# Course Announcement
# ECE 598 - Machine Learning in Silicon (Fall 2017)

**Instructor:** Naresh R. Shanbhag, http://shanbhag.ece.illinois.edu
**Credit:** 4 hours

**Prerequisites:** ECE 482 or equivalent. Students should be familiar with programming in MATLAB. HDL (VHDL/Verilog) programming experience is desirable.

**Textbook:** Instructor notes and assigned technical papers.

**Time and Place:** 11:00 -12.20, MW, **2074** ECEB

**Course Description**: This course will introduce the design and implementation of machine learning systems on *resource-constrained* platforms that are beginning to find use in emerging sensor-rich applications such as wearables, IoTs, autonomous vehicles, and biomedical devices. The course will begin with preliminaries including motivation and scope of the course; terminology, applications and platforms; taxonomy of inference tasks and learning. Algorithm, architecture and circuit trade-offs to meet desired system performance metrics such as accuracy, latency, throughput, will be studied under severe constraints on precision, memory, computation, and energy. The least mean squared (LMS) algorithm will be employed as a vehicle to understand the issues involved in mapping learning algorithms to architectures and circuits including – algorithmic properties (training, convergence); analytical estimation of bit precision requirements; use of data flow-graph (DFG) descriptors; algorithm-to-architecture mapping using DFG transforms; architectural energy and delay estimation via CMOS circuit models of arithmetic units, memory and interconnect; and case studies of CMOS prototypes of LMS. This path from algorithms-to-architectures-to-circuits will be taken for: single stage classifiers (support vector machine, decision trees), classifier ensembles (random forest, ADAboost), and deep neural networks (DNNs/CNNs). Finally, machine learning on silicon operating at limits of energy efficiency will be studied – properties of low-SNR/low-energy nanoscale fabrics; intrinsic error tolerance of machine learning algorithms; error-resilient computing; inexact computing; Shannon-inspired computing (statistical error compensation (SEC)); and case studies of CMOS implementations. Advanced topics include: emerging cognitive applications; deep in-memory architecture (DIMA); systems on beyond CMOS fabrics.

**Grading:** Course grade will be based on three homework assignments (25%) on the initial part of the course (preliminaries, LMS, single-stage classifiers), one paper presentation (25%), and a course project (50%).

**Instructor Office Hours:**

Wednesdays: 2pm-3pm, 414 CSL

Contact Prof. Shanbhag at shanbhag@illinois.edu, if consultation at a different time is needed.

**Course Web-Page:** http://courses.ece.uiuc.edu/ece598ns/fa2017
**TAs**: Ameya Patil (adpatil2@illinois.edu) and Charbel Sakr (sakr2@illinois.edu);
**TA Office Hours**: Patil, Sakr (2-3PM on Fridays); **ECEB 2036**

## Topical Outline

1. **Preliminaries (hwk1):** motivation and scope of the course; terminology, applications (vision, robotics, biomedical) and platforms (cloud, autonomous, human-centric); taxonomy of inference tasks (prediction, classification, recognition, clustering, density estimation); learning and adaptation (supervised, unsupervised).

2. **The least mean squared (LMS) learning algorithm (hwk2)**: described using the language of machine learning; connection to stochastic gradient descent (SGD) algorithm; applications (prediction, estimation, modeling) for time-series data; precision requirements; data-flow graphs (DFGs) and algorithm-to-architecture mapping via algorithm transforms (retiming, pipelining, block processing, folding, and unfolding); architectural level energy and delay estimation (delay and energy models of CMOS arithmetic units, SRAM, and interconnect); Case studies of CMOS implementations will be presented.

3. **Single stage classifiers** and **ensemble methods (hwk3):** ADALINE; perceptron; support vector machine (SVM); decision trees; random forest; boosting (AdaBoost); supervised learning algorithms; precision requirements; energy efficient architectures via algorithm transforms; Case studies of CMOS implementations;

4. **Deep neural networks (paper presentations):** fully-connected and convolutional networks; training via back-propagation algorithm; precision requirements; DNN architectures (DianNao, Tensor Processing Unit, Eyeriss); Case studies of CMOS implementations;

5. **non von Neumann computing for machine learning:** inherent error-tolerance of ML algorithms; low SNR circuits (near threshold voltage (NTV), all-spin logic, RRAM); error-resilient computing (RAZOR, error detection sequential (EDS)); inexact computing (approximate computing, neuromorphic computing, probabilistic computing); Shannon-inspired computing (statistical error compensation (SEC)); Case studies of CMOS implementations;

6. **The Future:** challenges and opportunities in designing machine learning systems on stochastic and low-SNR circuit fabrics – deep in-memory architecture (DIMA), systems on emerging beyond CMOS fabrics, fundamental limits on energy and reliability, and others.

**Lecture Schedule**

| # | Date | Topic | Remarks |
|---|---|---|---|
| 1 | 28-Aug | Introduction to ECE 598NS | |
| 2 | 30-Aug | The LMS Algorithm and SGD | |
| 3 | 4-Sep | **Labor Day - No Class** | |
| 4 | 6-Sep | Fixed-point Algorithms | |
| 5 | 11-Sep | Algorithm-to-Architecture Mapping Techniques | |
| 6 | 13-Sep | Energy-Delay Trade-offs | TA lecture |
| 7 | 18-Sep | Linear SVM and Case Study | |
| 8 | 20-Sep | Fixed-point, Training & Non-linear SVM | |
| 9 | 25-Sep | Ensemble Methods - ADABoost | |
| 10 | 27-Sep | Case Study - EACB | |
| 11 | 2-Oct | Deep Neural Networks and CNN | |
| 12 | 4-Oct | **No Class** | SONIC annual meeting |
| 13 | 9-Oct | Fixed-point issues and training DNNs | |
| 14 | 11-Oct | Low-complexity DNNs | |
| 15 | 16-Oct | Case Studies - TPU, DianNao | |
| 16 | 18-Oct | Case Studies - Eyeriss, Brooks | TA lecture |
| 17 | 23-Oct | Non von Neumann Architectures - Neurmorphic & Approximate Computing | |
| 18 | 25-Oct | Non von Neumann Architectures - Error-resilient, Stochastic and Hyperdimensional Computing | |
| 19 | 30-Oct | Student presentations | |
| 20 | 1-Nov | Student presentations | |
| 21 | 6-Nov | Student presentations | |
| 22 | 8-Nov | Student presentations | |
| 23 | 13-Nov | Student presentations | |
| 24 | 15-Nov | Student presentations | |
| | | 11/18/2017-11/26/2017 **Thanksgiving Break** | |
| 25 | 27-Nov | Shannon-inspired Computing | |
| 26 | 29-Nov | Shannon-inspired Computing | |
| 27 | 4-Dec | Deep In-memory Computing | |
| 28 | 6-Dec | Deep In-memory Computing | |
| 29 | 11-Dec | Inference on Nanoscale Stochastic Fabrics | |
| 30 | 13-Dec | Project Report & Presentations Due | |