Jeff Pool

jeffpool@gmail.com
http://www.cs.unc.edu/~jpool
Durham, NC

EDUCATION

The University of North Carolina **Chapel Hill, NC**
Department of Computer Science
Ph.D. in Computer Science, May 2012
M.S. in Computer Science, May 2009

The University of South Carolina Honors College **Columbia, SC**
College of Engineering and Information Technology
B.S. in Electrical Engineering, May 2007
Graduated Magna Cum Laude; Dean's and President's Lists.
Active member of Tau Beta Pi Engineering Honor Society, 2005 - 2007.

SKILLS

Strong: C, C++, Python, CUDA, Java, Caffe.
Familiar: Objective-C, Cocoa, iOS SDK, Perl, Torch.
Understanding of state-of-the-art graphics and rendering techniques and architectures, deep learning networks and workloads, and how they fit together.

EXPERIENCE

NVIDIA **Durham, NC**
http://www.nvidia.com
Senior GPU Architect *August 2012 – Present*
- Proposed features and performed their investigations for Maxwell, Pascal, and Volta architectures.
- Led a team looking into new areas of deep learning: research transitioned to the architecture.
- Collaborated with many different teams to test ideas requiring cooperation between SW and HW units.
- Added features to and tests to exercise both functional and performance simulators.
- Served as program committee member for and had work accepted to internal and external conferences.
- Built and maintained an investigation playground for ongoing use by several teams.

GPU Architecture Intern *May – August 2008, 2009, 2011*
Compute Architecture Intern *May – August 2010*
- Enhanced simulators and test environments, building several automation and reporting tools.
- Studied designs for features to be added to the Kepler architecture.
- Studied the performance characteristics of various aspects of the Fermi architecture.

Interactive Data Visualization, Inc. **Lexington, SC**
http://www.idvinc.com
Software Engineering Intern *January 2005 – May 2007*
- Added various plugins (.x3d scene importer, graph/chart types) to a data visualization tool.
- Developed a lightweight tree model browser for artists to use outside of the main CAD interface.

PUBLICATIONS

- H. Mao, S. Han, **J. Pool**, W. Li, X. Liu, Y. Wang, W. Dally, "Exploring the Regularity of Sparse Structure in Convolutional Neural Networks", TMCV workshop at the Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

- S. Han, **J. Pool**, S. Narang, H. Mao, E. Gong, S. Tang, E. Elsen, P. Vajda, M. Paluri, J. Tran, B. Catanzaro, W. Dally, "DSD: Dense-Sparse-Dense Training for Deep Neural Networks," International Conference on Learning Representations (ICLR), April 2017.

- S. Han, **J. Pool**, J. Tran, W. Dally, "Learning both Weights and Connections for Efficient Neural Networks," Neural Information Processing Systems (NIPS), December 2015.

- **J. Pool**, "Energy-Precision Tradeoffs in the Graphics Pipeline," Ph.D. Thesis, The University of North Carolina at Chapel Hill, Department of Computer Science, 2012.

- **J. Pool**, A. Lastra, and M. Singh, "Lossless Compression of Variable-Precision Floating-Point Buffers on GPUs," ACM Interactive 3D Graphics and Games (I3D), 9-11 March 2012.

- **J. Pool**, A. Lastra, and M. Singh, "Precision Selection for Energy-Efficient Pixel Shaders," High Performance Graphics, 5-7 Aug. 2011.

- **J. Pool**, A. Lastra, and M. Singh, "Power-Gated Arithmetic Circuits for Energy-Precision Tradeoffs in Mobile Graphics Processing Units," Journal of Low Power Electronics, Vol. 7, No. 2, 2011.

- **J. Pool**, A. Lastra, and M. Singh, "An Energy Model for Graphics Processing Units," 2010 IEEE International Conference on Computer Design, 3-6 Oct. 2010.

- **J. Pool**, A. Lastra, and M. Singh, "Energy-Precision Tradeoffs in Mobile Graphics Processing Units," 2008 IEEE International Conference on Computer Design, pp.60-67, 12-15 Oct. 2008.

WHITEPAPERS

- M. Rhu, M. O'Connor, N. Chatterjee, **J. Pool**, S. Keckler, "Compressing DMA Engine: Leveraging Activation Sparsity for Training Deep Neural Networks."

- F. Zhu, **J. Pool**, M. Andersch, J. Appleyard, F. Xie, "Sparse Persistent RNNs," Presented at the GPU Technology Conference 2017.