engineering at the University of Texas at Austin. His research focuses on improving the performance, efficiency, and scalability of computing systems through advances in memory systems, hardware architecture, software systems, and programming models. His current focus areas are architectures for machine learning, large-scale and high-performance computing, and memory systems. His work aims to improve cooperation across system layers and develop flexible and adaptive mechanisms for proportional resource usage. Mattan received a BSc in Electrical Engineering and a BA in Physics from the Technion, Israel Institute of Technology and his MS and PhD. in Electrical Engineering from Stanford University. He was awarded a Presidential Early Career Award for Scientists and Engineers from President Obama and received an Early Career Research Award from the Department of Energy and an NSF CAREER Award.