References

[1] Ramyad Hadidi, Bahar Asgari, Sam Jijina, Adriana Amyette, Nima Shoghi, and Hyesoon Kim. Quantifying the design-space tradeoffs in autonomous drones. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 661–673, 2021.

[2] Ramyad Hadidi, , Sam Jijina, Jun Chen, Zhen Jiang, Ashutosh Dhekne, and Hyesoon Kim. Capcomcom: Leveraging mobility of autonomous agents for distributed on-the-fly sensor data processing. In *International Conference on Information Processing in Sensor Networks (IPSN), Under Review*, 2022.

[3] Jiashen Cao, Keren Sarkar, Ramyad Hadidi, Joy Arulraj, and Hyesoon Kim. Figo: Fine-grained query optimization in video analytics. In *International Conference on Management of Data (SIGMOD)*, 2022.

[4] Ramyad Hadidi, Jiashen Cao, Matthew Woodward, Michael S Ryoo, and Hyesoon Kim. Distributed perception by collaborative robots. *IEEE Robotics and Automation Letters (RA-L), Invited to IROS*, 3(4):3709–3716, 2018.

[5] Ramyad Hadidi, Jiashen Cao, Michael S Ryoo, and Hyesoon Kim. Toward collaborative inferencing of deep neural networks on internet-of-things devices. *IEEE Internet of Things Journal*, 7(6):4950–4960, 2020.

[6] Ramyad Hadidi, Bahar Asgari, Jiashen Cao, Younmin Bae, Da Eun Shim, Hyojong Kim, Sung-Kyu Lim, Michael S Ryoo, and Hyesoon Kim. LCP: A low-communication parallelization method for fast neural network inference in image recognition. *arXiv preprint arXiv:2003.06464*, 2020.

[7] Ramyad Hadidi, , Jiashen Cao, Michael S Ryoo, and Hyesoon Kim. Reducing inference latency with concurrent architectures for image recognition. *arXiv preprint arXiv:2011.07092*, 2020.

[8] Bahar Asgari, Ramyad Hadidi, Hyesoon Kim, and Sudhakar Yalamanchili. Eridanus: Efficiently running inference of dnns using systolic arrays. *IEEE MICRO*, 39(5):46–54, 2019.

[9] Bahar Asgari, Ramyad Hadidi, Hyesoon Kim, and Sudhakar Yalamanchili. Lodestar: Creating locally-dense cnns for efficient inference on systolic arrays. In *Proceedings of the 56th Annual Design Automation Conference (DAC)*, pages 1–2, 2019.

[10] Ramyad Hadidi, Jiashen Cao, Yilun Xie, Bahar Asgari, Tushar Krishna, and Hyesoon Kim. Characterizing the deployment of deep neural networks on commercial edge devices. In *2019 IEEE International Symposium on Workload Characterization (IISWC)*, pages 35–48. IEEE, 2019.

[11] Matthew L Merck, Bingyao Wang, Lixing Liu, Chunjun Jia, Arthur Siqueira, Qiusen Huang, Abhijeet Saraha, Dongsuk Lim, Jiashen Cao, Ramyad Hadidi, et al. Characterizing the execution of deep neural networks on collaborative robots and edge devices. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (PEARC)*, pages 1–6. 2019.

[12] Bahar Asgari, Ramyad Hadidi, and Hyesoon Kim. Meissa: Multiplying matrices efficiently in a scalable systolic architecture. In *2020 IEEE 38th International Conference on Computer Design (ICCD)*, pages 130–137. IEEE, 2020.

[13] Bahar Asgari, Ramyad Hadidi, and Hyesoon Kim. Proposing a fast and scalable systolic array for matrix multiplication. In *2020 IEEE 28th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, pages 204–204. IEEE, 2020.

[14] Ramyad Hadidi, Jiashen Cao, Michael S Ryoo, and Hyesoon Kim. Robustly executing dnns in iot systems using coded distributed computing. In *Proceedings of the 56th Annual Design Automation Conference (DAC)*, 2019.

[15] Ramyad Hadidi, Jiashen Cao, and Hyesoon Kim. Creating robust deep neural networks with coded distributed computing for iot systems. *arXiv preprint arXiv:2104.04447*, 2021.

[16] Ramyad Hadidi, Nima Shoghi, Bahar Asgari, and Hyesoon Kim. Context-aware task handling in resource-constrained robots with virtualized execution. In *Proceedings of the 58th Annual Design Automation Conference (DAC)*, 2021.

[17] Ramyad Hadidi, Nima Shoghi Ghalehshahi, Bahar Asgari, and Hyesoon Kim. Context-aware task handling in resource-constrained robots with virtualization. *arXiv preprint arXiv:2104.04563*, 2021.

[18] Nima Shoghi, Ramyad Hadidi, Lee Jaewon, Jun Chen, Arthur Siqueria, Rahul Rajan, Shaan Dhawan, Pooya Shoghi, and Kim. Secure location-aware authentication and communication for intelligent transportation systems. *arXiv preprint arXiv:2011.08936*, 2020.

[19] Bahar Asgari, Ramyad Hadidi, Nima Shoghi Ghaleshahi, and Hyesoon Kim. Pisces: Power-aware implementation of slam by customizing efficient sparse algebra. In *Proceedings of the 57th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2020.

[20] Hyojong Kim, Jaewoong Sim, Prasun Gera, Ramyad Hadidi, and Hyesoon Kim. Batch-aware unified memory management in gpus for irregular workloads. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 1357–1370, 2020.

[21] Hyojong Kim, Ramyad Hadidi, Lifeng Nai, Hyesoon Kim, Nuwan Jayasena, Yasuko Eckert, Onur Kayiran, and Gabriel Loh. Coda: Enabling co-location of computation and data for multiple gpu systems. *ACM Transactions on Architecture and Code Optimization (TACO)*, 15(3):1–23, 2018.

[22] Bahar Asgari, Ramyad Hadidi, Tushar Krishna, Hyesoon Kim, and Sudhakar Yalamanchili. Alrescha: A lightweight reconfigurable sparse-computation accelerator. In *2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 249–260. IEEE, 2020.

[23] Bahar Asgari, Ramyad Hadidi, Jiashen Cao, Sung-Kyu Lim, Hyesoon Kim, et al. Fafnir: Accelerating sparse gathering by using efficient near-memory intelligent reduction. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 908–920. IEEE, 2021.

[24] Lifeng Nai, Ramyad Hadidi, Jaewoong Sim, Hyojong Kim, Pranith Kumar, and Hyesoon Kim. Graphpim: Enabling instruction-level pim offloading in graph computing frameworks. In *2017 IEEE International symposium on high performance computer architecture (HPCA)*, pages 457–468. IEEE, 2017.

[25] Ramyad Hadidi, Lifeng Nai, Hyojong Kim, and Hyesoon Kim. Cairo: A compiler-assisted technique for enabling instruction-level offloading of processing-in-memory. *ACM Transactions on Architecture and Code Optimization (TACO)*, 14(4):1–25, 2017.

[26] Lifeng Nai, Ramyad Hadidi, He Xiao, Hyojong Kim, Jaewoong Sim, and Hyesoon Kim. Coolpim: Thermal-aware source throttling for efficient pim instruction offloading. In *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 680–689. IEEE, 2018.

[27] Lifeng Nai, Ramyad Hadidi, He Xiao, Hyojong Kim, Jaewoong Sim, and Hyesoon Kim. Thermal-aware processing-in-memory instruction offloading. *Journal of Parallel and Distributed Computing (JPDC)*, 130:193–207, 2019.

[28] Ramyad Hadidi, Bahar Asgari, Burhan Ahmad Mudassar, Saibal Mukhopadhyay, Sudhakar Yalamanchili, and Hyesoon Kim. Demystifying the characteristics of 3d-stacked memories: A case study for hybrid memory cube. In *2017 IEEE international symposium on Workload characterization (IISWC)*, pages 66–75. IEEE, 2017.

[29] Ramyad Hadidi, Bahar Asgari, Jeffrey Young, Burhan Ahmad Mudassar, Kartikay Garg, Tushar Krishna, and Hyesoon Kim. Performance implications of nocs on 3d-stacked memories: Insights from the hybrid memory cube. In *2018 IEEE international symposium on Performance analysis of systems and software (ISPASS)*, pages 99–108. IEEE, 2018.

[30] Bahar Asgari, Ramyad Hadidi, Joshua Dierberger, Charlotte Steinichen, and Hyesoon Kim. Copernicus: Characterizing the performance implications of compression formats used in sparse workloads. In *2021 IEEE international symposium on Workload characterization (IISWC)*. IEEE, 2021.