UNIVERSITY OF AMSTERDAM

QuEP: Quantum Expectation Propagation

Max Welling

Microsoft

Evgenii Egorov

UNIVERSITY OF AMSTERDAM

Antonio Rotundo

Leibniz Universität Hannover

Ido Niesen

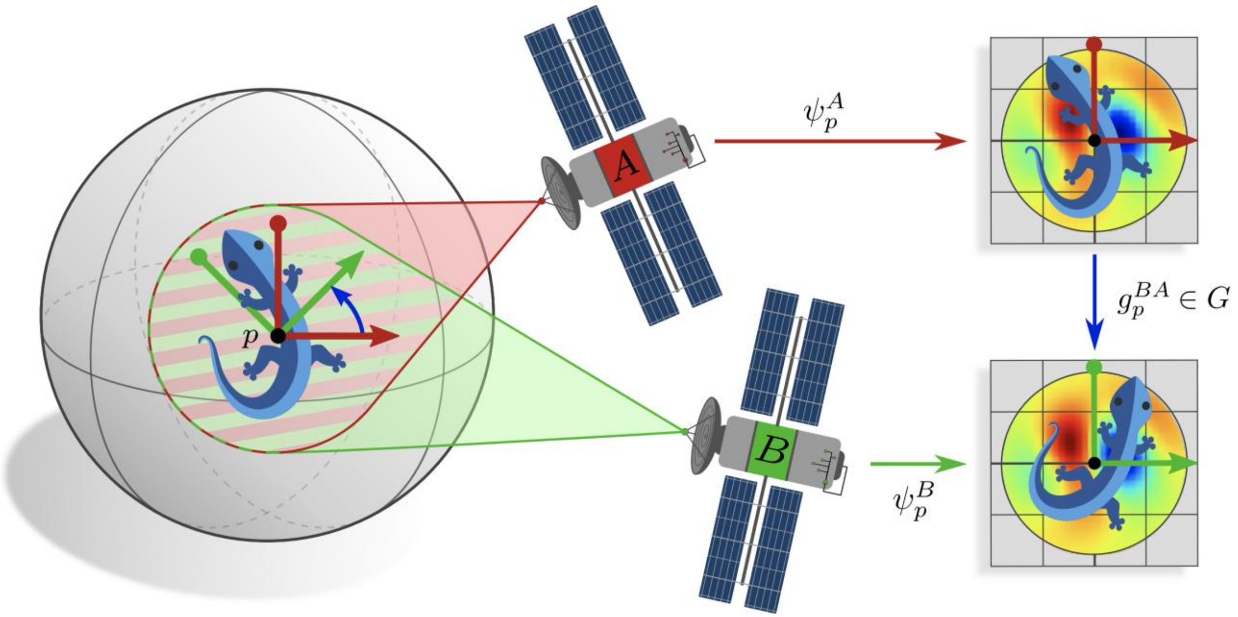
fermioniq

Roberto Bondesan

Imperial College London

Overview

- The Free Energy
- Symmetries & QEC
- Classical BP & GBP
- Quantum BP & GBP
- The Hinton
- Conclusion



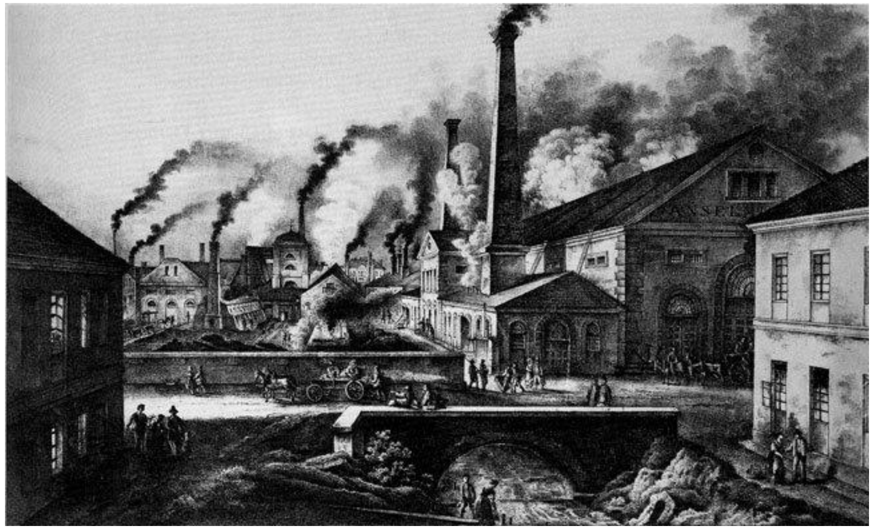
Picture created by Maurice Weiler

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

Free Energy = Energy - Entropy

Ability to perform physical work

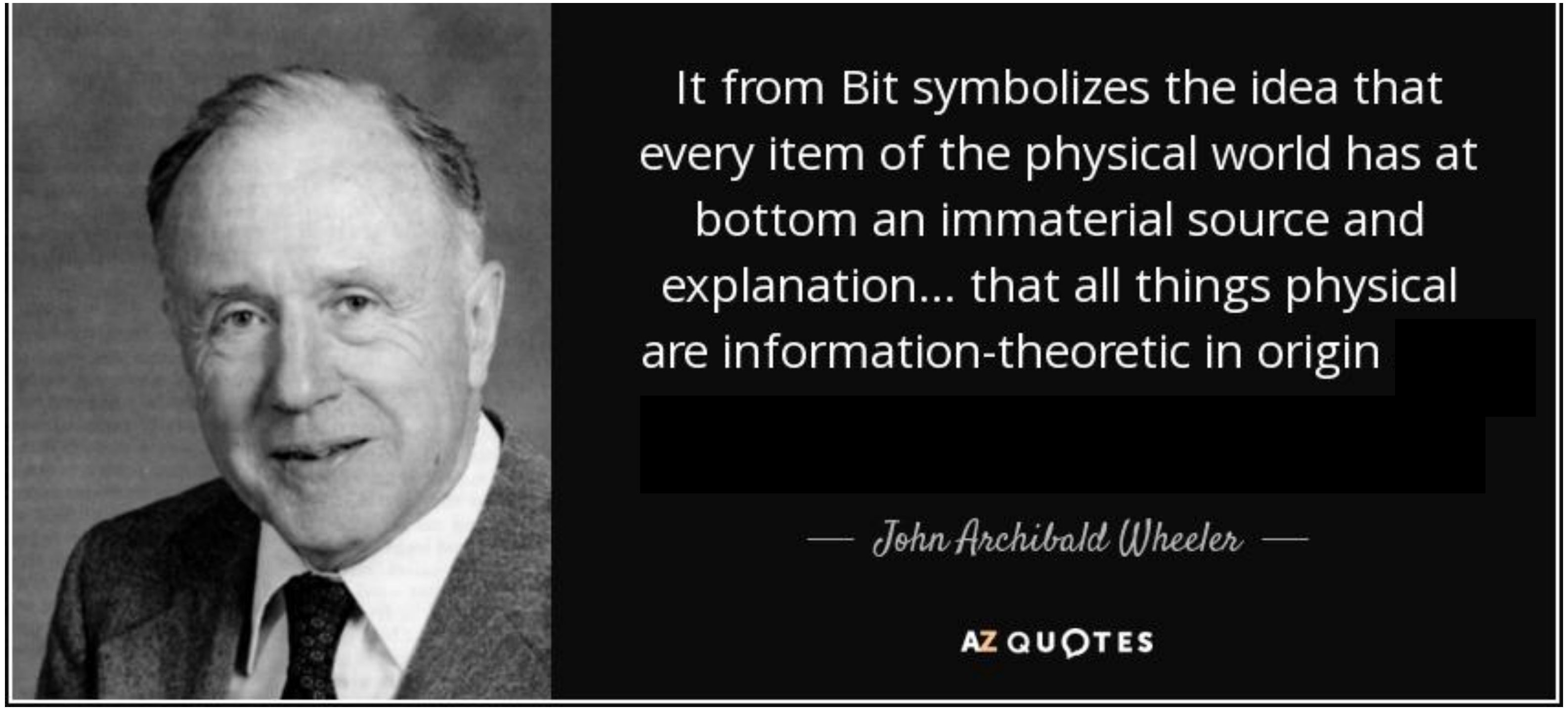
Level of organization, information of a system



Industrial Revolution: ±1820

Information Revolution: ±1940

"It From Bit"


It from Bit symbolizes the idea that every item of the physical world has at bottom an immaterial source and explanation... that all things physical are information-theoretic in origin [redacted]
— John Archibald Wheeler —
AZ QUOTES

E.T. Jaynes

PHYSICAL REVIEW VOLUME 106, NUMBER 4 MAY 15, 1957

Information Theory and Statistical Mechanics

E. T. Jaynes
Department of Physics, Stanford University, Stanford, California

(Received September 4, 1956; revised manuscript received March 4, 1957)

Entropy is ***our*** *degree of ignorance about the microscopic degrees of freedom of a system (same as in AI)*

Free Energies in Physics/Chemistry and ML

$$-\log P_X \leq -\mathbb{E}_{Q_{Z|X}}(\log P_{X,Z}) - S(Q_{Z|X})$$

(Evidence Lower Bound: ELBO)

$$KL(Q||P) \geq 0$$

$$F = -T\log Z \leq \mathcal{F}(Q) = \mathbb{E}_Q(H) - TS(Q)$$

(Variational Free Energy)

$$P = \frac{1}{Z}e^{-H/T}$$

(Boltzmann Distribution)

ML

Physics/chemistry

ML as Nonequilibrium Thermodynamics

$$-\log P_X \leq -\mathbb{E}_{Q_{Z|X}}(\log P_{X,Z}) - S(Q_{Z|X})$$

$$F = -T\log Z \leq \mathcal{F}(Q) = \mathbb{E}_Q(H) - TS(Q)$$

Expectation Maximization:

E-step: Update Q to minimize Bound

M-step: Update P(X,Z) to minimize Bound

Nonequilibrium Thermodynamics:

Heat: Relax Q to minimize F

Work: Change H at fixed Q

Generative AI as Nonequilibrium Statistical Mechanics

2015

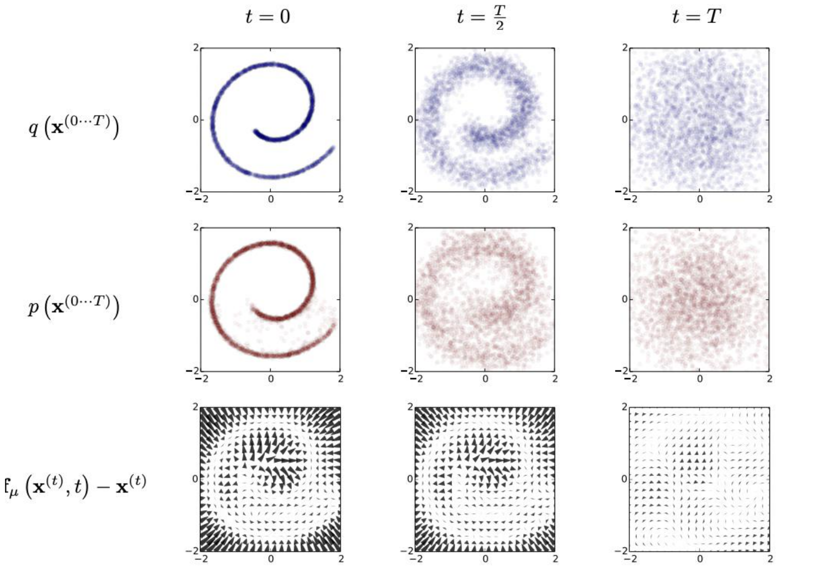
Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EAWEISS@BERKELEY.EDU
University of California, Berkeley

Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University



Example of a VAE: Diffusion Based Models

2021

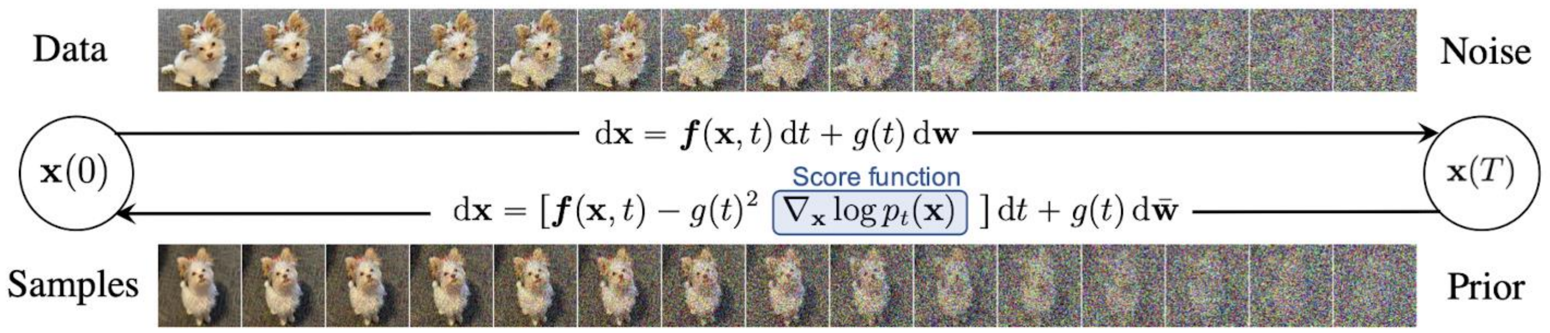
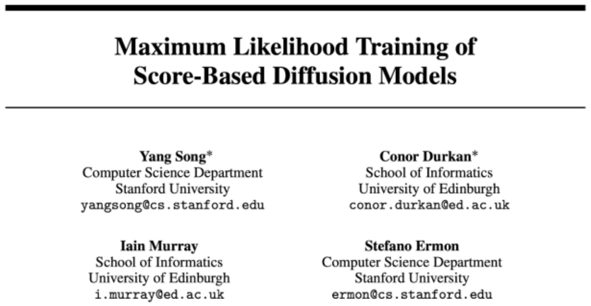


Figure 1: We can use an SDE to diffuse data to a simple noise distribution. This SDE can be reversed once we know the score of the marginal distribution at each intermediate time step, $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$.



Maximum Likelihood Training of Score-Based Diffusion Models

Yang Song* — Computer Science Department, Stanford University, yangsong@cs.stanford.edu

Conor Durkan* — School of Informatics, University of Edinburgh, conor.durkan@ed.ac.uk

Iain Murray — School of Informatics, University of Edinburgh, i.murray@ed.ac.uk

Stefano Ermon — Computer Science Department, Stanford University, ermon@cs.stanford.edu

Generative AI: Images

Generative AI: Art

Artist Wins Photography Contest After Submitting AI-Generated Image, Then Forfeits Prize

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

Generative AI: Videos

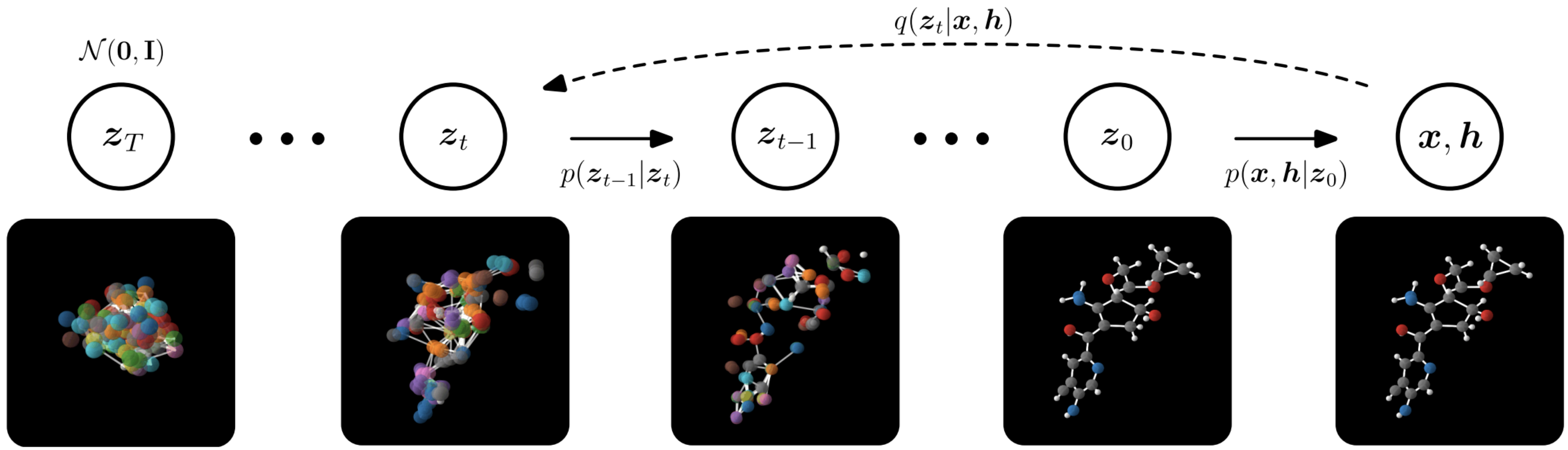
"A shot following a hiker through jungle brush."

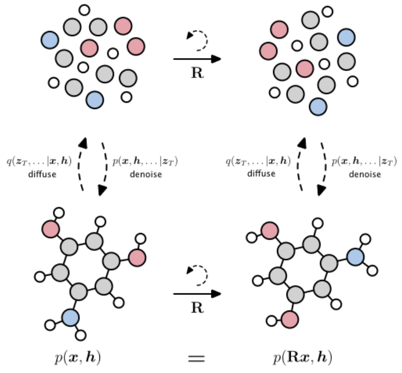
"An aerial shot of a mountain landscape."

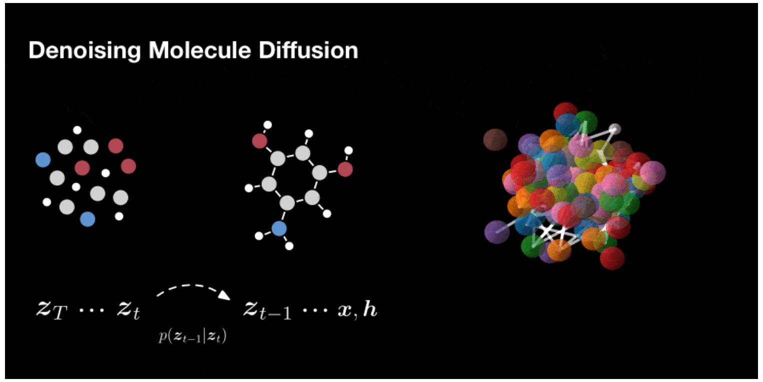
Generative AI: Molecules!

Equivariant Diffusion for Molecule Generation in 3D

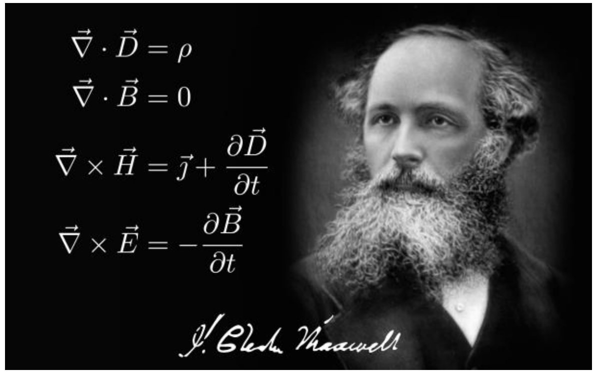
Emiel Hoogeboom [*1] Victor Garcia Satorras [*1] Clément Vignac [*2] Max Welling [1]



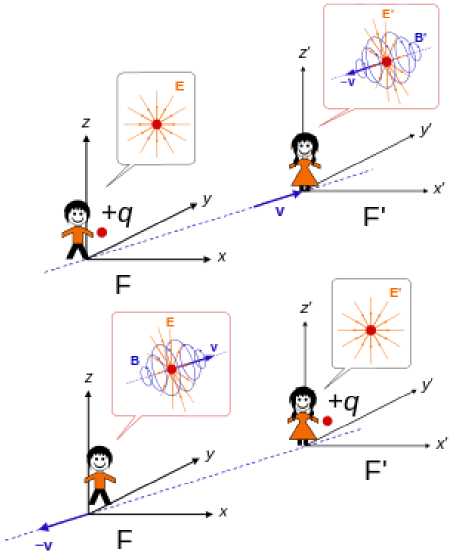




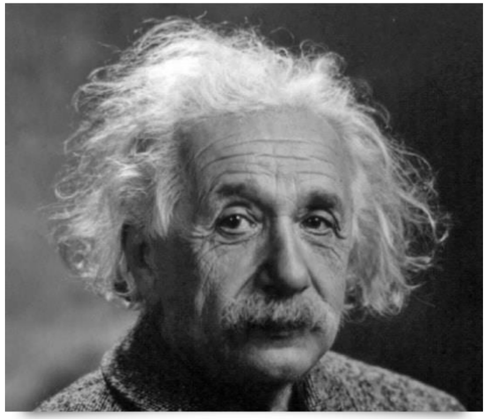
Symmetries & Equivariance



Electricity = Magnetism

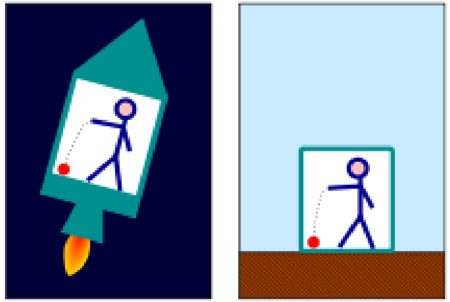


- Lead to **Special Relativity**: electric field = magnetic field.
- Lead to **General Relativity**: gravity = acceleration.
- Led to **Standard Model** of elementary particles!

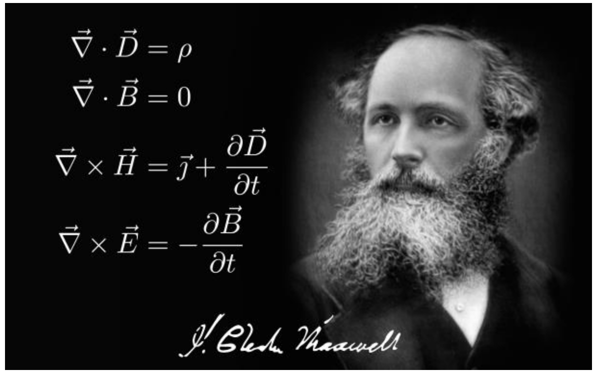


Gravity = Acceleration

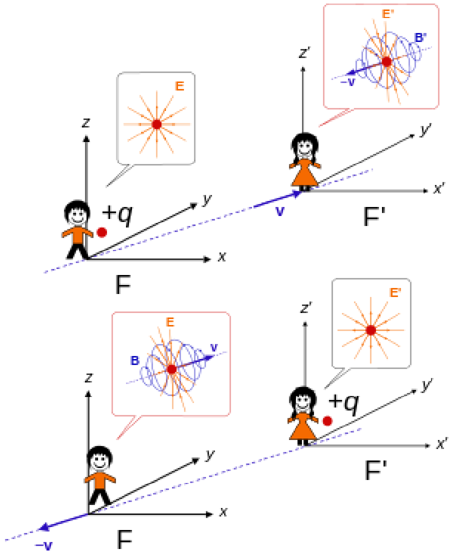
$$R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} + \Lambda\, g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

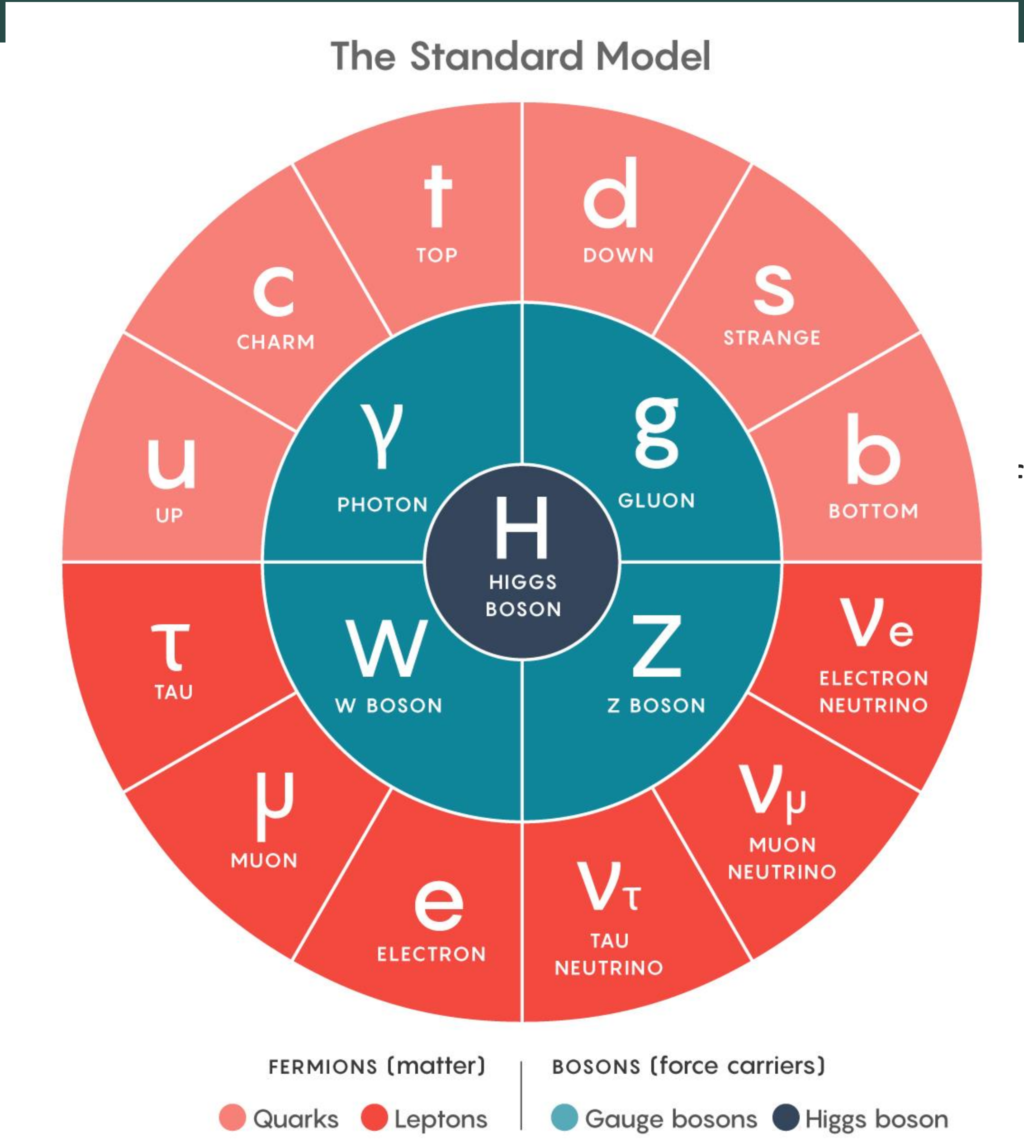


Symmetries & Equivariance

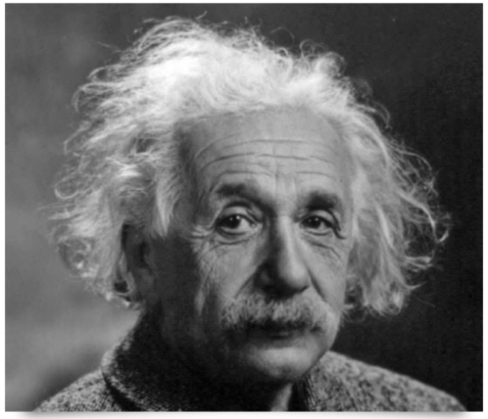


Electricity = Magnetism



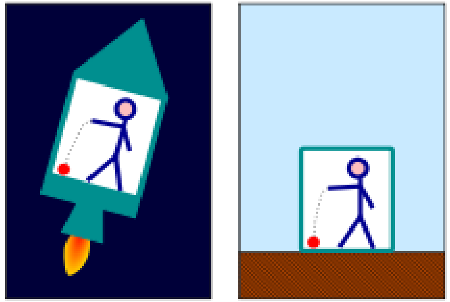


- ield.
-
-

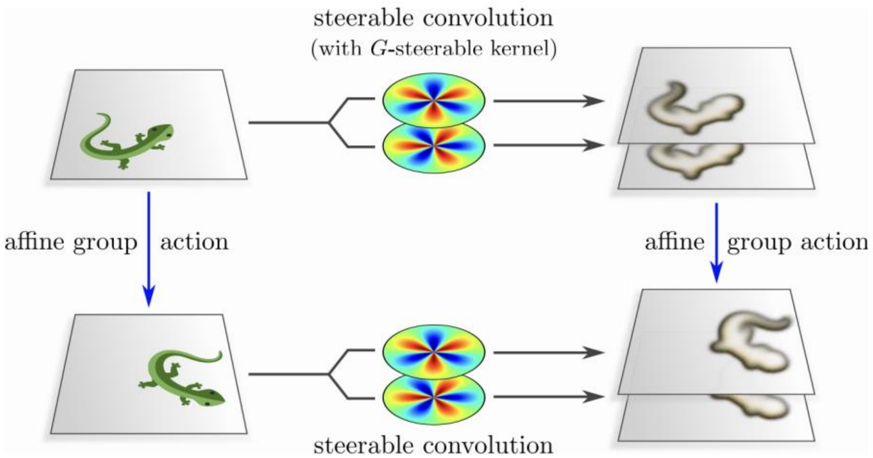


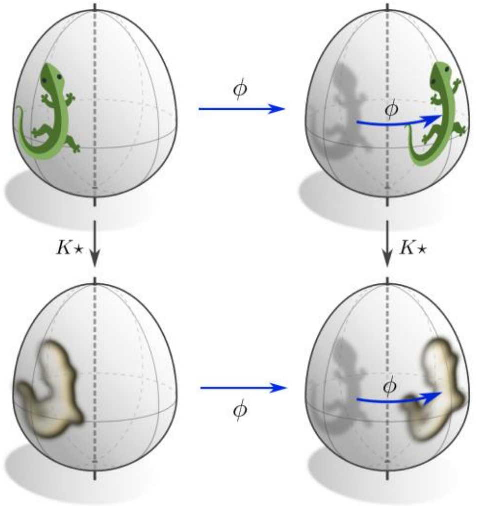
Gravity = Acceleration

$$R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} + \Lambda\, g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

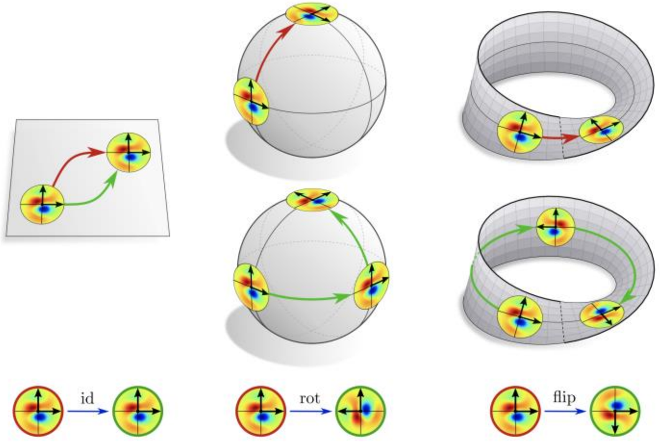


Equivariance





Equivariance on manifold



Gauge symmetries are needed to define proper convolutions on manifolds

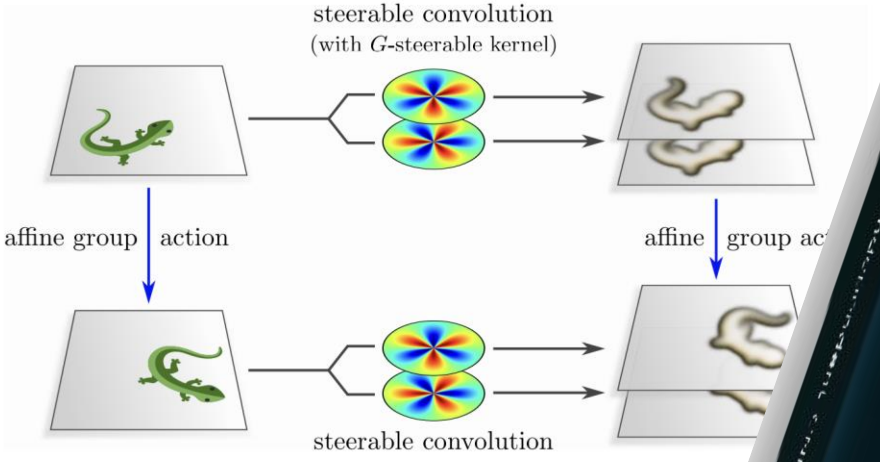
- Equivariance is good for:
  - Data efficiency
  - Disentangling pose and presence
  - Creates easy patterns for next layer

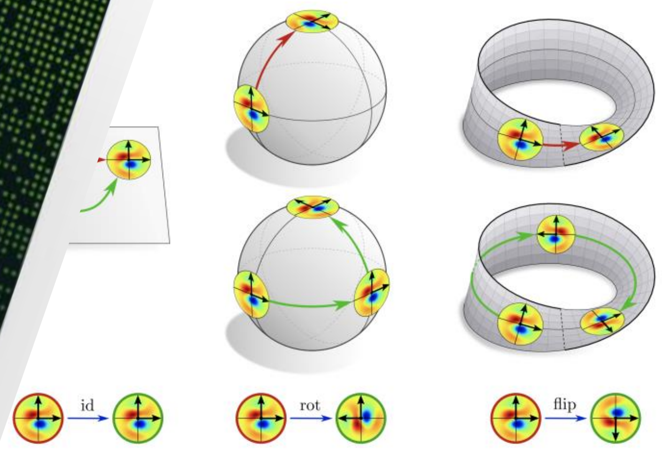
- First appearance in ML: Group CNNs
  Cohen & W. '16, Dieleman et al, '16

Picture created by Maurice Weiler

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

Equivariance

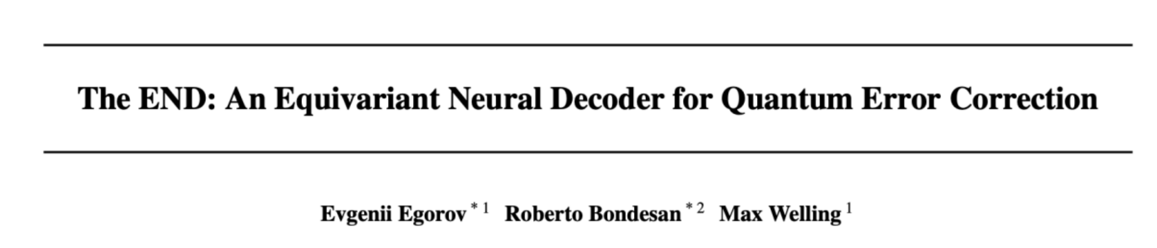




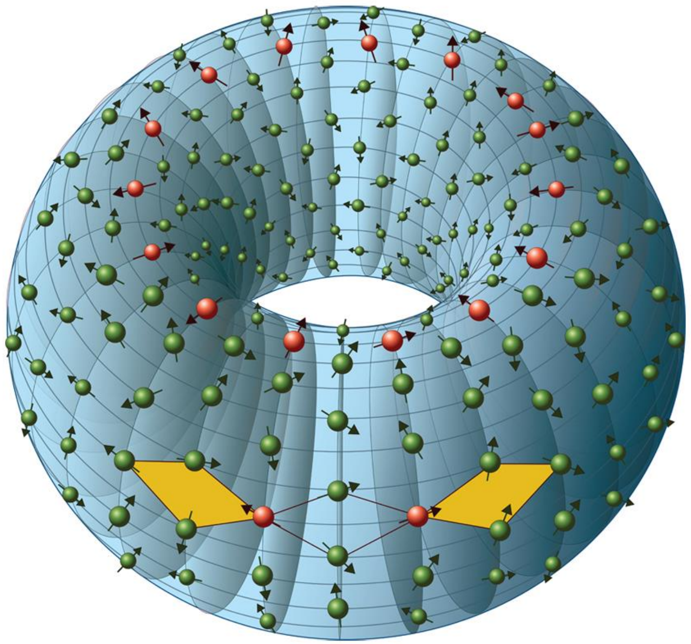
Gauge symmetries are needed to define proper convolutions on manifolds

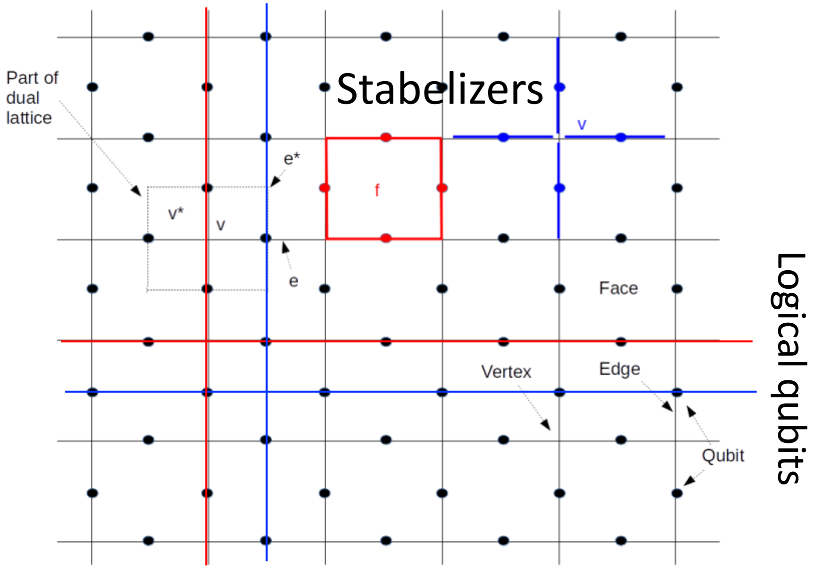
- Equivariance is good for:
  - Data efficiency
  - Disentangling pose and
  - Creates easy patterns fo
- First appearance in ML: Group CNNs
  Cohen & W. '16, Dieleman et al, '16

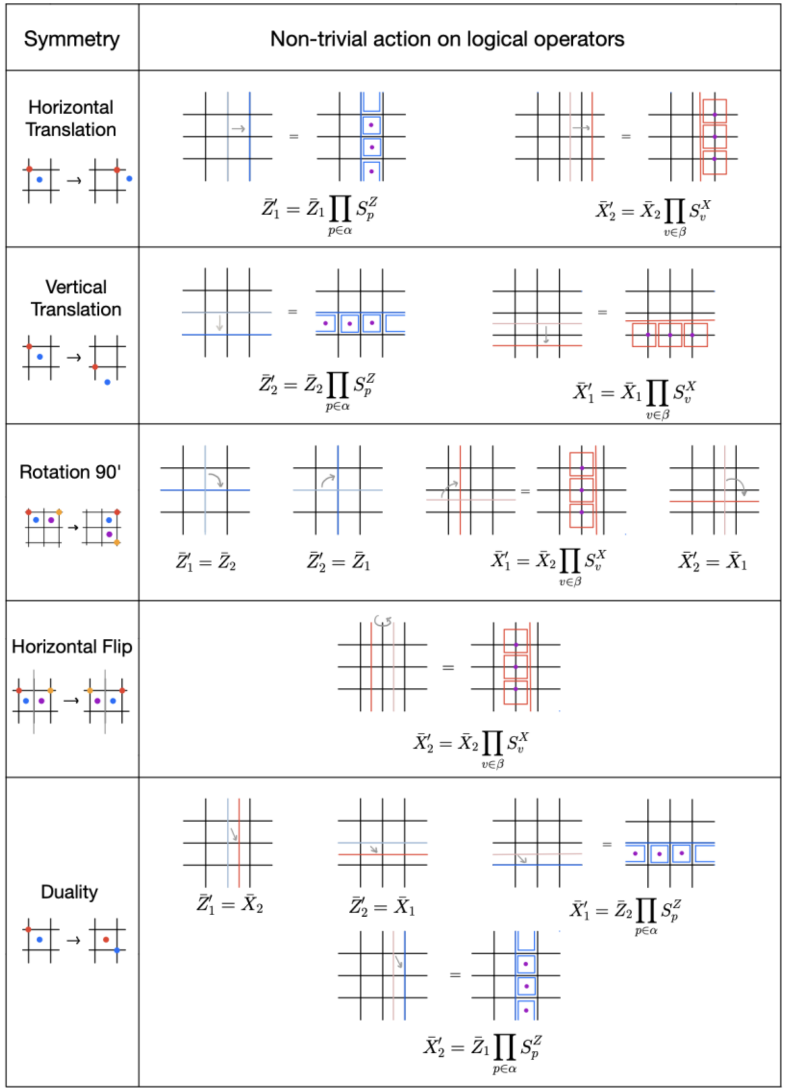
Quantum Error Correction


The END: An Equivariant Neural Decoder for Quantum Error Correction

Evgenii Egorov [*1] Roberto Bondesan [*2] Max Welling [1]

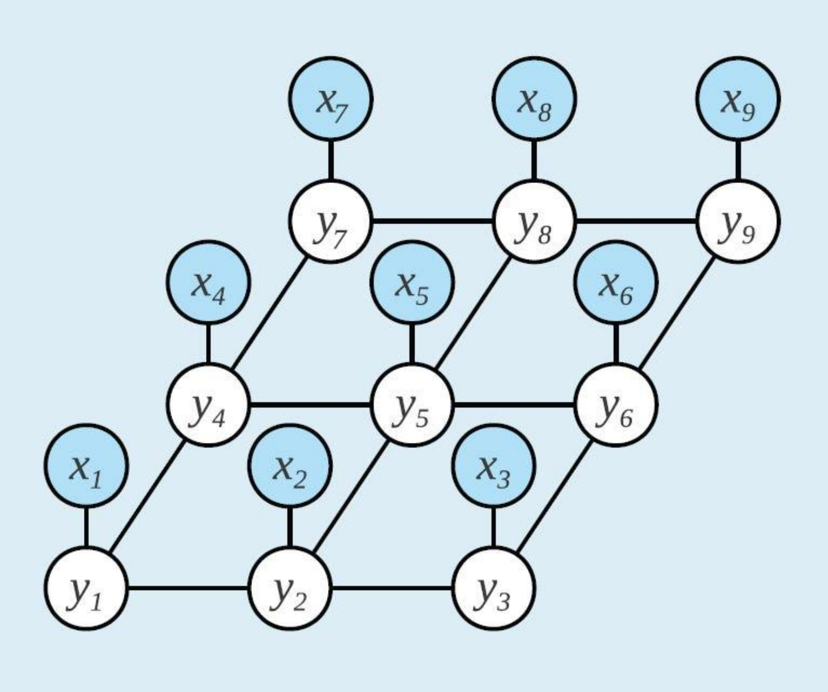






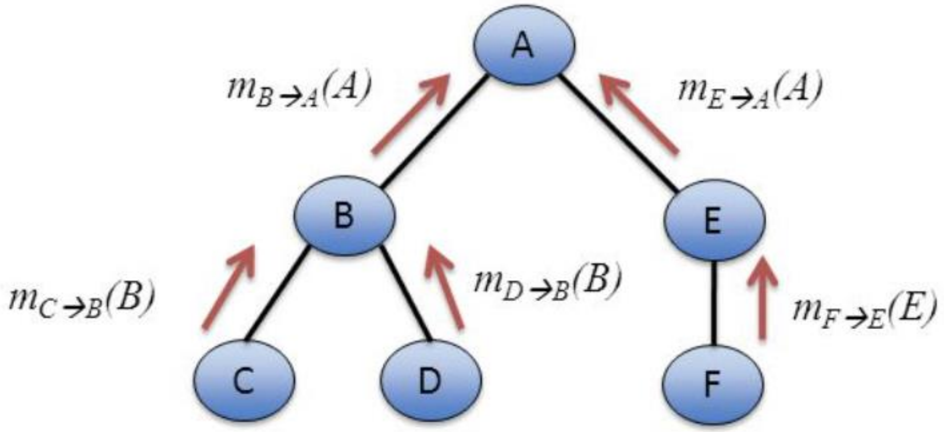
Build symmetries of toricode + noise channel into decoder
that predicts probabilities of logical qubits given measured syndrome.

Inference in Graphical Models



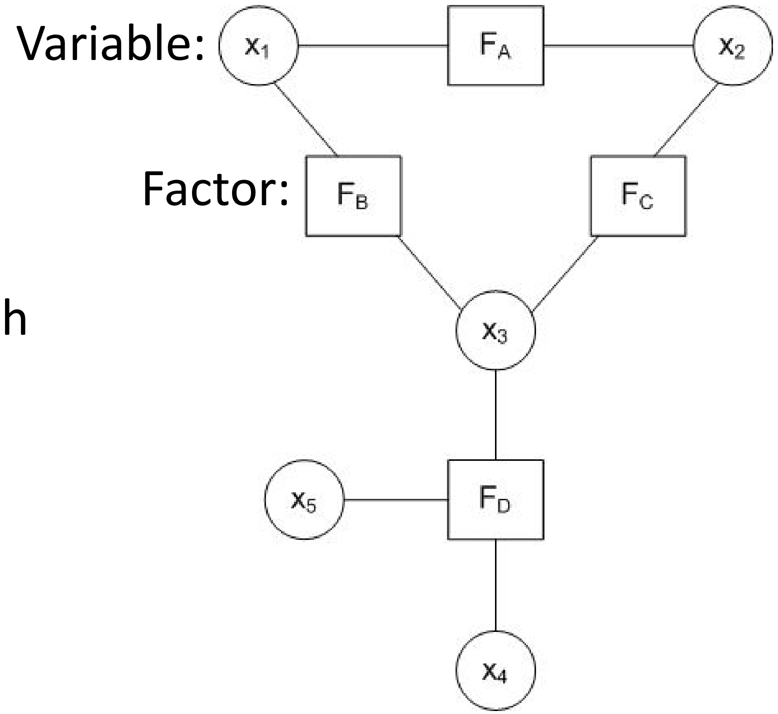
$$p(x, y) = \frac{1}{Z} \prod_f \psi_f(x_f, y_f)$$

Task: e.g. compute marginal: $p(x_\alpha) = \sum_{x \setminus x_\alpha} p(x)$



Belief propagation for inference:

Bethe Approximation to Free Energy



Factor graph

$$\log \hat{Z}_{BP} = -\mathcal{F}(\{\psi_f\}, \{b_f, b_i\}) \tag{3}$$

$$= \sum_{f \in F} \mathbb{E}_{b_f}[\log \psi_f] + \sum_{f \in F} \mathbb{H}(b_f) + \sum_{i \in V} (1 - |F_i|)\mathbb{H}(b_i)$$

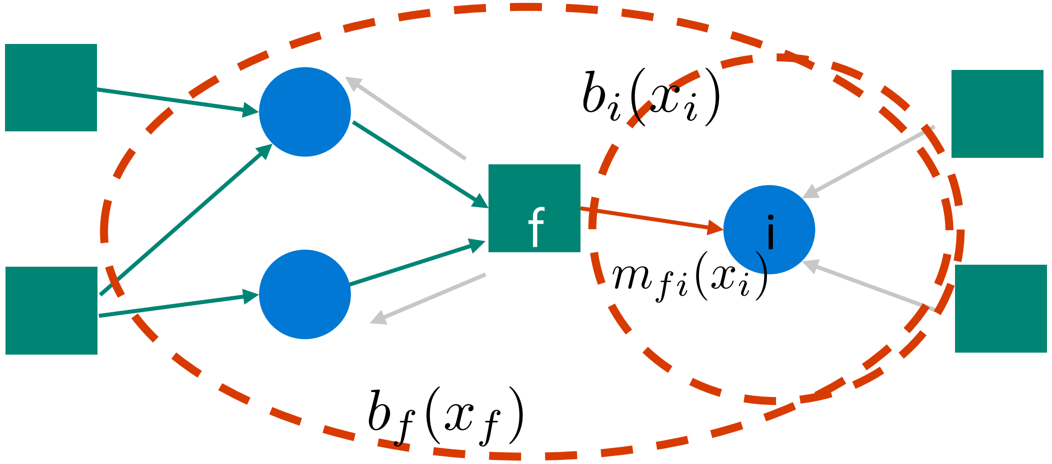
- Variational Free Energy with
  - Approximate entropy
  - Replace global constraint:

$$b_f(x_f) = \sum_{x \setminus x_f} b(x) \quad \forall f$$

with local constraints:

$$b_i(x_i) = \sum_{x_f \setminus x_i} b_f(x_f) \quad \forall f, i$$

Loopy BP



$$b_i(x_i) \propto \prod_{f \in F_i} m_{fi}(x_i)$$

$$b_f(x_f) \propto \psi_f(x_f) \prod_{i \in f} \prod_{g \in F_i \setminus f} m_{gi}(x_i)$$

$$m_{fi}^{\text{new}}(x_i) \leftarrow \delta(x_i) m_{fi}^{\text{old}}(x_i), \quad \delta(x_i) \doteq \frac{\sum_{x_f \setminus x_i} b_f(x_f)}{b_i(x_i)}$$

Loopy BP:

$$m_{fi}^{\text{new}}(x_i) = \sum_{x_f \setminus x_i} \psi_f(x_f) \prod_{j \in f \setminus i} \prod_{g \in F_j \setminus f} m_{gj}(x_j)$$

- Reparameterization of: $p(x) \propto \prod_f \psi_f(x_f)$
- After update $b_i(x_i) = \sum_{x_f \setminus x_i} b_f(x_f)$
- Minimizes Bethe Free Energy!

Generalizations: Kikuch Appriximation→ GBP

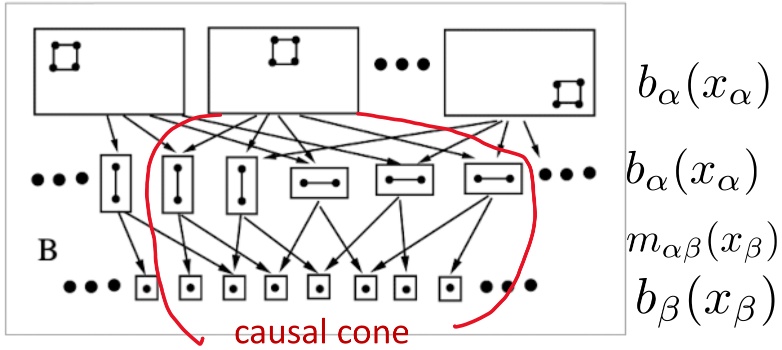
$$\log \hat{Z}_{GBP} = \sum_{f \in F} \mathbb{E}_b[\log \psi_f] + \sum_{\alpha \in \mathcal{R}} c_\alpha \mathbb{H}(b_\alpha).$$

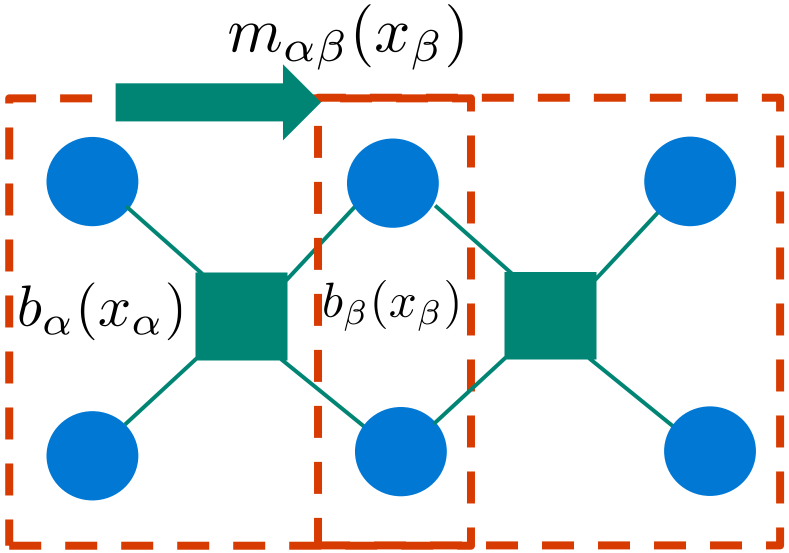
$c_\alpha = 1$ (top region)

$c_\alpha = 1 - \sum_{\beta \in an(\alpha)} c_\beta.$ (child regions)

+consistency between all parent-child regions:

$$b_\beta(x_\beta) = \sum_{x_\alpha \backslash x_\beta} b_\alpha(x_\alpha) \quad \forall \alpha \in \text{Parent}(\beta)$$





$$b_\alpha(x_\alpha) = \frac{1}{Z_\alpha} \prod_{f \in \alpha} \psi_f(x_f) \prod_{\substack{\gamma \in \{an(\Delta_\alpha) \backslash \Delta_\alpha\} \\ \beta \in \Delta_\alpha}} m_{\gamma\beta}(x_\beta)$$

(factor times all incoming messages into causal cone of a region)

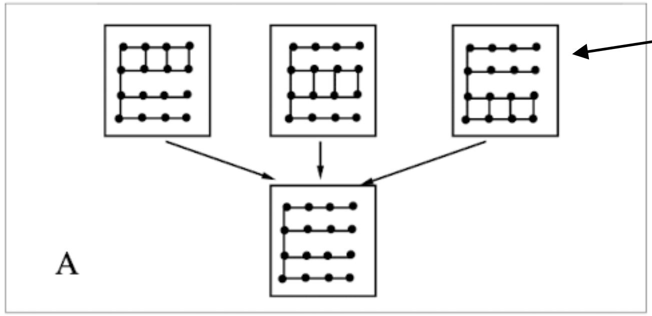
$$m^{\text{new}}_{\alpha\beta}(x_\beta) \leftarrow \delta(x_\beta) m^{\text{old}}_{\alpha\beta}(x_\beta) = \frac{\sum_{x_\alpha \backslash x_\beta} b_\alpha(x_\alpha)}{b_\beta(x_\beta)} m^{\text{old}}_{\alpha\beta}(x_\beta)$$

Generalization: EP

$$p(x) \propto \prod_f \psi_f(x_f)$$

$$q(x) = \prod_{fa} m_{fa}(x_a)$$

q(x) is some tractable distribution that you get by splitting factors.



q(x)

- Iterate:
  - Compute "Cavity Distribution"

$$q^{\text{cavity}}_{-(fa)}(x) = \frac{q(x)}{m_{fa}(x_a)}$$

  - Insert exact factor and project back to q

$$q^{\text{new}}(x_a) = \sum_{x \backslash x_a} \psi_f(x_f) q^{\text{cavity}}_{-(fa)}(x)$$

  - Recompute message

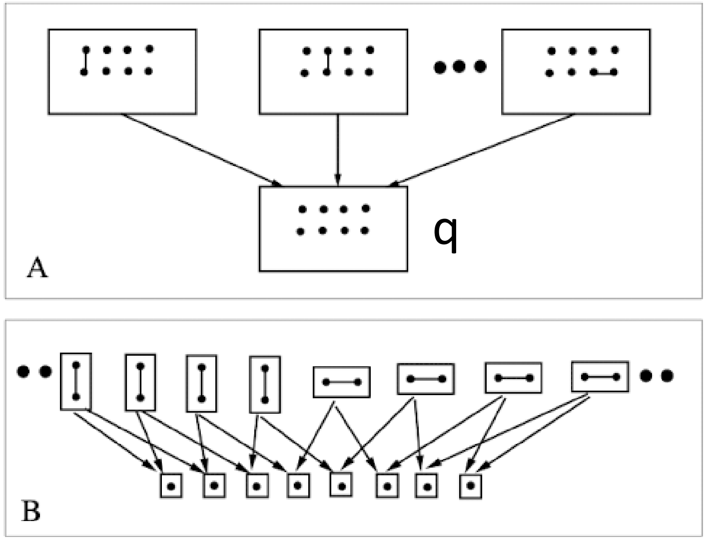
$$m_{fa}(x_a) \propto \frac{q^{\text{new}}(x_a)}{q^{\text{cavity}}_{-(fa)}(x_a)}$$

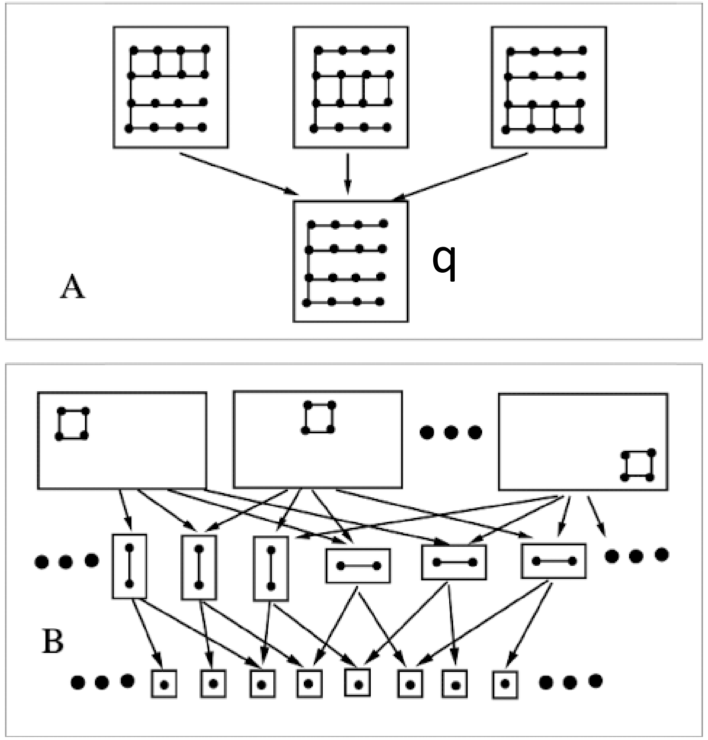
Region Graphs

EP region graph

Same fixed points!

BP region graph





TreeEP region graph

Same fixed pints!

GBP region graph

Tensor Networks

Gauging tensor networks with belief propagation

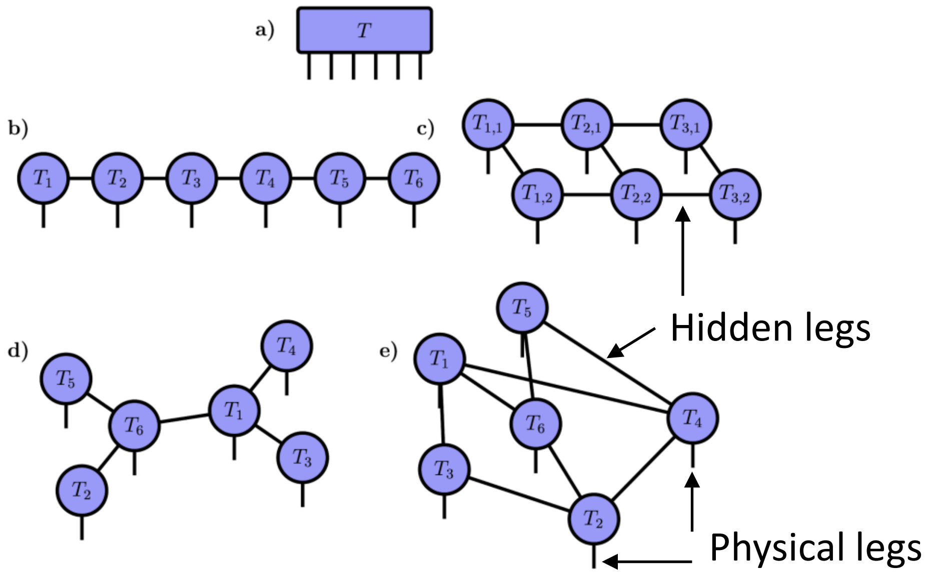
Joseph Tindall, Matthew Fishman

Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

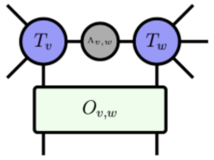
July 3, 2023

(figures from)

$$|\psi> = \sum_{x_1,..x_n} T_1(x_1) \circ T_2(x_2) \circ ... \circ T_n(x_n)|x_1, ..., x_n>$$



- T can be vector, matrix, or higher order tensor
- Much like graphical models in ML
- To compute expectations over operators you need to contract the tensors (similar to inference in GMs)



- TNs are very good approximations to quantum states if entanglement is limited.

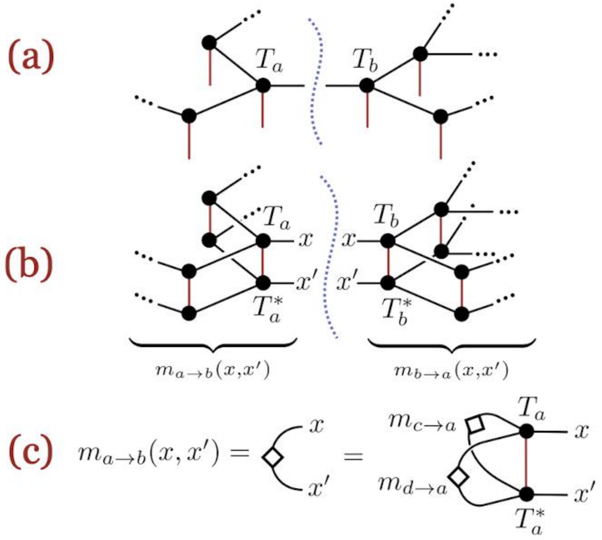
Quantum BP

Tensor Networks contraction and the Belief Propagation algorithm

R. Alkabetz[1] and I. Arad[1]

[1] *Department of Physics, Technion, 3200003 Haifa, Israel*

(Dated: August 25, 2020)



- Replace distributions with density (PSD) matrices
- Factors are now given by PSD tensors:

$$\psi_i(\{z_e, z'_e\}) = \sum_{x_i} T_i(x_i, \{z_e\}) T^*(x_i, \{z'_e\}), \quad z_e, z'_e \in N(i)$$

- Message updates:

$$m_{a\to b}^{(t+1)} \stackrel{\text{def}}{=} \mathrm{Tr}\left(T_a T_a^* \prod_{a' \in N_a \setminus \{b\}} m_{a'\to a}^{(t)}\right).$$

QBP

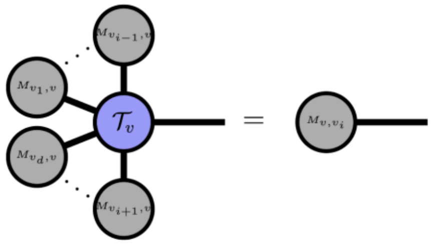
Gauging tensor networks with belief propagation

Joseph Tindall, Matthew Fishman

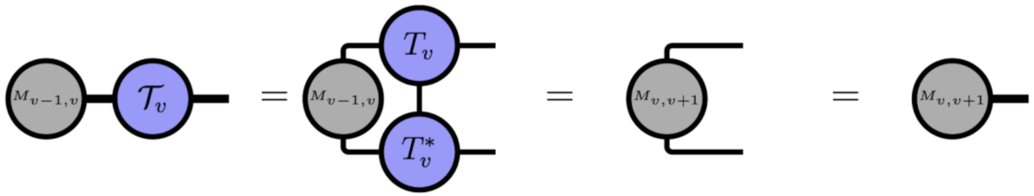
Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

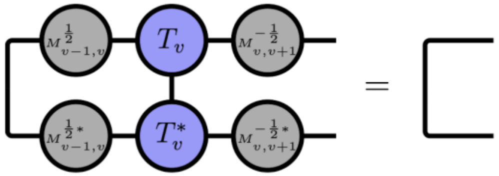
July 3, 2023

- Simple (trivial) update has same fixed point as QBP (even on loopy graphs).



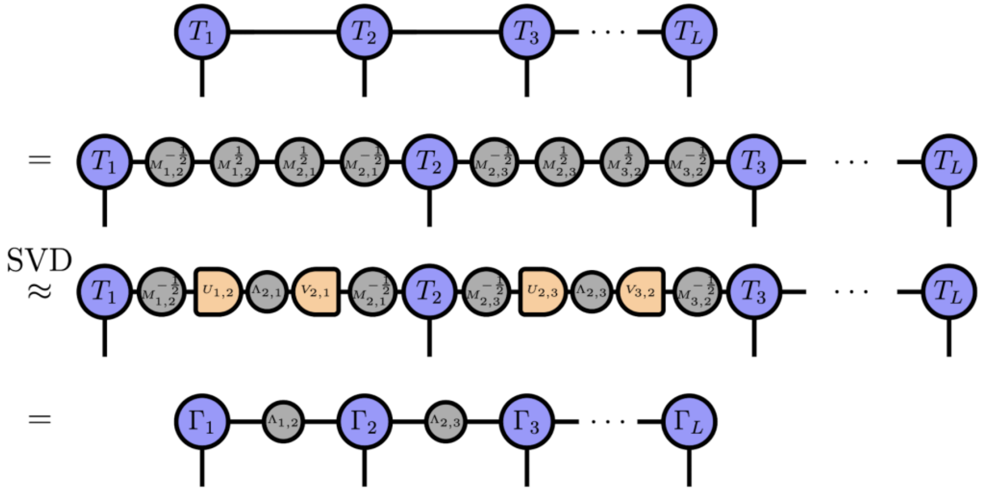
$$m_{a\to b}^{(t+1)} \stackrel{\text{def}}{=} \mathrm{Tr}\left(T_a T_a^* \prod_{a'\in N_a \setminus \{b\}} m_{a'\to a}^{(t)}\right).$$





Isometric tensor

After convergence, use BP messages to transform to canonical (Vidal) gauge:

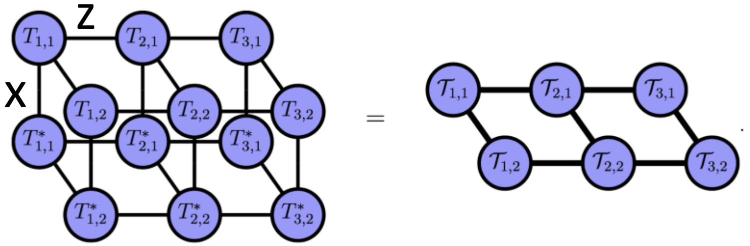


Generalized QBP (QEP)

(disclaimer: not yet tested)

Density Matrix approximated by TN:

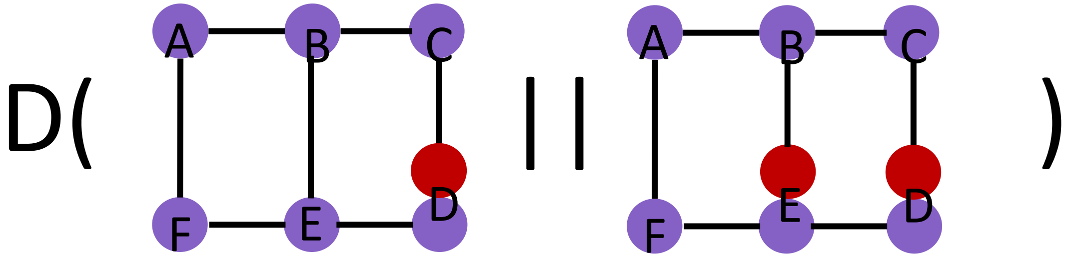
$$\rho(Z) = \frac{1}{Z}\prod_A \sum_{x_A} T_A(\{z_A\}, x_A) T^*(\{z'_A\}, x_A) = \frac{1}{Z}\prod_A T_A(z_A)$$

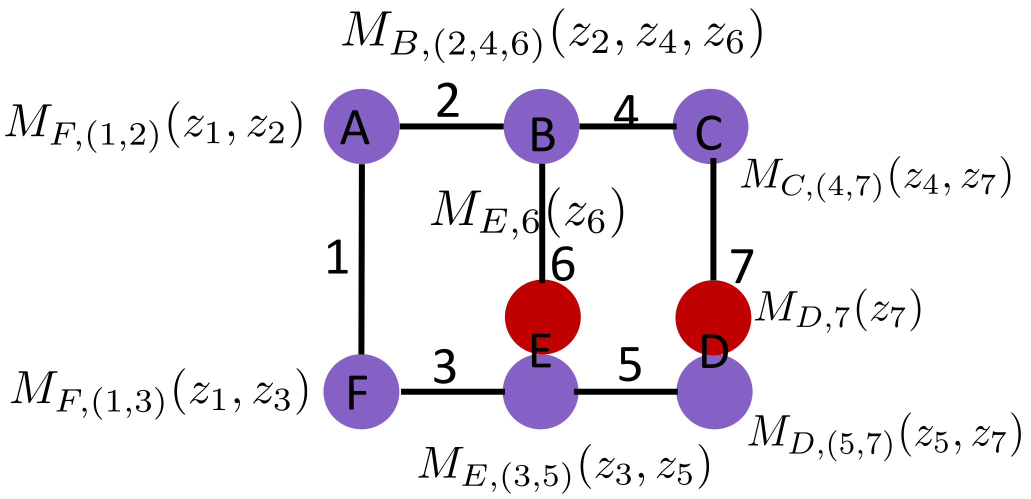


Define messages by splitting tensors T into subset of variables:

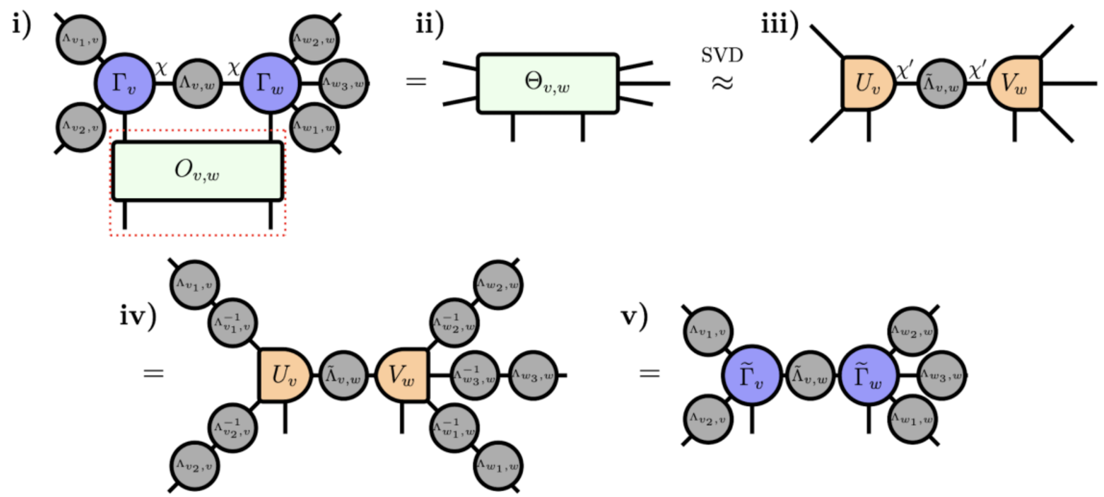
$$M_{A\beta}(z_\beta) = \sum_y m_{A\beta}(\{z_\beta\}, y) m^*_{A\beta}(\{z'_\beta\}, y)$$

Message Update: $M^{\text{new}}_{A\beta}(z_\beta) = \underset{M_{A\beta}(z_\beta)}{\arg\min}\, D(\frac{1}{Z} T_A \prod_{F\backslash A} M_F || \frac{1}{Z'} \prod_F M_F)$





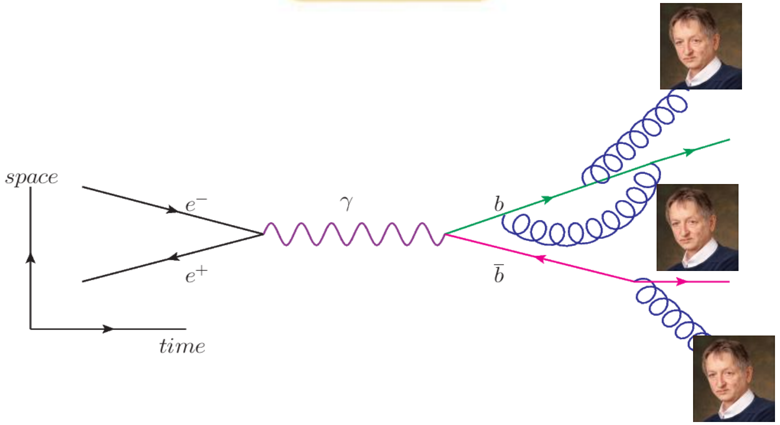
Quantum Evolution (or learning)



- Insert Trotterized unitary evolution: $O = e^{-i\delta H}$
- Recompute new (approximate) tensors T
- Rerun QBP/QEP to equilibrium
- Repeat
- Done for quantum circuit simulation
- It's nonequilibrium thermodynamics again (in z)!
  - Evolution step = M-step = Work-step
  - QBP step = E-step = Heat-step

Quantum Field Theory for Deep Learning, A New Particle?

Photon, Gluon, Z/W-Boson, Electron, Muon, Fermion,…



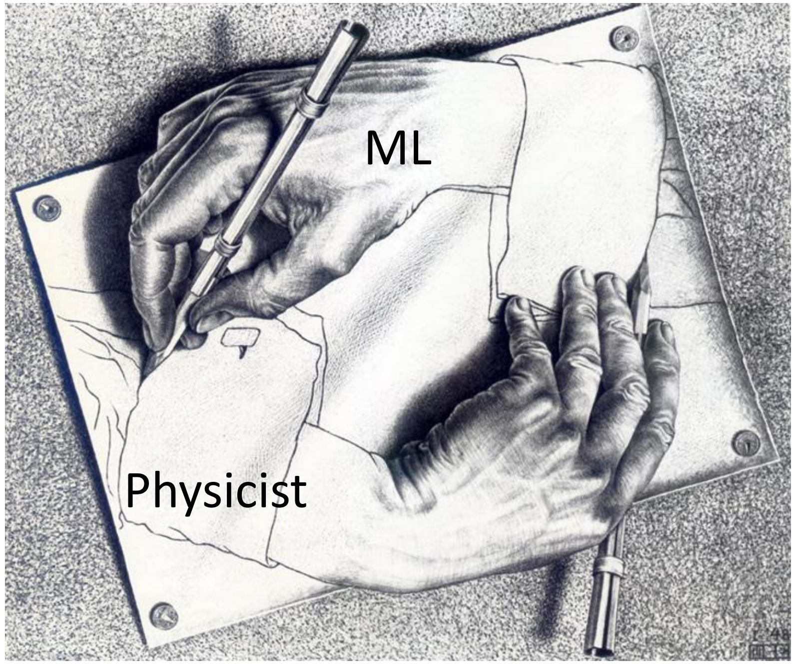
The Hintons in your Neural Network:
a Quantum Field Theory View of Deep Learning

Roberto Bondesan [1] **Max Welling** [1]

Large Hinton Collider

Conclusions

Deep connection between AI and physics: also quantum physics?



GBP from ML is a great hammer for tensor network computations

Conclusions

r
ns