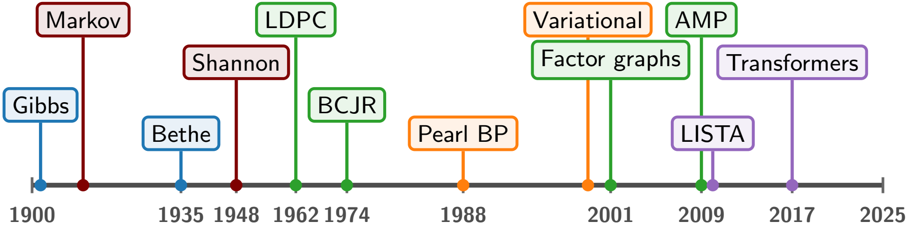
ECE 590.17 Lecture 1: History of Factor Graphs, Inference, and Machine Learning

Duke University, Spring 2026
Instructor: Henry Pfister

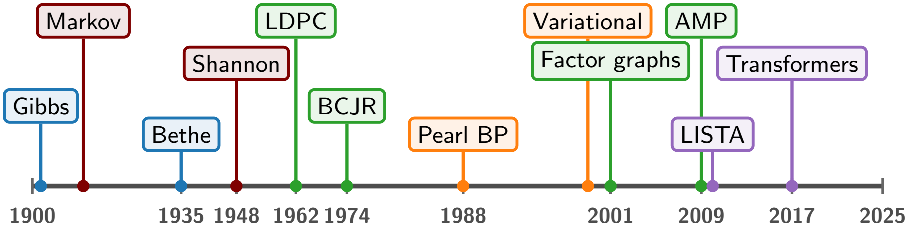
Last Modified: 01/09/2026

Overview



- **Probability and Information:** Bayesian updates, entropy, conditional independence.
- **Statistical physics:** local interactions (Gibbs), tree/cluster approximations (Bethe/Kikuchi), free energy views.
- **Signal processing & coding:** dynamic prog on trellises; iterative decoding (LDPC/turbo), $\Rightarrow$ message passing.

Overview

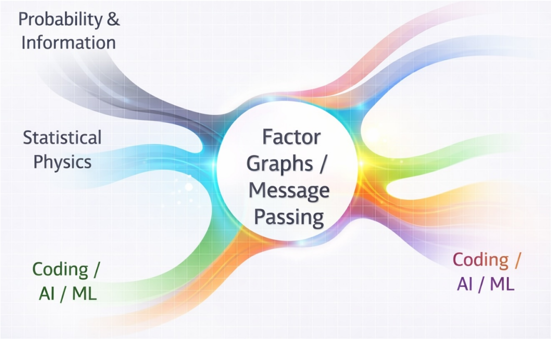


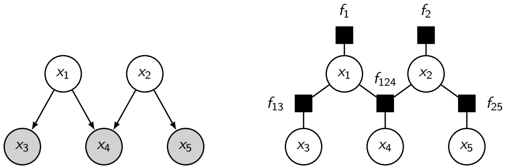
- **Probability and Information:** Bayesian updates, entropy, conditional independence.
- **Statistical physics:** local interactions (Gibbs), tree/cluster approximations (Bethe/Kikuchi), free energy views.
- **Signal processing & coding:** dynamic prog on trellises; iterative decoding (LDPC/turbo), $\Rightarrow$ message passing.
- **AI / graphical models:** Bayes nets, belief propagation, junction trees (exact on trees; approximate on loopy graphs).
- **Modern ML:** learned message passing (GNNs/Transformers), deep unfolding (unrolled inference/optimization).
- **Unifying idea:** Represent global function by *factorization* and pass *local messages*.

Graphical Models Origin Story

- **One idea, many reinventions:** combine local evidence to reason about a global hypothesis.
- **Three parent disciplines:**
  - **Probability & information:** uncertainty, likelihood, entropy, conditional independence.
  - **Statistical physics:** local interactions, tree/cluster approximations, free energy views.
  - **Coding / signal processing:** decoding & filtering as fast dynamic programming on graphs.
- **Modern synthesis:** PGMs + factor graphs + learned message passing.

Unifying lens: make the *factorization* explicit ⇒ compute by *local messages*.





(a) Bayesian Network (b) Factor Graph

From Exact Inference to Learned Message Passing

- **Part I: Trees (exact)**

  sum-product / max-product as dynamic programming; what structure buys you.

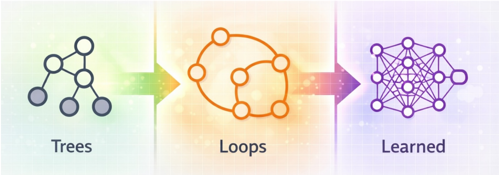
- **Part II: Loops (approximate)**

  loopy BP as a heuristic + variational / free-energy interpretations. Also, MCMC gives sampling alternatives.

- **Part III: Learned (hybrid)**

  AMP + state evolution; deep unfolding (train the iterations); GNNs/Transformers as learned message passing.

Goal: recognize the *same computation template* across inference, optimization, and modern ML.



Early milestones (pre-1982)

Physics & probability

- 1902 **Gibbs**: distributions from local energies (proto-factorization).
- 1906 **Markov**: dependence with local structure (chains $\rightarrow$ HMMs).
- 1925 **Ising**: canonical lattice interaction model (pairwise MRF archetype).
- 1935 **Bethe**: tree-like approximations $\rightarrow$ later Bethe free energy/BP links.
- 1953/1970 **Metropolis–Hastings**: MCMC sampling as computation (inference).

Early milestones (pre-1982)

Physics & probability

- 1902 **Gibbs**: distributions from local energies (proto-factorization).
- 1906 **Markov**: dependence with local structure (chains $\rightarrow$ HMMs).
- 1925 **Ising**: canonical lattice interaction model (pairwise MRF archetype).
- 1935 **Bethe**: tree-like approximations $\rightarrow$ later Bethe free energy/BP links.
- 1953/1970 **Metropolis–Hastings**: MCMC sampling as computation (inference).

Information measures

- 1927 **von Neumann entropy**, 1948 **Shannon entropy**: common currency for inference/learning/coding.

Early milestones (pre-1982)

Physics & probability

- 1902 **Gibbs**: distributions from local energies (proto-factorization).
- 1906 **Markov**: dependence with local structure (chains $\rightarrow$ HMMs).
- 1925 **Ising**: canonical lattice interaction model (pairwise MRF archetype).
- 1935 **Bethe**: tree-like approximations $\rightarrow$ later Bethe free energy/BP links.
- 1953/1970 **Metropolis–Hastings**: MCMC sampling as computation (inference).

Information measures

- 1927 **von Neumann entropy**, 1948 **Shannon entropy**: common currency for inference/learning/coding.

Signal processing & coding

- 1960 **Kalman filter**: recursive Bayesian inference on chains (linear-Gaussian).
- 1967 **Viterbi**: max-product DP on trellis.
- 1974 **BCJR**: sum-product (forward–backward) on a trellis.
- 1962 **Gallager LDPC**: sparse constraints + iterative decoding ideas.
- 1981 **Tanner graphs**: explicit bipartite constraint graph representation.

Early milestones (pre-1982)

Physics & probability

- 1902 **Gibbs**: distributions from local energies (proto-factorization).
- 1906 **Markov**: dependence with local structure (chains $\rightarrow$ HMMs).
- 1925 **Ising**: canonical lattice interaction model (pairwise MRF archetype).
- 1935 **Bethe**: tree-like approximations $\rightarrow$ later Bethe free energy/BP links.
- 1953/1970 **Metropolis–Hastings**: MCMC sampling as computation (inference).

Information measures

- 1927 **von Neumann entropy**, 1948 **Shannon entropy**: common currency for inference/learning/coding.

Signal processing & coding

- 1960 **Kalman filter**: recursive Bayesian inference on chains (linear-Gaussian).
- 1967 **Viterbi**: max-product DP on trellis.
- 1974 **BCJR**: sum-product (forward–backward) on a trellis.
- 1962 **Gallager LDPC**: sparse constraints + iterative decoding ideas.
- 1981 **Tanner graphs**: explicit bipartite constraint graph representation.

Theme: many “special-purpose” algorithms were already message passing on graphs.

Unification: 1983–2001 (graphical models, AI, coding)

AI / probabilistic graphical models

- 1988 **Pearl**: belief propagation as local computation (exact on trees/polytrees).
- 1988 **Lauritzen–Spiegelhalter**: junction tree / clique tree for exact inference.
- 1989–1990 **HUGIN** + local computation architectures: inference MP in software.

Unification: 1983–2001 (graphical models, AI, coding)

AI / probabilistic graphical models

- 1988 **Pearl**: belief propagation as local computation (exact on trees/polytrees).
- 1988 **Lauritzen–Spiegelhalter**: junction tree / clique tree for exact inference.
- 1989–1990 **HUGIN** + local computation architectures: inference MP in software.

Unification: factor graphs & distributive law

- 2000 **Aji–McEliece (GDL)**: one template (semirings) for many graph algorithms.
- 2001 **Kschischang–Frey–Loeliger**: factor graphs + sum-product as a universal inference algorithm.
- 2001 **Forney**: normal realizations / codes on graphs (system-theoretic duality).

Unification: 1983–2001 (graphical models, AI, coding)

AI / probabilistic graphical models

- 1988 **Pearl**: belief propagation as local computation (exact on trees/polytrees).
- 1988 **Lauritzen–Spiegelhalter**: junction tree / clique tree for exact inference.
- 1989–1990 **HUGIN** + local computation architectures: inference MP in software.

Unification: factor graphs & distributive law

- 2000 **Aji–McEliece (GDL)**: one template (semirings) for many graph algorithms.
- 2001 **Kschischang–Frey–Loeliger**: factor graphs + sum-product as a universal inference algorithm.
- 2001 **Forney**: normal realizations / codes on graphs (system-theoretic duality).

Iterative coding revolution

- 1993 **Turbo codes**: iterative decoding near Shannon limits.
- 1996 **MacKay–Neal**: LDPC renaissance with modern BP decoding experiments.
- 1996 **Wiberg**: "codes on graphs" decoding viewpoint for sum/max-product on graphs.
- 1998 **McEliece–MacKay–Cheng**: turbo decoding as Pearl-style BP.

Unification: 1983–2001 (graphical models, AI, coding)

AI / probabilistic graphical models

- 1988 **Pearl**: belief propagation as local computation (exact on trees/polytrees).
- 1988 **Lauritzen–Spiegelhalter**: junction tree / clique tree for exact inference.
- 1989–1990 **HUGIN** + local computation architectures: inference MP in software.

Unification: factor graphs & distributive law

- 2000 **Aji–McEliece (GDL)**: one template (semirings) for many graph algorithms.
- 2001 **Kschischang–Frey–Loeliger**: factor graphs + sum-product as a universal inference algorithm.
- 2001 **Forney**: normal realizations / codes on graphs (system-theoretic duality).

Iterative coding revolution

- 1993 **Turbo codes**: iterative decoding near Shannon limits.
- 1996 **MacKay–Neal**: LDPC renaissance with modern BP decoding experiments.
- 1996 **Wiberg**: "codes on graphs" decoding viewpoint for sum/max-product on graphs.
- 1998 **McEliece–MacKay–Cheng**: turbo decoding as Pearl-style BP.

Takeaway: "BP" becomes a *general algorithm*, not just an AI or ECC decoding trick.

Modern: 2001-2025 (approximate inference, high-dimension, learning)

Loopy BP optimizes variational free-energy

- 2001 **Yedidia–Freeman–Weiss**: BP fixed points $\leftrightarrow$ stationary points Bethe free energy.
- 2001 **EP (Minka)** and 2005 **VMP (Winn–Bishop)**: broader approximate message passing toolkits.
- 2004 **Loeliger**: factor graphs become standard engineering language.

Modern: 2001-2025 (approximate inference, high-dimension, learning)

Loopy BP optimizes variational free-energy

- 2001 **Yedidia–Freeman–Weiss**: BP fixed points $\leftrightarrow$ stationary points Bethe free energy.
- 2001 **EP (Minka)** and 2005 **VMP (Winn–Bishop)**: broader approximate message passing toolkits.
- 2004 **Loeliger**: factor graphs become standard engineering language.

Dense-graph message passing: AMP line

- 2009 **Donoho–Maleki–Montanari**: AMP for compressed sensing.
- 2011 **Bayati–Montanari**: state evolution (rigorous prediction of AMP dynamics).
- 2010+ **Rangan**: GAMP for generalized channels/priors.

Modern: 2001-2025 (approximate inference, high-dimension, learning)

Loopy BP optimizes variational free-energy

- 2001 **Yedidia–Freeman–Weiss**: BP fixed points $\leftrightarrow$ stationary points Bethe free energy.
- 2001 **EP (Minka)** and 2005 **VMP (Winn–Bishop)**: broader approximate message passing toolkits.
- 2004 **Loeliger**: factor graphs become standard engineering language.

Dense-graph message passing: AMP line

- 2009 **Donoho–Maleki–Montanari**: AMP for compressed sensing.
- 2011 **Bayati–Montanari**: state evolution (rigorous prediction of AMP dynamics).
- 2010+ **Rangan**: GAMP for generalized channels/priors.

Modern ML: learned message passing

- **Deep unfolding / unrolling:** 2010 **LISTA** (Gregor–LeCun) and many descendants (train the iterations).
- **Learned BP decoders:** keep the Tanner graph, learn weights/damping/schedules.
- **GNNs:** learned message passing on sparse graphs (graph conv / MPNNs).
- **Transformers:** self-attention as learned message passing on a dense token graph (data-dependent edges).

Modern: 2001-2025 (approximate inference, high-dimension, learning)

Loopy BP optimizes variational free-energy

- 2001 **Yedidia–Freeman–Weiss**: BP fixed points $\leftrightarrow$ stationary points Bethe free energy.
- 2001 **EP (Minka)** and 2005 **VMP (Winn–Bishop)**: broader approximate message passing toolkits.
- 2004 **Loeliger**: factor graphs become standard engineering language.

Dense-graph message passing: AMP line

- 2009 **Donoho–Maleki–Montanari**: AMP for compressed sensing.
- 2011 **Bayati–Montanari**: state evolution (rigorous prediction of AMP dynamics).
- 2010+ **Rangan**: GAMP for generalized channels/priors.

Modern ML: learned message passing

- **Deep unfolding / unrolling:** 2010 **LISTA** (Gregor–LeCun) and many descendants (train the iterations).
- **Learned BP decoders:** keep the Tanner graph, learn weights/damping/schedules.
- **GNNs:** learned message passing on sparse graphs (graph conv / MPNNs).
- **Transformers:** self-attention as learned message passing on a dense token graph (data-dependent edges).

Current synthesis: model-based structure + trainable params $\Rightarrow$ fast stable "neural BP."

The 2D Ising Model in Action (32×32, $J = 1$, $T = 2.3$)