# Candidacy Exam: Proposed Syllabus

Davide Giri

## Accelerator Integration in Heterogeneous Architectures

### Why accelerators?

1. Y. S. Shao, D. Brooks, "**Research Infrastructures for Hardware Accelerators**," *Synthesis Lectures on Computer Architecture, Morgan & Claypool*, 2015, chapters 1-2.

### Accelerators taxonomy

2. C. Cascaval, S. Chatterjee, H. Franke, K. J. Gildea, P. Pattnaik, "**A Taxonomy of Accelerator Architectures and their Programming Models**," *IBM Journal of Research and Development*, 2010.
3. A. Parashar, M. Rhu, A. Mukkara, A. Puglielli, R. Venkatesan, B. Khailany, J. Emer, S. W. Keckler, W. J. Dally, "**SCNN: An Accelerator for Compressed-sparse Convolutional Neural Networks**," *International Symposium on Computer Architecture (ISCA)*, 2017.
4. L. Wu, A. Lottarini, T. K. Paine, M. A. Kim, K. A. Ross, "**Q100: The Architecture and Design of a Database Processing Unit**," *Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2014.
5. A. Lottarini, J. P. Cerqueira, T. J. Repetti, S. A. Edwards, K. A. Ross, M. Seok, M. A. Kim, "**Master of None Acceleration: a Comparison of Accelerator Architectures for Analytical Query Processing**," *International Symposium on Computer Architecture (ISCA)*, 2019.

### Accelerators integration

#### Invocation, synchronization and address translation

6. Y. S. Shao, S. L. Xi, V. Srinivasan, G. Wei, D. Brooks, "**Co-designing Accelerators and SoC Interfaces using Gem5-Aladdin**," *International Symposium on Microarchitecture (MICRO)*, 2016.
7. P. Mantovani, E. G. Cota, C. Pilato, G. Di Guglielmo, L. P. Carloni, "**Handling Large Data Sets for High-performance Embedded Applications in Heterogeneous Systems-on-Chip**," *International Conference on Compilers, Architectures, and Synthesis of Embedded Systems (CASES)*, 2016.

8. J. Cong, M. A. Ghodrat, M. Gill, B. Grigorian, G. Reinman, "**Architecture Support for Accelerator-Rich CMPs**," *Design Automation Conference (DAC)*, 2012.
9. Y. Chen, J. Cong, M. A. Ghodrat, M. Huang, C. Liu, B. Xiao, Y. Zou, "**Accelerator-rich CMPs: From Concept to Real Hardware**," *International Conference on Computer Design (ICCD)*, 2013.
10. G. Kornaros, M. Coppola, "**Enabling Efficient Job Dispatching in Accelerator-extended Heterogeneous Systems with Unified Address Space**," *International Symposium on Computer Architecture (ISCA)*, 2018.
11. S. Haria, M. D. Hill, M. M. Swift, "**Devirtualizing Memory in Heterogeneous Systems**," *Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2018.
12. Y. Hao, Z. Fang, G. Reinman, J. Cong, "**Supporting Address Translation for Accelerator-Centric Architectures**," *International Symposium on Computer Architecture (ISCA)*, 2017.

#### Relevant techniques from general purpose computing

13. Y. Fu, T. M. Nguyen, D. Wentzlaff, "**Coherence Domain Restriction on Large Scale Systems**," *International Symposium on Microarchitecture (MICRO)*, 2015.
14. D. Lustig and M. Martonosi, "**Reducing GPU Offload Latency Via Fine-grained CPU-GPU Synchronization**," *International Symposium on High-Performance Computer Architecture (HPCA)*, 2013.

### Memory hierarchy and data movement

15. E. G. Cota, P. Mantovani, G. Di Guglielmo, L. P. Carloni, "**An Analysis of Accelerator Coupling in Heterogeneous Architectures**," *Design Automation Conference (DAC)*, 2015.
16. M. Lyons, M. Hempstead, G. Wei, D. Brooks, "**The Accelerator Store: A Shared Memory Framework for Accelerator-based Systems**," *ACM Transactions on Architecture and Code Optimization (TACO)*, 2012.
17. M. Pellauer, Y. S. Shao, J. Clemons, N. Crago, K. Hegde, R. Venkatesan, S. W. Keckler, C. W. Fletcher, J. Emer, "**Buffets: An Efficient and Composable Storage Idiom for Explicit Decoupled Data Orchestration**," *Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2019.
18. S. Kumar, A. Shriraman, N. Vedula, "**Fusion: Design Tradeoffs in Coherent Cache Hierarchies for Accelerators**," *International Symposium on Computer Architecture (ISCA)*, 2015.
19. J. H. Kelm, D. R. Johnson, W. Tuohy, S. S. Lumetta, S. J. Patel, "**Cohesion: An Adaptive Hybrid Memory Model for Accelerators**," *IEEE Micro*, 2011.

20. C. F. Fajardo, Z. Fang, R. Iyer, G. F. Garcia, S. E. Lee L. Zhao, "**Buffer-Integrated-Cache: A Cost-effective SRAM Architecture for Handheld and Embedded Platforms**," *Design Automation Conference (DAC)*, 2011.
21. M. Peemen, A. A. A. Setio, B. Mesman, H. Corporaal, "**Memory-centric Accelerator Design for Convolutional Neural Networks**," *International Conference on Computer Design (ICCD)*, 2013.

**Relevant techniques from general purpose computing**

22. R. Komuravelli, M. D. Sinclair, J. Alsop, M. Huzaifa, M. Kotsifakou, P. Srivastava, S. V. Adve, V. S. Adve, "**Stash: Have Your Scratchpad and Cache It Too**," *International Symposium on Computer Architecture (ISCA)*, 2015.
23. J. Alsop, M. Sinclair, S. Adve, "**Spandex: A Flexible Interface for Efficient Heterogeneous Coherence**," *International Symposium on Computer Architecture (ISCA)*, 2018.
24. T. J. Ham, J. L. Aragón, M. Martonosi, "**DeSC: Decoupled Supply-compute Communication Management for Heterogeneous Architectures**," *International Symposium on Microarchitecture (MICRO)*, 2015.

**Power management**

25. Paolo Mantovani, Emilio G. Cota, Kevin Tien, Christian Pilato, Giuseppe Di Guglielmo, Ken Shepard, and Luca P. Carloni, "**An FPGA-based Infrastructure for Fine-grained DVFS Analysis in High-performance Embedded Systems**," *Design Automation Conference (DAC)*, 2016.

**Relevant techniques from general purpose computing**

26. A. Vega, A. Buyuktosunoglu, P. Bose, "**Invited paper: Secure swarm intelligence: A new approach to many-core power management**," *International Symposium on Low Power Electronics and Design (ISLPED)*, 2017.