Electrical and Computer Engineering Course Syllabus

Instructor:	Prof. Henry Pfister	**E-mail:**	`henry.pfister@duke.edu`
Office:	Gross Hall 305	**Office Hour:**	TBD
Class Room:	Gross Hall 318	**Class Time:**	MoWe 1:25-2:40pm

Course Name: **ECE 590.17**

Course Title: Factor Graphs and Machine Learning: From Belief-Propagation to Transformers

Prerequisite(s): graduate level course in applied probability

Required Text(s): Course handouts and academic papers

Other Text(s): *Information, Physics, and Computation* by Mezard and Montanari (IPC)
Information Theory, Inference and Learning Algorithms by MacKay (ITILA)
Modern Coding Theory by Richardson and Urbanke (MCT)

Course Objectives:

Many real-world systems can be modeled by a large number of dependent random variables. For such systems, inference problems often require computing the marginal distribution of a small subset of the random variables. Graphical models provide a unifying framework for these systems based on local interactions (i.e., factorization) and lead to efficient algorithms (i.e., belief propagation) for both exact and approximate inference.

On the other hand, one drawback of graphical models is the difficulty of performing exact inference given a general model and the difficulty of learning models of sufficient complexity when given large amounts of training data. For those reasons, machine learning models based on neural networks is often favored over methods based on graphical models. A useful middle ground is given by the fact that many architectures in machine learning can be interpreted as combining learnable parameters with core ideas from graphical models.

This course builds a unified view based on factor graphs and belief propagation that shows how these principles reappear in neural approximations. In particular, we consider algorithm unrolling, multi-scale image networks (U-Nets), graph neural networks (as learned message passing), and Transformers (global kernels as soft messages). Applications include image processing, inference of graph signals, compressed sensing, and error-correcting codes. Some emphasis will be placed understanding the weaknesses of approaches based purely on graphical models and when/why adding learned parameters can improve performance.

At the end of the course, the student should be able to:

1. Describe basic properties of directed and undirected graphical models.
2. Define maximum-likelihood (ML), maximum-a-posteriori (MAP), and a-posteriori-probability (APP) estimation for inference problems.
3. Construct factor-graph representations for standard inference problems.
4. Implement message-passing algorithms for ML, MAP, and APP estimation and understand why they are optimal on tree factor graphs. Compare performance with unrolled and neuralized versions.
5. Implement approximate message passing (AMP) for compressed sensing. Compare with unrolled neuralized versions of the same algorithm.
6. Understand why locally-optimal algorithms (e.g., message passing) are not globally optimal. This will be discussed in general and described precisely for the simple example of LDPC codes on the erasure channel.

7. Understand U-Nets and their connection to message passing. For a simple model of images, implement and compare message-passing with a learned U-Net.
8. Discuss Graph Transformers and their relationships with factor graphs and message passing.
9. Identify situations in engineering, computer science, and statistics where where factor graphs and their neural approximations can be used to obtain useful results.

Student Evaluation:

Midterm	25%	Homework	25%
Final	25%	Project	25%

- Homework will include programming assignments.

Rules and Guidelines:

The class shall follow all established policies of Duke University.