

ECE 598 List of references (Fall 2017)

Introduction

- Shanbhag, Naresh R. "Energy-efficient machine learning in silicon: A communications-inspired approach." In: International Conference on Machine Learning 2016.
- LeCun, Bengio and Hinton, "Deep learning." Nature 2015.
- Michael Jordan and Tom Mitchell, "Machine learning: Trends, perspectives, and prospects." Science 2015
- Dreslinski, Ronald G., et al. "Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits." Proceedings of the IEEE 98.2 (2010): 253-266.
- Shanbhag, book chapter on Algorithms Transforms. Available on the course website.

SGD, ADALINE, LMS

- Leon Bottou. "Large-scale machine learning with stochastic gradient descent". In: Proceedings of COMPSTAT'2010. Springer, 2010, pp. 177–186.
- Leon Bottou. "Stochastic gradient descent tricks". In: Neural Networks: Tricks of the Trade. Springer, 2012, pp. 421–436.
- Adriana Dumitras and George Moschytz. Thinking about thinking: the discovery of the LMS algorithm. 2005.
- Paul J Werbos. "Backpropagation through time: what it does and how to do it". In: Proceedings of the IEEE 78.10 (1990), pp. 1550–1560.
- Bernard Widrow, Marcian E Hoff, et al. "Adaptive switching circuits". In: IRE WESCON convention record. Vol. 4. 1. New York. 1960, pp. 96–104.
- Bernard Widrow and Michael A Lehr. "30 years of adaptive neural networks: perceptron, madaline, and backpropagation". In: Proceedings of the IEEE 78.9 (1990), pp. 1415–1442.
- Shanbhag, book chapter on Adaptive Filters. Available on the course website.
- Shanbhag, Naresh R., and Keshab K. Parhi. "VLSI implementation of a 100 MHz pipelined ADPCM codec chip." *VLSI Signal Processing*, VI, 1993. Available on the course website.
- Tan, Loke Kun, et al. "A 70-Mb/s variable-rate 1024-QAM cable receiver IC with integrated 10-b ADC and FEC decoder." *IEEE Journal of Solid-State Circuits* 33.12 (1998): 2205-2218.

Single Stage Classifiers

- Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: Machine learning 20.3 (1995), pp. 273–297.
- Sakr, Charbel, et al. "Minimum precision requirements for the SVM-SGD learning algorithm." Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE, 2017.
- Kyong Ho Lee and Naveen Verma. "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals". In: IEEE Journal of Solid-State Circuits 48.7 (2013), pp. 1625–1637.

Ensemble Methods

- Leo Breiman. "Random forests". In: Machine learning 45.1 (2001), pp. 5–32.

- J. Ross Quinlan. "Induction of decision trees". In: Machine learning 1.1 (1986), pp. 81–106.
- Robert E Schapire. "The boosting approach to machine learning: An overview". In: Nonlinear estimation and classification. Springer, 2003, pp. 149–171.
- Paul Viola and Michael Jones. "Rapid object detection using a boosted cascade of simple features". In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE. 2001, pp. I–511.
- Yuya Hanai et al. "A versatile recognition processor employing Haar-like feature and cascaded classifier". In: 2009 IEEE International Solid-State Circuits Conference-Digest of Technical Papers. IEEE. 2009, pp. 148–149.
- Warren Rieutort-Louis et al. "A large-area image sensing and detection system based on embedded thin-film classifiers". In: IEEE Journal of Solid-State Circuits 51.1 (2016), pp. 281–290.
- Kang et al. "A 19.4 nJ/Decision 364K Decisions/s In-Memory Random Forest Classifier in 6T SRAM Array". In: 43rd European Solid-State Circuits Conference, 2017." ESSCIRC, 2017.

Deep Learning

- LeCun, Bengio and Hinton, "Deep learning." Nature 2015.
- Yann LeCun et al. "Gradient-based learning applied to document recognition". In: Proceedings of the IEEE 86.11 (1998), pp. 2278–2324.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: Advances in neural information processing systems. 2012, pp. 1097–1105.
- Sze, Vivienne, et al. "Efficient processing of deep neural networks: A tutorial and survey." arXiv preprint arXiv:1703.09039 (2017).
- Suyog Gupta et al. "Deep learning with limited numerical precision". In: arXiv preprint arXiv: 1502.02551 392 (2015).
- Matthieu Courbariaux and Yoshua Bengio. "Binarynet: Training deep neural networks with weights and activations constrained to+ 1 or-1". In: arXiv preprint arXiv: 1602.02830 (2016).
- Charbel Sakr et al. "Analytical Guarantees on Numerical Precision of Deep Neural Networks". In ICML 2017.
- Reagen, Brandon, et al. "Minerva: Enabling low-power, highly-accurate deep neural network accelerators." Proceedings of the 43rd International Symposium on Computer Architecture. IEEE Press, 2016.
- Chen, Tianshi, et al. "Diannao: A small-footprint high-throughput accelerator for ubiquitous machine-learning." Proceedings of the 19th international conference on ASPLOS. ACM, 2014.
- Chen, Yunji, et al. "Dadiannao: A machine-learning supercomputer." Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture. IEEE Computer Society, 2014.
- Du, Zidong, et al. "ShiDianNao: Shifting vision processing closer to the sensor." ACM SIGARCH Computer Architecture News. Vol. 43. No. 3. ACM, 2015.
- Liu, Daofu, et al. "Pudiannao: A polyvalent machine learning accelerator." ACM Proceedings of the Twentieth International Conference on ASPLOS, ACM, 2015.
- Yu-Hsin Chen, Joel Emer, and Vivienne Sze. "Eyeriss: A Spatial Architecture for Energy-Efficient Dataflow for Convolutional Neural Networks". In: International Symposium on Computer Architecture (ISCA). IEEE. 2016, pp. 367–379.

- Jouppi, Norman P., et al. "In-Datcenter Performance Analysis of a Tensor Processing Unit." Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA). ACM, 2017.
- Chen, Yu-Hsin, et al. "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks." IEEE Journal of Solid-State Circuits 52.1 (2017): 127-138.
- Whatmough, Paul N., et al. "A 28nm SoC with a 1.2 GHz 568nJ/prediction sparse deep-neural-network engine with > 0.1 timing error rate tolerance for IoT applications." Solid-State Circuits Conference (ISSCC), 2017 IEEE International. IEEE, 2017.
- Moons, Bert, et al. "14.5 Envision: A 0.26-to-10TOPS/W subword-parallel dynamic-voltage-accuracy-frequency-scalable Convolutional Neural Network processor in 28nm FDSOI." Solid-State Circuits Conference (ISSCC), 2017 IEEE International. IEEE, 2017.

Non von Neumann Computing for Machine Learning

Stochastic and Approximate Computing:

- Han, Jie, and Michael Orshansky. "Approximate computing: An emerging paradigm for energy-efficient design." Test Symposium (ETS), 2013 18th IEEE European. IEEE, 2013.
- Esmaeilzadeh, Hadi, et al. "Neural acceleration for general-purpose approximate programs." Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture. IEEE Computer Society, 2012.
- Venkataramani, Swagath, et al. "AxNN: energy-efficient neuromorphic systems using approximate computing." Proceedings of the 2014 international symposium on Low power electronics and design. ACM, 2014.
- Chippa, Vinay K., et al. "StoRM: a stochastic recognition and mining processor." Proceedings of the 2014 international symposium on Low power electronics and design. ACM, 2014.
- Venkatesan, Rangharajan, et al. "Spintastic: spin-based stochastic logic for energy-efficient computing." Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015. IEEE, 2015.

Error Resilient Computing:

- Ernst, Dan, et al. "Razor: A low-power pipeline based on circuit-level timing speculation." Microarchitecture, 2003. MICRO-36. Proceedings. 36th Annual IEEE/ACM International Symposium on. IEEE, 2003.
- Tschanz, James, et al. "Resilient circuits—Enabling energy-efficient performance and reliability." Computer-Aided Design-Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on. IEEE, 2009.
- Cho, Hyungmin, Larkhoon Leem, and Subhasish Mitra. "ERSA: Error resilient system architecture for probabilistic applications." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 31.4 (2012): 546-558.

Neuromorphic Computing:

- Merolla, Paul A., et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface." Science 345.6197 (2014): 668-673.
- Indiveri, Giacomo, et al. "Neuromorphic silicon neuron circuits." Frontiers in neuroscience 5 (2011).

- Yu, Theodore, and Gert Cauwenberghs. "Analog VLSI biophysical neurons and synapses with programmable membrane channel kinetics." *IEEE Transactions on Biomedical circuits and Systems* 4.3 (2010): 139-148.
- Choudhary, Swadesh, et al. "Silicon neurons that compute." *Artificial neural networks and machine learning–ICANN 2012* (2012): 121-128.

Shannon-inspired Computing:

- Shanbhag, Naresh R., et al. "Stochastic computation." *Proceedings of the 47th Design Automation Conference*. ACM, 2010.
- Hegde, Rajamohana, and Naresh R. Shanbhag. "Soft digital signal processing." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 9.6 (2001): 813-823.
- Shim, Byonghyo, Srinivasa R. Sridhara, and Naresh R. Shanbhag. "Reliable low-power digital signal processing via reduced precision redundancy." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 12.5 (2004): 497-510.
- Eric P Kim et al. "Error Resilient and Energy Efficient MRF Message-Passing-Based Stereo Matching". In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 24.3 (2016), pp. 897–908.
- Zhuo Wang and Naveen Verma. "Enabling hardware relaxations through statistical learning". In: *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE. 2014, pp. 319–326.
- Zhuo Wang, Robert E Schapire, and Naveen Verma. "Error adaptive classifier boosting (EACB): leveraging data-driven training towards hardware resilience for signal inference". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 62.4 (2015), pp. 1136–1145.
- Wang, Zhuo, Kyong Ho Lee, and Naveen Verma. "Overcoming computational errors in sensing platforms through embedded machine-learning kernels." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 23.8 (2015): 1459-1470.
- Rajamohana Hegde and Naresh R Shanbhag. "A voltage overscaled low-power digital filter IC". In: *IEEE Journal of Solid-State Circuits* 39.2 (2004), pp. 388–391.
- Rami A. Abdallah and Naresh R Shanbhag. "An energy-efficient ECG processor in 45-nm CMOS using statistical error compensation". In: *IEEE Journal of Solid-State Circuits* 48.11 (2013), pp. 2882–2893.
- Eric P Kim et al. "A 3.6-mW 50-MHz PN code acquisition filter via statistical error compensation in 180-nm cmos". In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 23.3 (2013), pp. 598–602.

Advanced Topics

Deep In-Memory Architectures

- Kang, Mingu, et al. "An energy-efficient VLSI architecture for pattern recognition via deep embedding of computation in SRAM." *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014.
- M. Kang, S. Gonugondla, A. Patil, and N. Shanbhag, "A 481pJ/decision 3.4M decision/s multifunctional deep in-memory inference processor using standard 6T SRAM array." <https://arxiv.org/abs/1610.07501>.

- Zhang, Jintao, Zhuo Wang, and Naveen Verma. "In-Memory Computation of a Machine-Learning Classifier in a Standard 6T SRAM Array." *IEEE Journal of Solid-State Circuits* 52.4 (2017): 915-924.
- Shin, Dongjoo, et al. "DNPU: An 8.1 TOPS/W reconfigurable CNN-RNN processor for general-purpose deep neural networks." *Solid-State Circuits Conference (ISSCC), 2017 IEEE International. IEEE, 2017.*

Inference using Nanoscale Devices

- Patil et al, "Shannon-inspired Statistical Computing to Enable Spintronics." arXiv 2017.
- Sengupta, Abhronil, Yong Shim, and Kaushik Roy. "Proposal for an All-Spin Artificial Neural Network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets." *IEEE transactions on biomedical circuits and systems* 10.6 (2016): 1152-1160.
- Shafiee et al., "ISAAC- A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars." ISCA 2016.
- Behin-Aein et al., "A building block for hardware belief networks." *Scientific Reports* 2016.
- Camsari, Kerem Yunus, et al. "Stochastic p-bits for Invertible Logic." *Physical Review X* 7.3 (2017): 031014.
- Yu, Shimeng, et al. "Binary neural network with 16 Mb RRAM macro chip for classification and online training." *Electron Devices Meeting (IEDM), 2016 IEEE International. IEEE, 2016.*
- Gao, Ligang, Pai-Yu Chen, and Shimeng Yu. "Demonstration of convolution kernel operation on resistive cross-point array." *IEEE Electron Device Letters* 37.7 (2016): 870-873.
- Li, Haitong, et al. "Hyperdimensional computing with 3D VRRAM in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition." *Electron Devices Meeting (IEDM), 2016 IEEE International. IEEE, 2016.*
- Sheridan, Patrick M., et al. "Sparse coding with memristor networks." *Nature nanotechnology* (2017).