Free Energy is All You Need

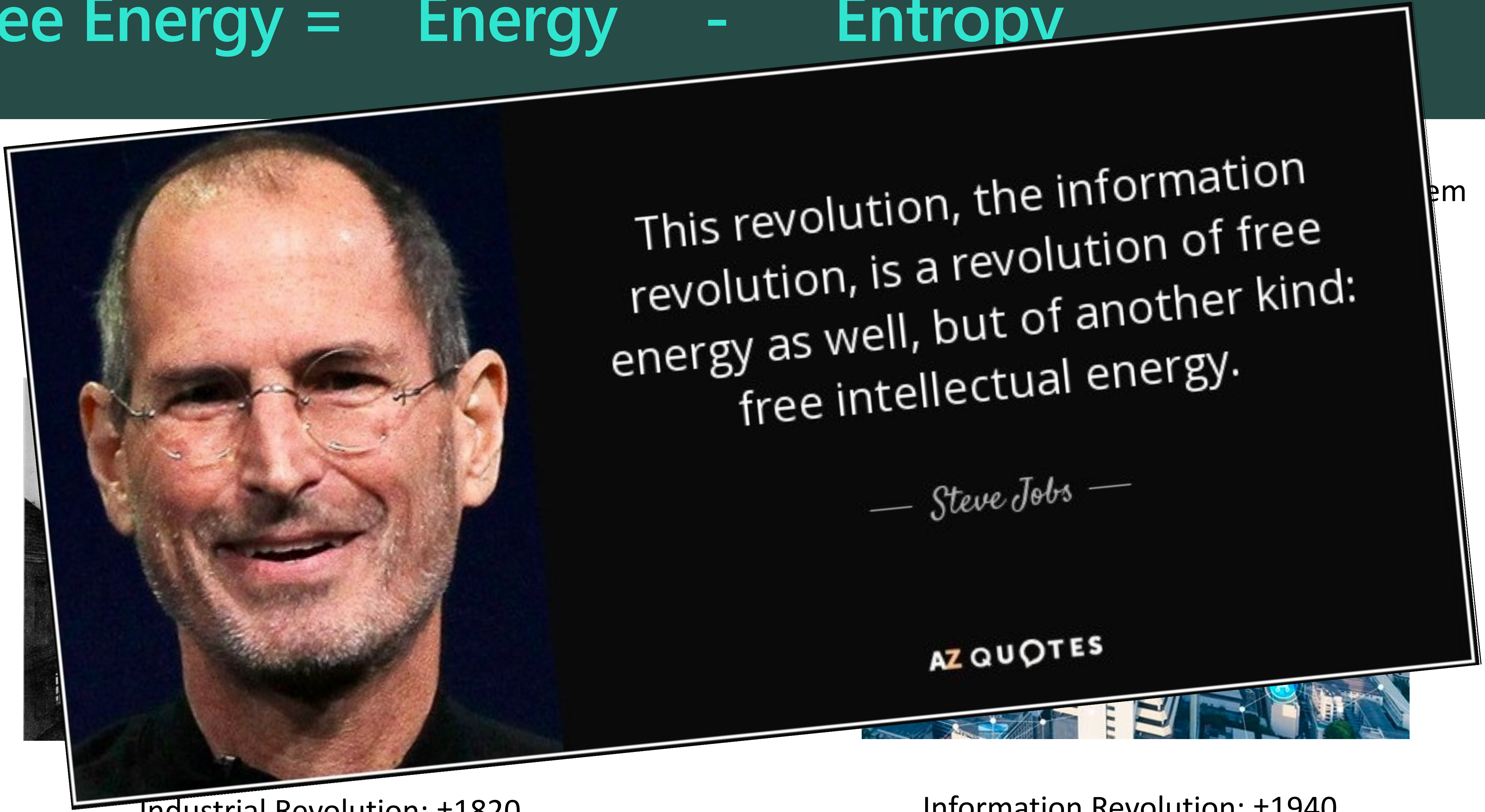
Max Welling

Today’s Menu

- AI4Science: A new paradigm for scientific discovery?
- Free Energies in Physics and Machine Learning
- Examples of AI4Science:
    - Transition Path Sampling
    - Classical Density Functional Theory
    - Quantum Variation Monte Carlo
    - Gravitational Lensing
- Outlook

Free Energy = Energy - Entropy

em

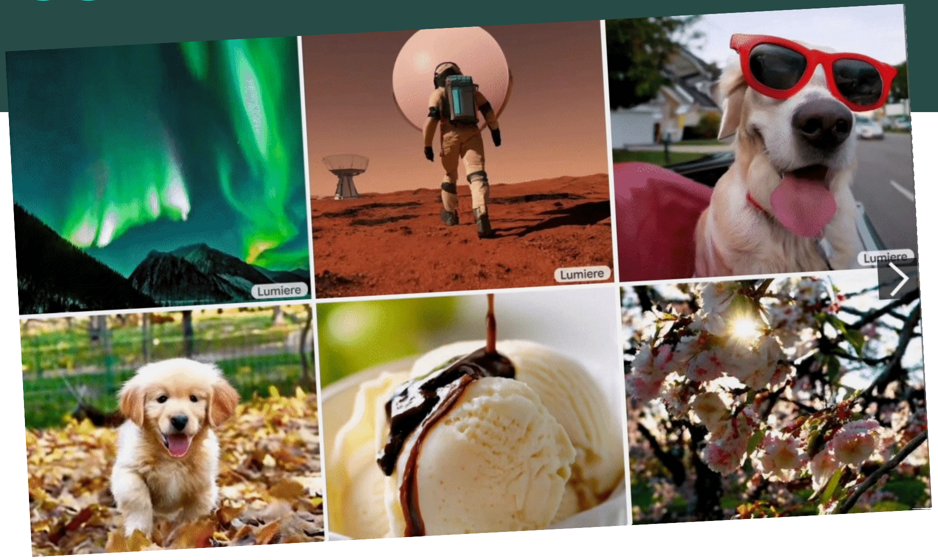


Industrial Revolution: ±1820

Information Revolution: ±1940

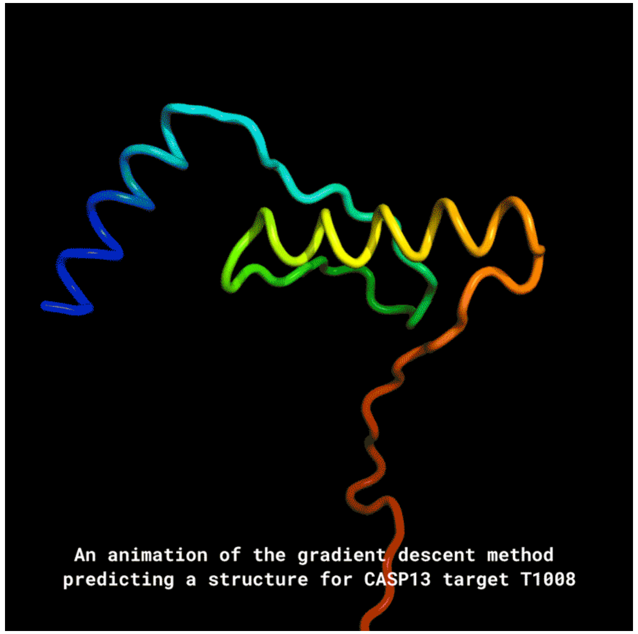
Generative AI: Images

Generative AI: Videos



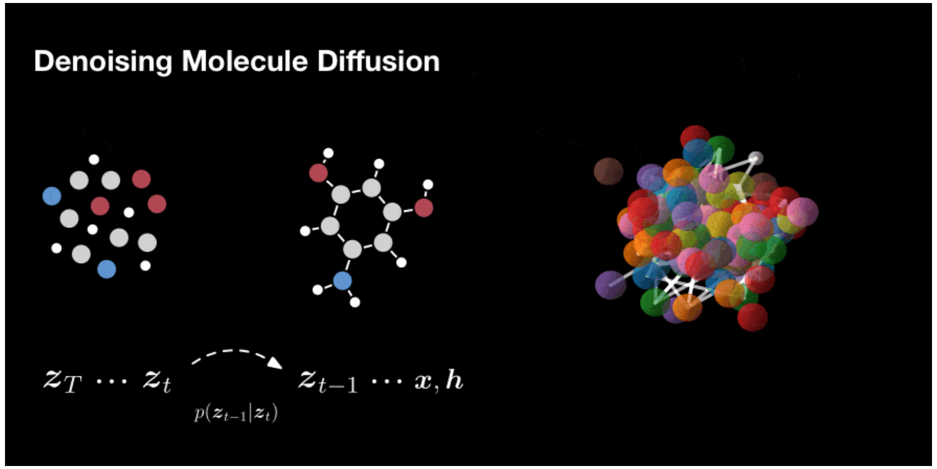
"A shot following a hiker through jungle brush."

Deep learning in the natural sciences



Molecule Generation

Protein Folding



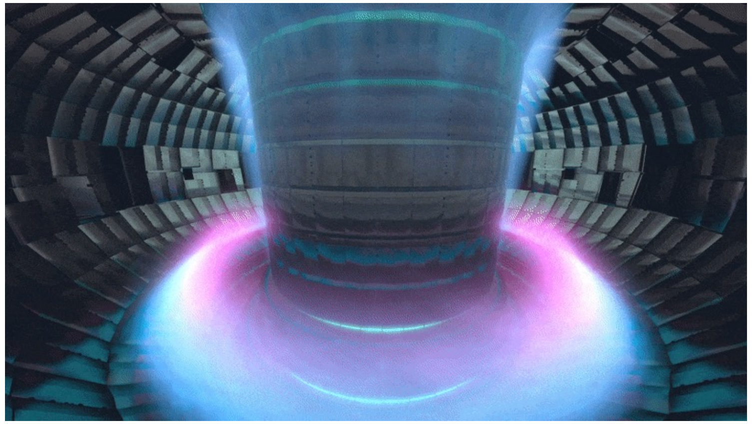
Equivariant Diffusion for Molecule Generation in 3D

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

Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, ... Demis Hassabis + Show authors

Nature **596**, 583–589 (2021) | Cite this article



Plasma Control

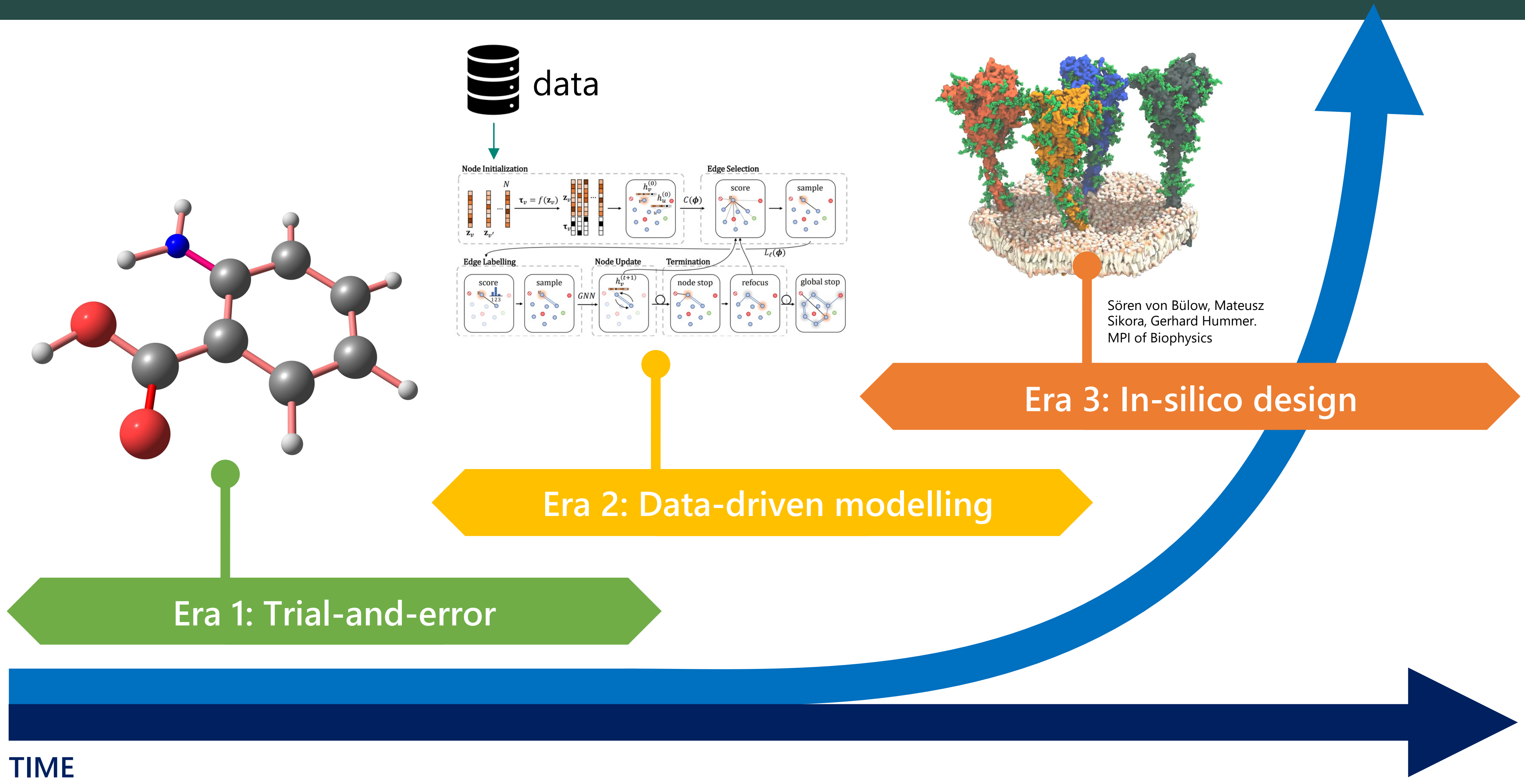
Magnetic control of tokamak plasmas through deep reinforcement learning

Jonas Degrave, Federico Felici, ... Martin Riedmiller + Show authors

Nature **602**, 414–419 (2022) | Cite this article

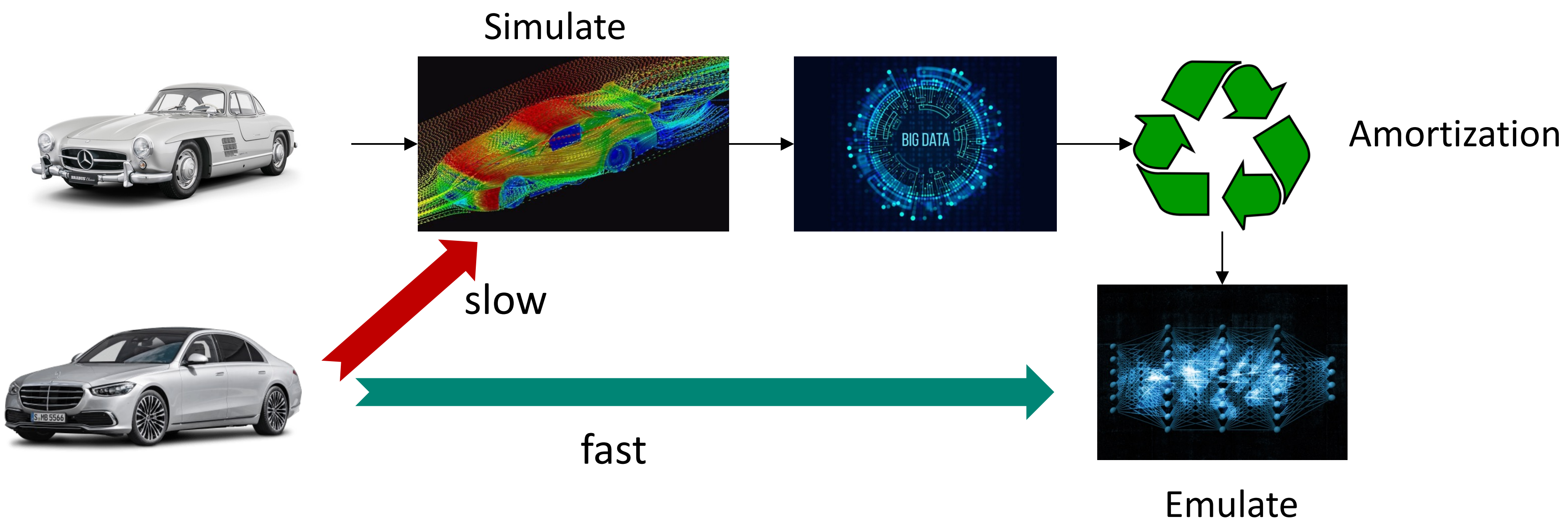
From Experiment to Computer Simulation

COMPUTATIONAL COMPLEXITY



From Simulation to Emulation

Simulate → Train NN Surrogate → Emulate



From Physics to 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)

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

ML

Physics/Chemistry/Biology

$$P(z|x) = \frac{P(x|z)P(z)}{P(x)}$$

$$KL[Q||P] = \int dz\ Q(z)[\log Q(z) + E_Q[-\log(P(x|z)P(z))] + \log P(x)] \geq 0$$

$$-\log P(x) \equiv F(x) \leq \mathcal{F}(Q;x) \equiv E_Q[-\log(P(x|z)P(z))] - S(Q)$$

$$P(z) = \frac{\exp[-H(z)]}{Z}$$

$$KL[Q||P] = \int dz\ Q(z)[\log Q(z) + H(z) + \log Z] \geq 0$$

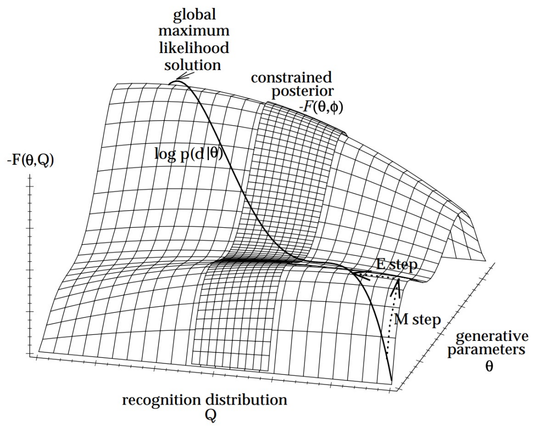
$$-\log Z \equiv F \leq \mathcal{F}(Q) \equiv U(Q) - S(Q)$$

Expectation Maximization as Stochastic Thermodynamics

EM-algorithm:

$$\text{E-step:} \quad \min_{Q} \mathcal{F}(Q, \theta)$$

$$\text{M-step:} \quad \min_{\theta} \mathcal{F}(Q, \theta)$$



(from 1994 "Helmholtz Machine" paper)

Thermodynamics:

$$\delta\mathcal{F} = \frac{\delta U}{\delta\theta}\delta\theta + \frac{\delta U}{\delta Q}\delta Q - \frac{\delta S}{\delta Q}\delta Q$$

$$= \text{Work} + \text{Heat} + \text{System Entropy change}$$

$$= \text{M-step} + \text{E-step}$$

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)

Bayesian View of Statistical Mechanics:
*Entropy is **our** degree of ignorance about the microscopic degrees of freedom of a system*

Generative AI

2015

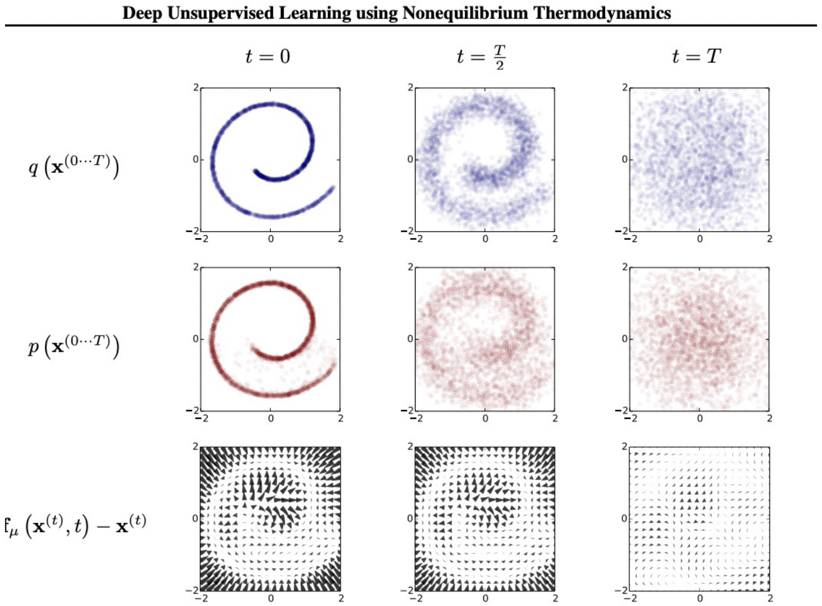
Deep Unsupervised Learning using Nonequilibrium Thermodynamics

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



Diffusion Based Models

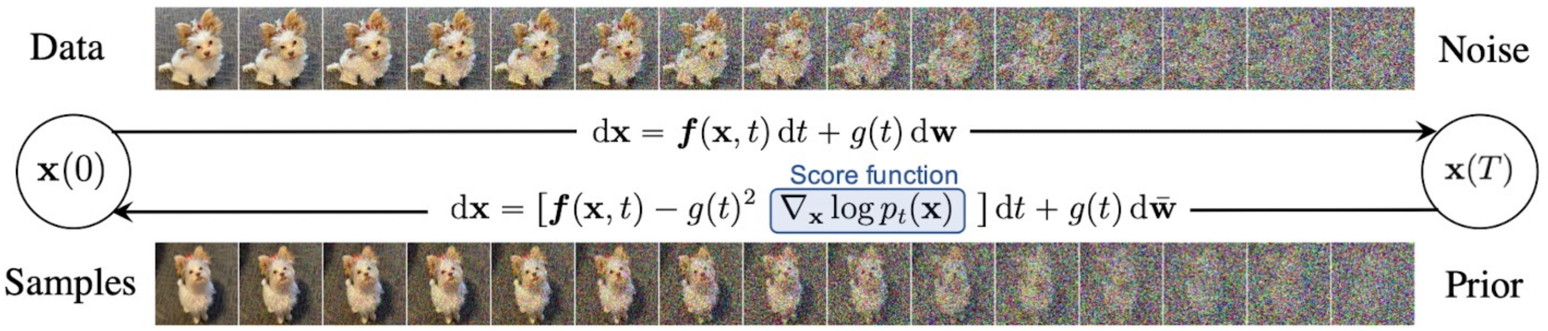
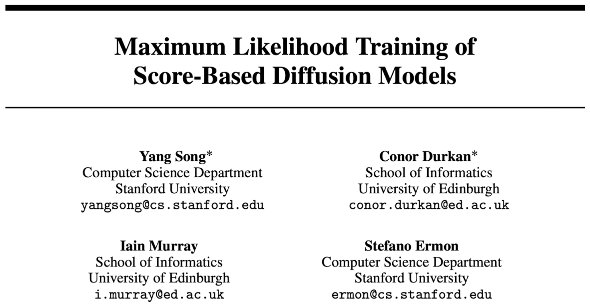


Figure 1: We can use an SDE to diffuse data to a simple noise distribution. This SDE can be reversed once we know the score of the marginal distribution at each intermediate time step, $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$.

2021



Maximum Likelihood Training of Score-Based Diffusion Models

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

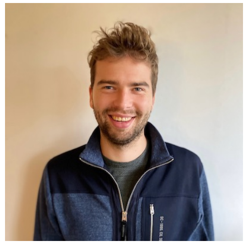
Conor Durkan* School of Informatics University of Edinburgh conor.durkan@ed.ac.uk

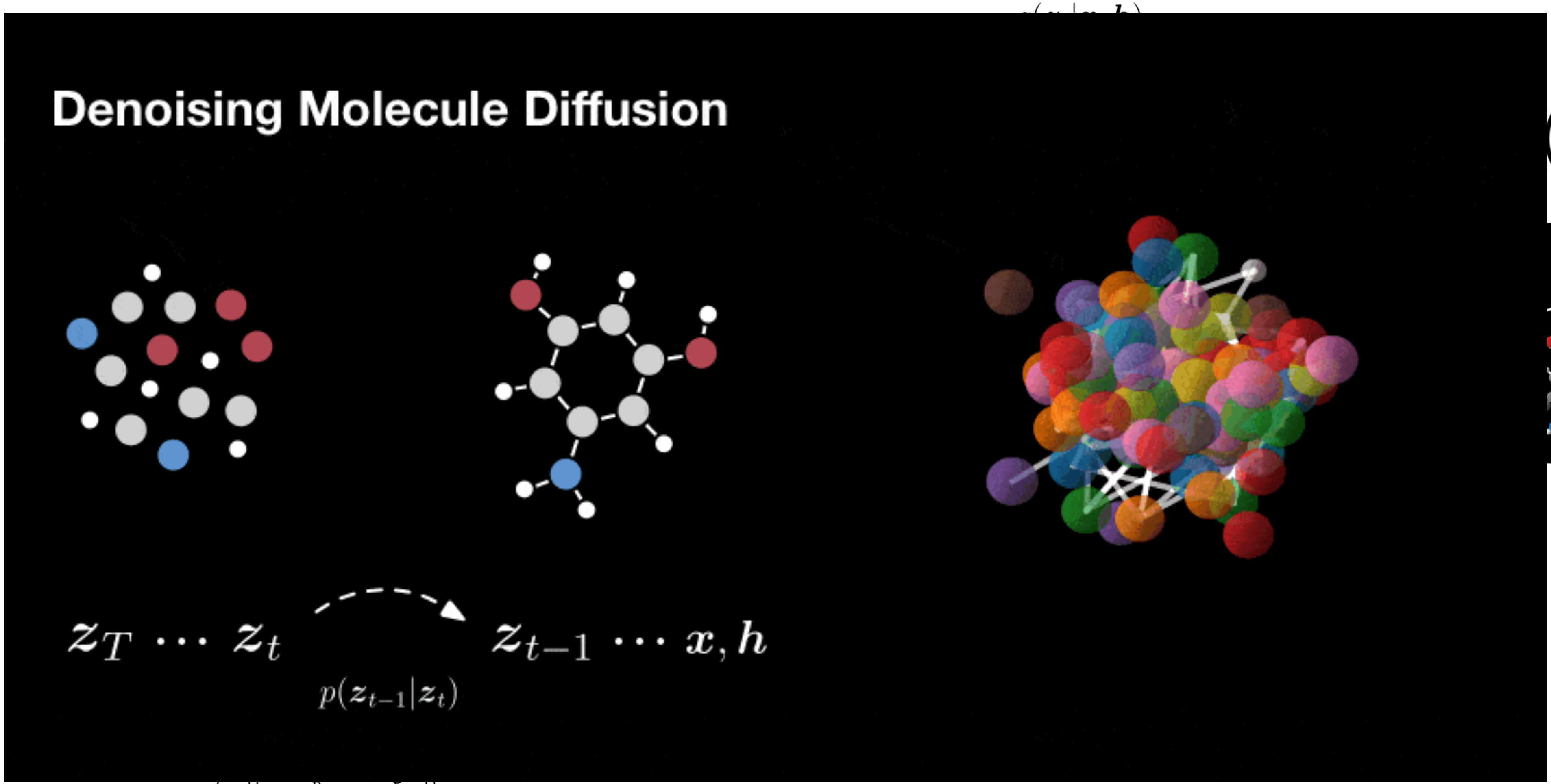
Iain Murray School of Informatics University of Edinburgh i.murray@ed.ac.uk

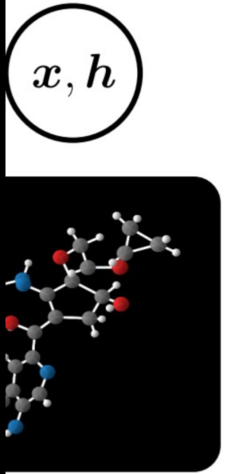
Stefano Ermon Computer Science Department Stanford University ermon@cs.stanford.edu

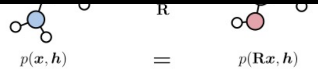
- Crooks Fluctuation Theorem: $\frac{P(A \to B)}{P(A \leftarrow B)} = \exp[\beta(W_{A\to B} - \Delta F)]$
- The faster you want to generate, the more work you dissipate to the environment.

Generative AI: Molecules!









Free Energy is all you need

Variational Inference
MCMC Sampling

Optimal Control
G-Flows

Free Energy

Schrodinger Bridges
Optimal Transport

Diffusion Models
Normalizing Flows

- Objective is to minimize KL(Q||P) (a.k.a. Free Energy)

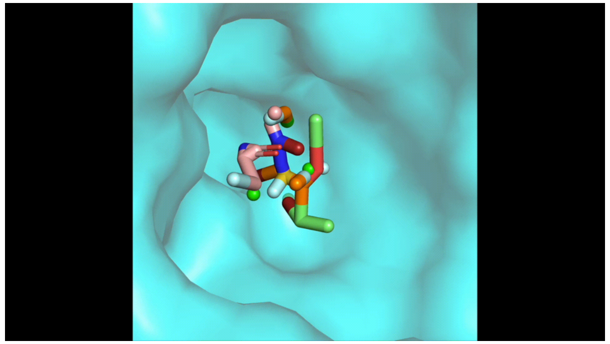
$$\mathcal{F}(Q;x) \equiv E_Q[-\log(P(x|z)P(z))] - S(Q)$$

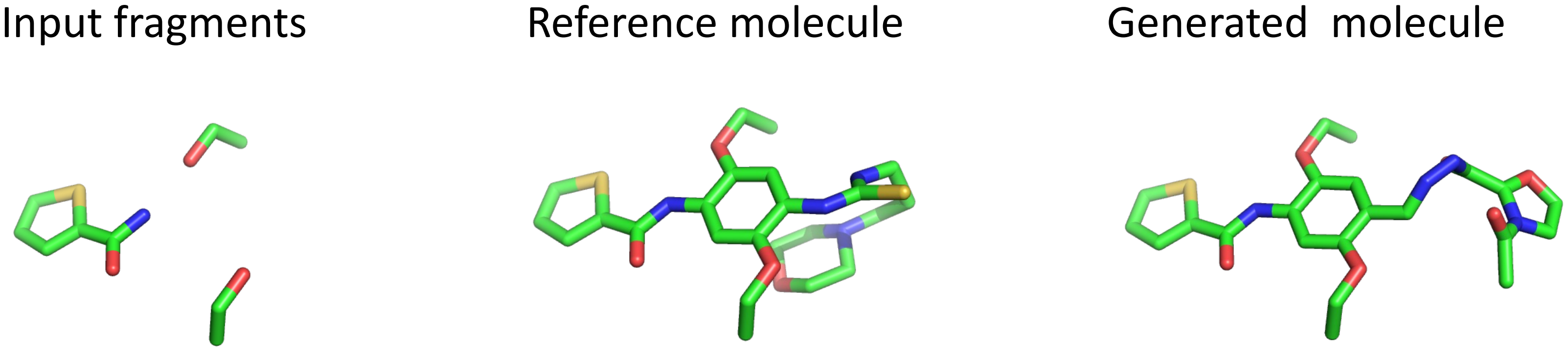
- Q & P are Markov Chains:

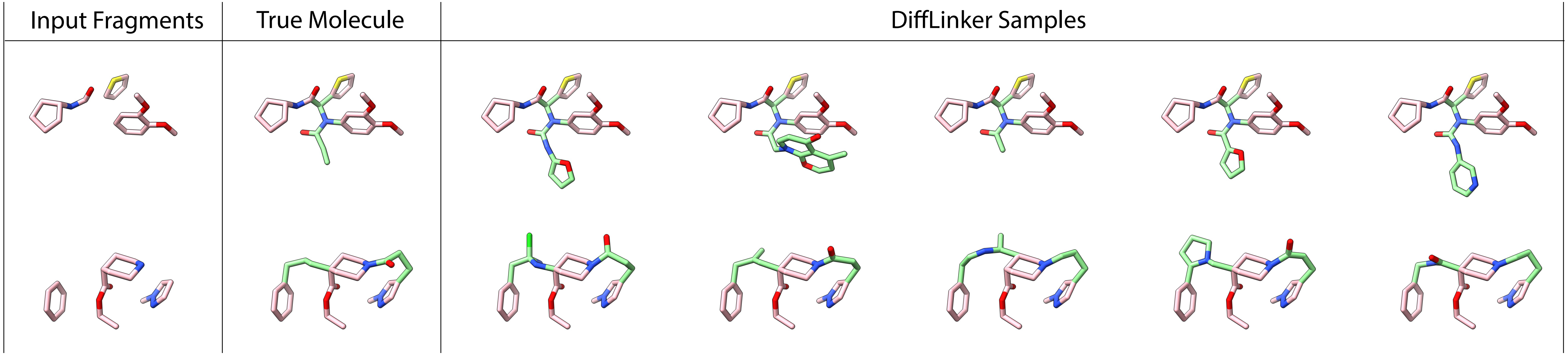
$$Q(Z) = Q_0(z_0)\prod_{t=1}^{T} F_t(z_t|z_{t-1})$$

$$P(Z) = P_T(z_T)\prod_{t=1}^{T} B_t(z_{t-1}|z_t)$$

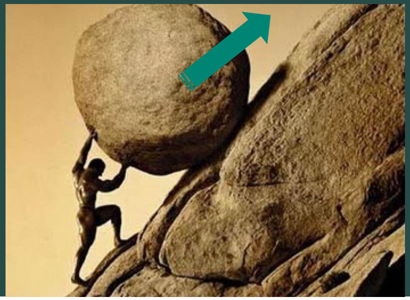
DiffLinker: Molecular Linker Design







Transition Path sampling



Project Sisyphus

Sampling transition paths between molecular conformations

PIPS : Path Integral Path Sampling:

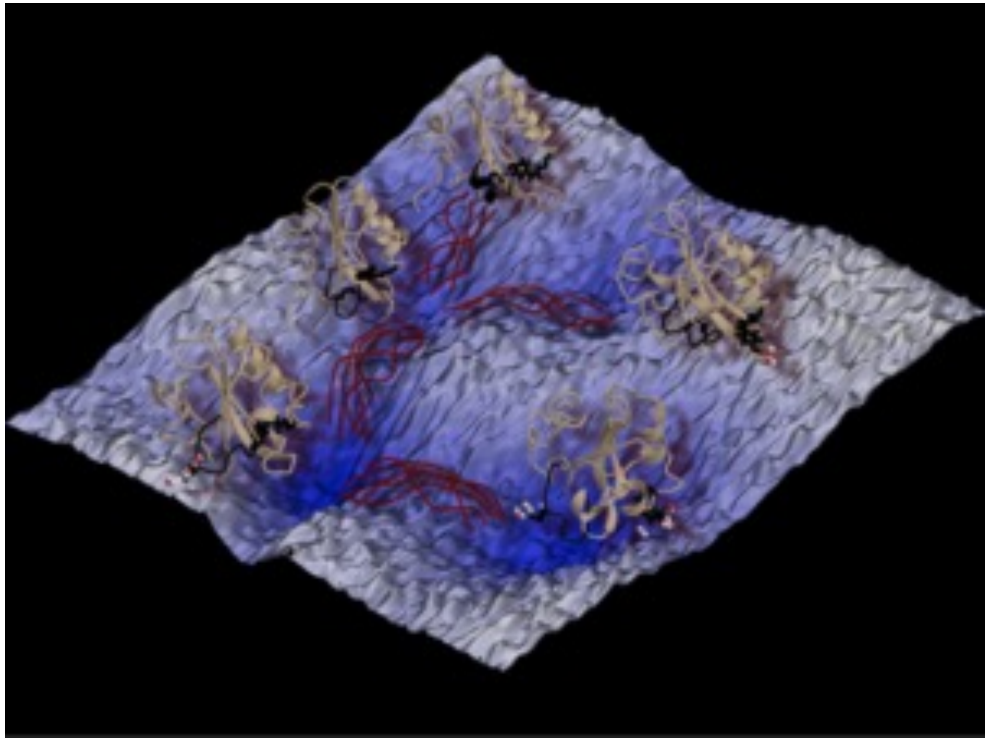
Given initial state r_0 and target state r_T find the series of intermediate states $\{r_1, r_2, \ldots, r_{T-1}\}$ that describe the transition path of minimal energy.

Langevin dynamics:

extra force term

$$\underbrace{\begin{pmatrix} \mathrm{d}\boldsymbol{r}_t \\ \mathrm{d}\boldsymbol{v}_t \end{pmatrix}}_{\mathrm{d}\boldsymbol{x}_t} = \underbrace{\begin{pmatrix} \boldsymbol{v}_t \\ -\nabla_{\boldsymbol{r}_t}\boldsymbol{U}(\boldsymbol{r}_t) \end{pmatrix}}_{\boldsymbol{f}(\boldsymbol{x}_t,t)} \mathrm{d}t + \underbrace{\begin{pmatrix} 0_{3n} \\ \mathbb{I}_{3n} \end{pmatrix}}_{\boldsymbol{G}(\boldsymbol{x}_t,t)} \cdot \left(\boldsymbol{u}(\boldsymbol{x}_t, t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{\varepsilon}_t\right), \qquad t \in [0, \tau]$$

Train policy to force molecule over energy barrier



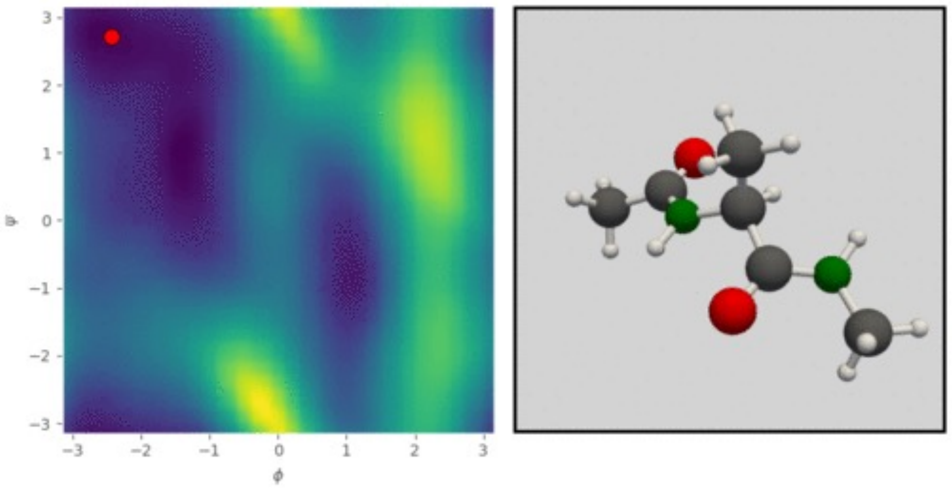
Source: https://www.e-cam2020.eu/rare-events-story/

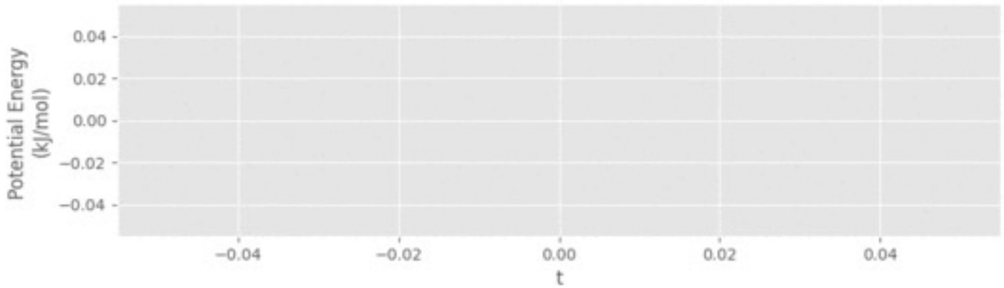
Alanine Dipeptide

- Extensively studied molecule with known collective variables

Collective Variables:

Dihedral angles ψ and ϕ





With Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Bernd Ensing

Learning the Free Energy of Classical DFT

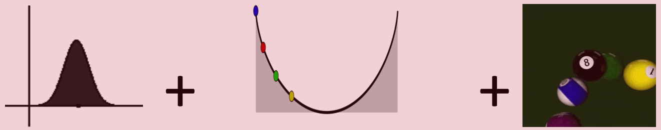
Jacobus Dijkman:
(with B. Ensing
J.W. van Meent
M. Dijkstra
R. Van Roij)

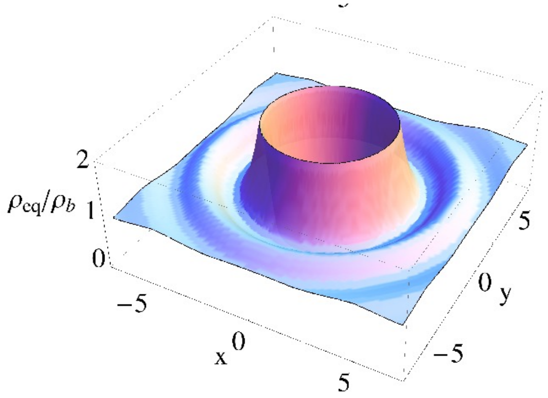
- Consider liquid/gas with the following free energy:

$$\Omega[\rho] = \mathscr{F}[\rho] + \int \mathrm{d}r\rho(\mathbf{r})\left(V_{\text{ext}}\left(\mathbf{r}\right) - \mu\right)$$

Grand Potential Functional, Internal Free Energy, Particle Density, External Potential, Chemical Potential

$$F(\rho(x,t)) = F^{\text{id}}[\rho(x,t)] + \int(V^{\text{ext}} - \mu)(x)\rho(x,t)dx + F^{\text{excess}}[\rho(x,t)]$$





- Learn $\mathscr{F}[\rho]$ from MD data
- minimize $\Omega[\rho]$ over density for new V
- Dynamics: $\partial_t \rho(x,t) = \gamma \nabla_x \cdot \left(\rho(x,t)\nabla_x \frac{\delta\Omega[\rho]}{\delta\rho(x,t)}\right)$

Learning Neural Free-Energy Functionals with Pair-Correlation Matching

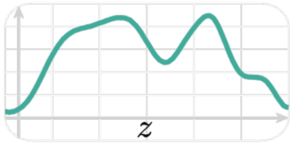
Jacobus Dijkman,[1,2] Marjolein Dijkstra,[3] René van Roij,[4]
Max Welling,[2] Jan-Willem van de Meent,[2] and Bernd Ensing[1,5]

Training a free-energy approximator: naïve approach

1. Generate training data from MC simulations

simulation data

sampled non-uniform density



$$\begin{pmatrix}\rho(z_0)\\ \rho(z_1)\\ \vdots \\ \rho(z_n)\end{pmatrix} \quad -\beta\frac{\delta \mathscr{F}^{\text{exc}}_{\text{sim}}}{\delta\rho_i} = \ln \Lambda^3\rho(z_i) + \beta V_{ext}(z_i) - \beta\mu \longrightarrow \frac{\delta \mathscr{F}^{\text{exc}}_{\text{sim}}}{\delta\rho_i}$$

2. Train the neural network

$$\begin{pmatrix}\rho(z_0)\\ \rho(z_1)\\ \vdots \\ \rho(z_n)\end{pmatrix} \;\; \text{CNN}_\theta \;\; F_\theta^{\text{exc}} \;\; \text{autodiff} \;\; \begin{pmatrix}\frac{\partial F_\theta^{\text{exc}}}{\partial\rho_0}\\ \vdots \\ \frac{\partial F_\theta^{\text{exc}}}{\partial\rho_n}\end{pmatrix}$$

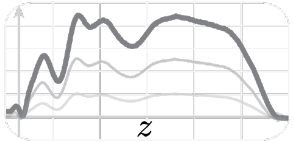
update model parameters

$$\theta_{l+1} \leftarrow \theta_l + \nabla_\theta \left\| \frac{\delta \mathscr{F}^{\text{exc}}_{\text{sim}}}{\delta\rho_i} - \frac{1}{\Delta z}\frac{\partial F_\theta^{\text{exc}}}{\partial\rho_i} \right\|$$

3. Classical DFT at fixed $\beta\mu$ and $\beta V^{\text{ext}}(z)$

$$\begin{pmatrix}\rho(z_0)\\ \rho(z_1)\\ \vdots \\ \rho(z_n)\end{pmatrix} \;\; \text{CNN}_\theta \;\; F_\theta^{\text{exc}} \;\; \text{autodiff} \;\; \begin{pmatrix}\frac{\partial F_\theta^{\text{exc}}}{\partial\rho_0}\\ \vdots \\ \frac{\partial F_\theta^{\text{exc}}}{\partial\rho_n}\end{pmatrix}$$

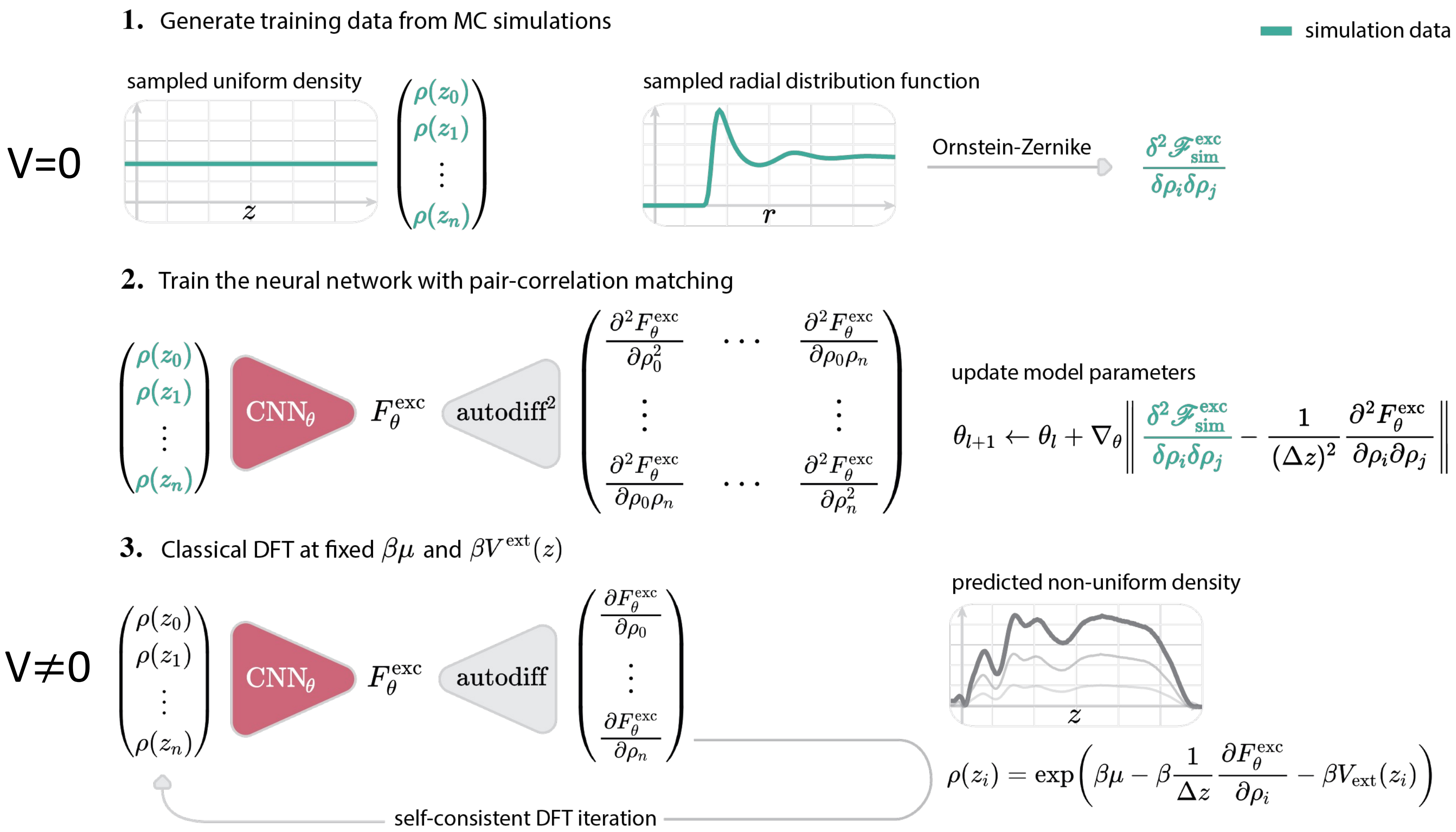
predicted non-uniform density



$$\rho(z_i) = \exp\left(\beta\mu - \beta\frac{1}{\Delta z}\frac{\partial F_\theta^{\text{exc}}}{\partial\rho_i} - \beta V_{\text{ext}}(z_i)\right)$$

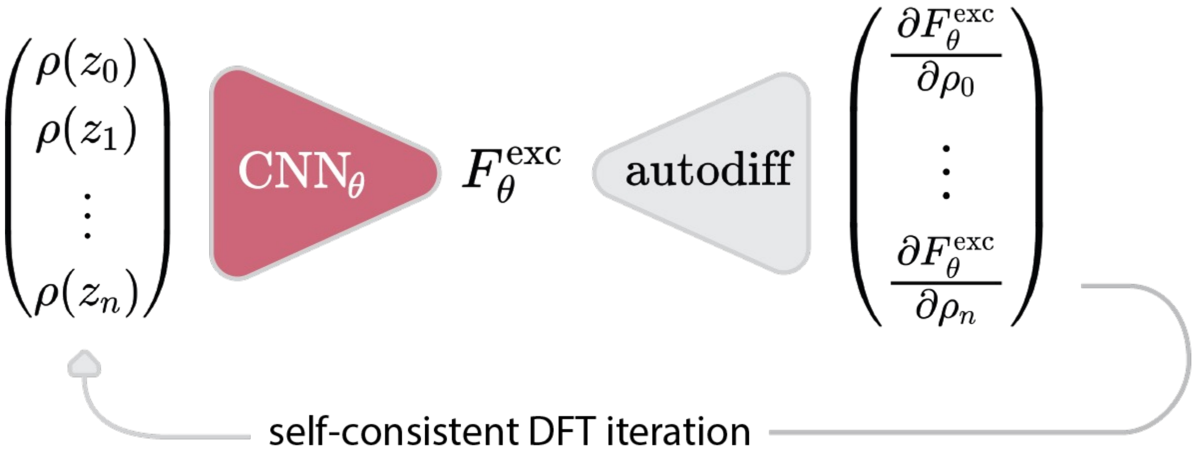
self-consistent DFT iteration

Training a free-energy approximator: pair-correlation matching

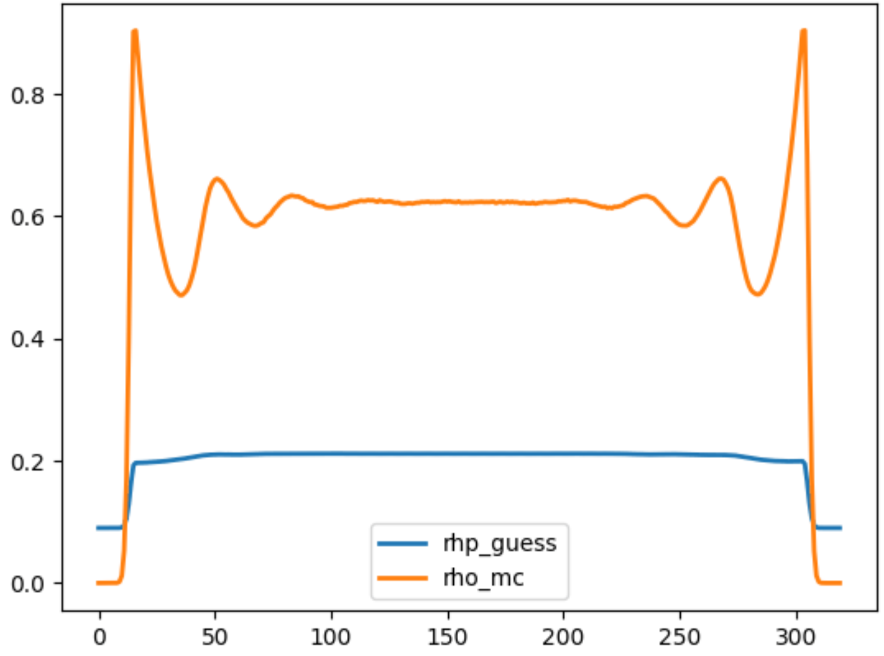


Classical DFT application

Classical DFT at fixed $\beta\mu$ and $\beta V^{\text{ext}}(z)$

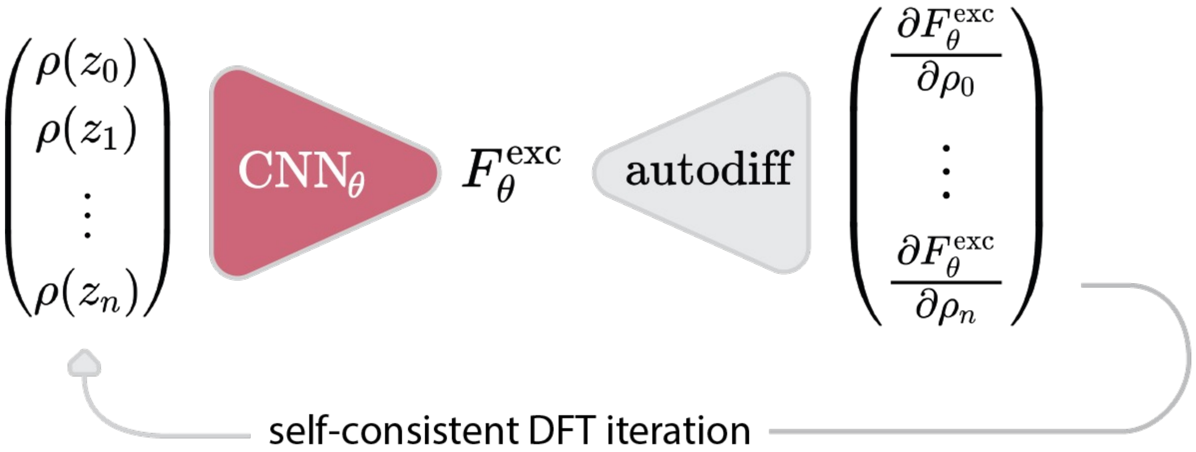


$$\rho(z_i) = \exp\left(\beta\mu - \beta\frac{1}{\Delta z}\frac{\partial F_\theta^{\text{exc}}}{\partial \rho_i} - \beta V_{\text{ext}}(z_i)\right)$$

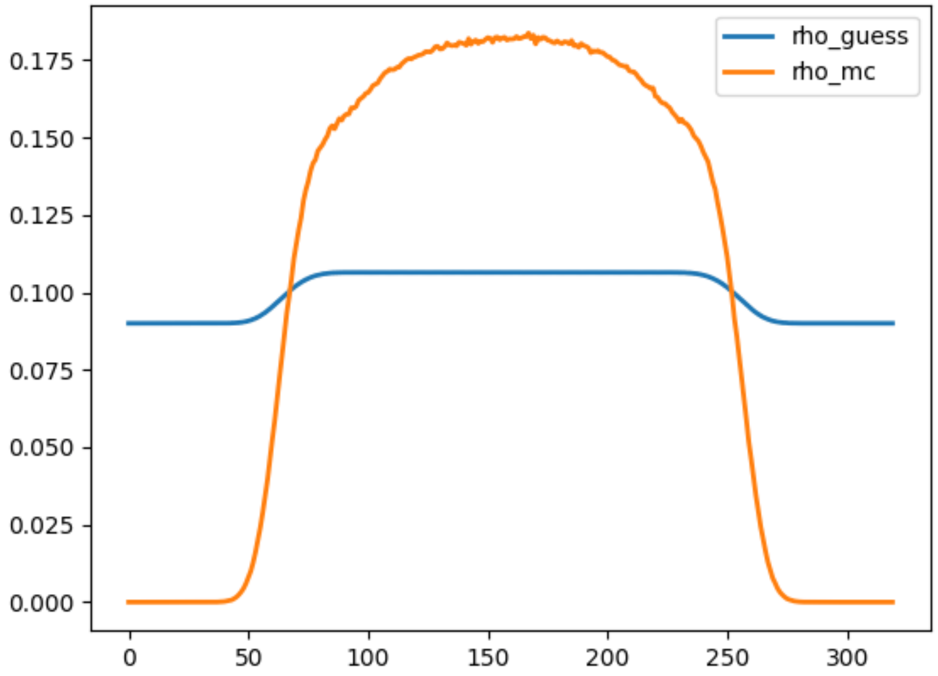


Classical DFT application

Classical DFT at fixed $\beta\mu$ and $\beta V^{\text{ext}}(z)$

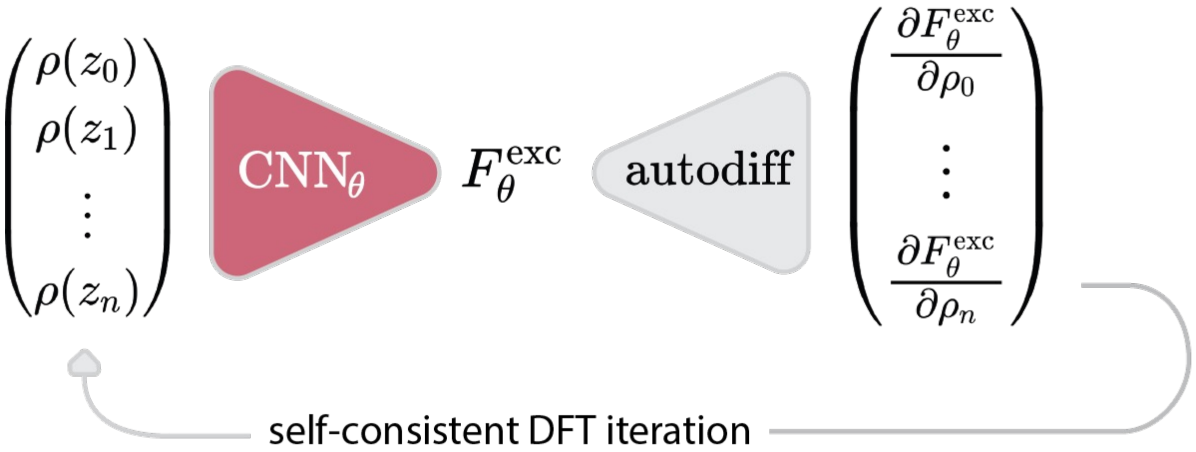


$$\rho(z_i) = \exp\left(\beta\mu - \beta\frac{1}{\Delta z}\frac{\partial F_\theta^{\text{exc}}}{\partial \rho_i} - \beta V_{\text{ext}}(z_i)\right)$$

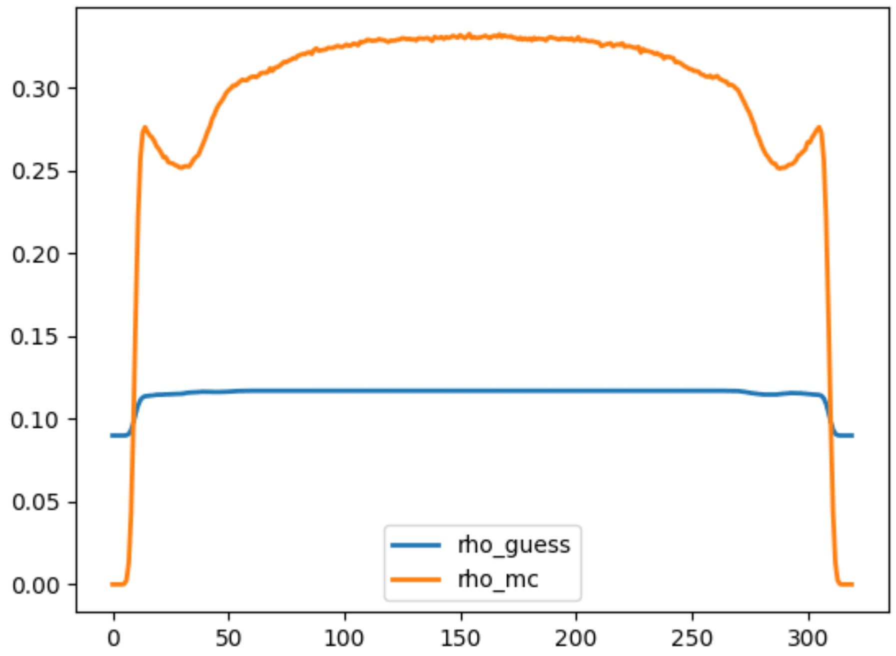


Classical DFT application

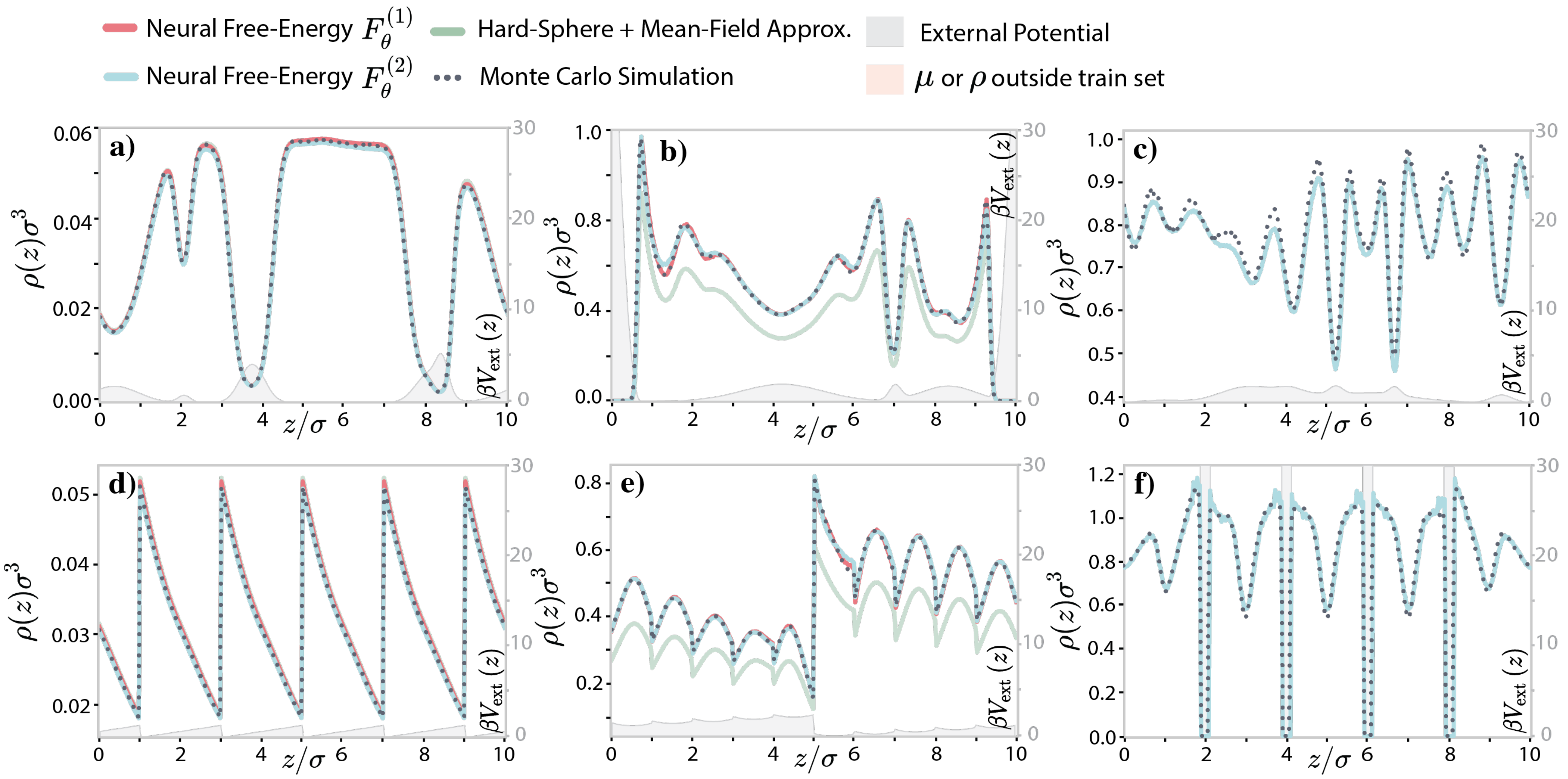
Classical DFT at fixed $\beta\mu$ and $\beta V^{\text{ext}}(z)$



$$\rho(z_i) = \exp\left(\beta\mu - \beta\frac{1}{\Delta z}\frac{\partial F_\theta^{\text{exc}}}{\partial \rho_i} - \beta V_{\text{ext}}(z_i)\right)$$

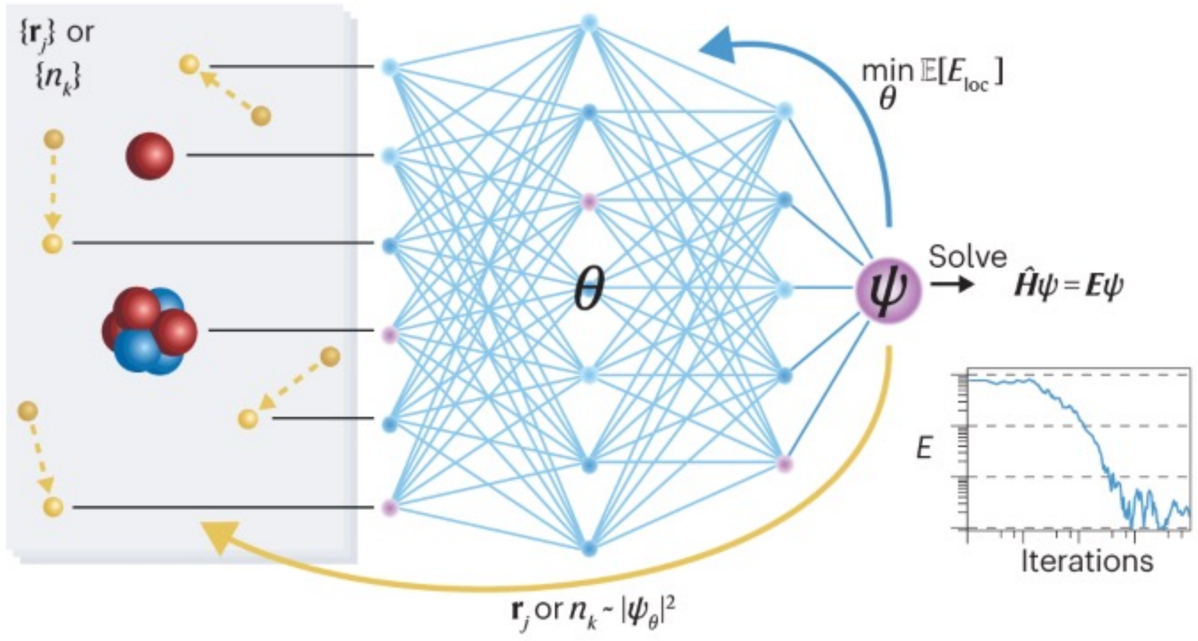


Some Results

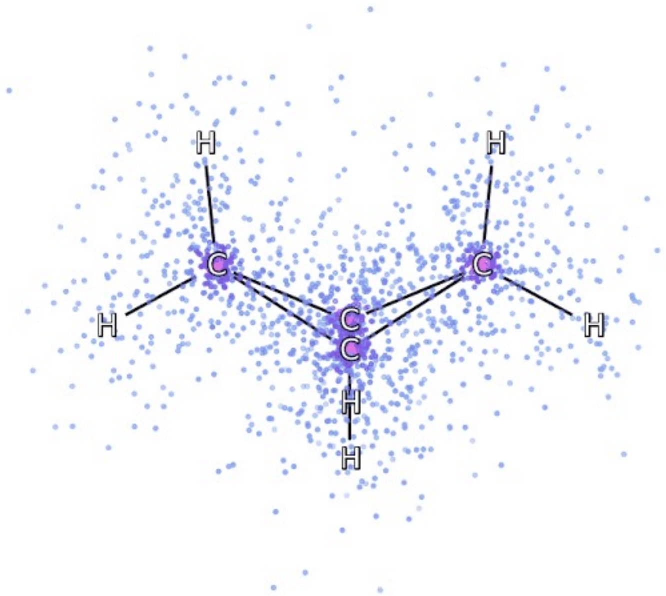


Wasserstein Variational Monte Carlo

- Train a Neural Network to represent the (quantum) wavefunction of the ground state



Source: https://www.nature.com/articles/s41557-020-0544-y



Source: https://deepmind.google/discover/blog/ferminet-quantum-physics-and-chemistry-from-first-principles/

Quantum Variational Monte Carlo

- Minimize Energy of physical system described by Hamiltonian H over wavefunction:

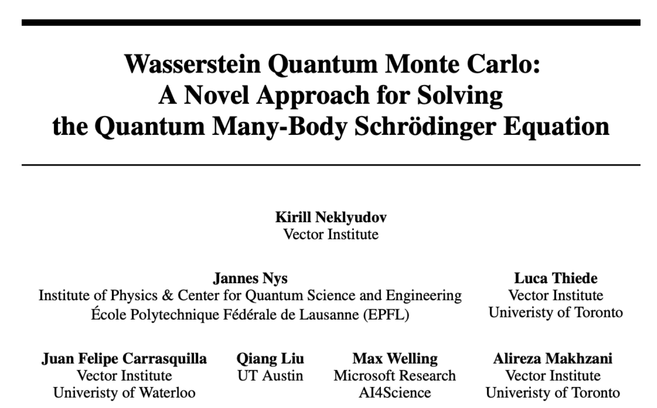
$$E[\psi] = \int dx\ \psi^*(x) H \psi(x) = \int dx\ |\psi(x)|^2\ \frac{H\psi(x)}{\psi(x)} = E_{x\sim|\psi|^2}[\mathrm{E}_{\mathrm{loc}}[\psi]]$$

- Define neural representation for $q(x,\theta) = |\psi(x,\theta)|^2$ and follow gradient:

$$\nabla_\theta E[q(\theta)] = \mathbb{E}_{q(x,\theta)}\left[\left(E_{\mathrm{loc}}(x,\theta) - \mathbb{E}_{q(x,\theta)}[E_{\mathrm{loc}}(x,\theta)]\right)\nabla_\theta \log q(x,\theta)\right]$$

- Claim: this can be viewed as a gradient flow of E[q] in a Fisher-Rao metric + a KL projection.

Geometrical Viewpoint



Wasserstein Quantum Monte Carlo: A Novel Approach for Solving the Quantum Many-Body Schrödinger Equation

Kirill Neklyudov
Vector Institute

Jannes Nys
Institute of Physics & Center for Quantum Science and Engineering
École Polytechnique Fédérale de Lausanne (EPFL)

Luca Thiede
Vector Institute
Univeristy of Toronto

Juan Felipe Carrasquilla
Vector Institute
Univeristy of Waterloo

Qiang Liu
UT Austin

Max Welling
Microsoft Research
AI4Science

Alireza Makhzani
Vector Institute
Univeristy of Toronto

- Introduce manifold for densities by defining a distance between densitities:

$$\mathrm{WFR}(p_0,p_1)^2 := \inf_{v_t,g_t,q_t}\int_0^1 \mathbb{E}_{q_t(x)}\Big[\overbrace{\frac{1}{2}\|v_t(x)\|^2}^{\text{Wasserstein metric}} + \overbrace{\frac{1}{2}g_t(x)^2}^{\text{Fisher-Rao metric}}\Big]dt, \quad \text{subj. to}$$

$$\frac{\partial q_t(x)}{\partial t} = -\nabla\cdot(q_t(x)v_t(x)) + g_t(x)q_t(x)\,, \quad \text{and } q_0(x)=p_0(x),\;\; q_1(x)=p_1(x)$$

Kirill Neklyudov

- Wasserstein metric moves probability mass over a vector field (like a physical flow)
- Fisher-Rao metric allows for nonlocal teleportation of probability mass

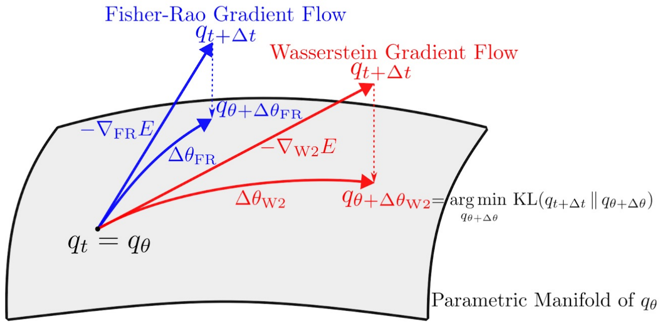
Two Flows to Minimize F[q]

- By using the gradient flow of the Wasserstein or Fisher-Rao metrics we get two gradient flows to minimize F:

$$\frac{\partial q_t}{\partial t}(x) = -\nabla \cdot \left(q_t(x) \left(-\nabla_x \frac{\delta F[q_t]}{\delta q_t}(x) \right) \right), \qquad \text{2-Wasserstein Gradient Flow,}$$

$$\frac{\partial q_t}{\partial t}(x) = -\left(\frac{\delta F[q_t]}{\delta q_t}(x) - \mathbb{E}_{q_t(y)} \left[\frac{\delta F[q_t]}{\delta q_t}(y) \right] \right) q_t(x), \qquad \text{Fisher–Rao Gradient Flow.}$$

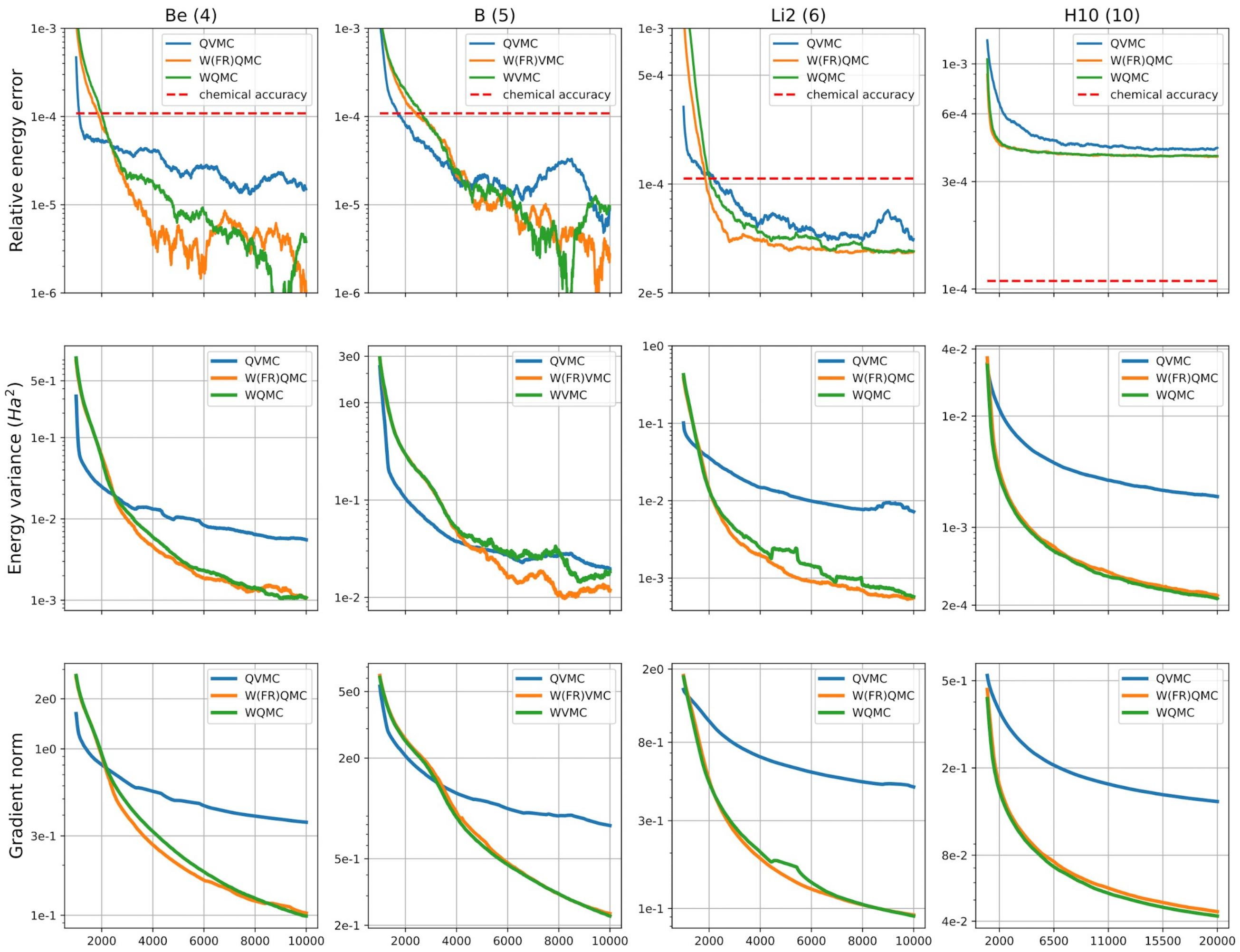
- Equilibrium (in both cases) if: $\left\| \nabla_x \frac{\delta F[q_t]}{\delta q_t}(x) \right\| = 0 \iff \frac{\delta F[q_t]}{\delta q_t}(x) \equiv \text{constant}\,.$



$$\nabla_\theta E[q(\theta)] = \mathbb{E}_{q(x,\theta)} \left[\left(E_{\mathrm{loc}}(x,\theta) - \mathbb{E}_{q(x,\theta)}[E_{\mathrm{loc}}(x,\theta)] \right) \nabla_\theta \log q(x,\theta) \right]$$

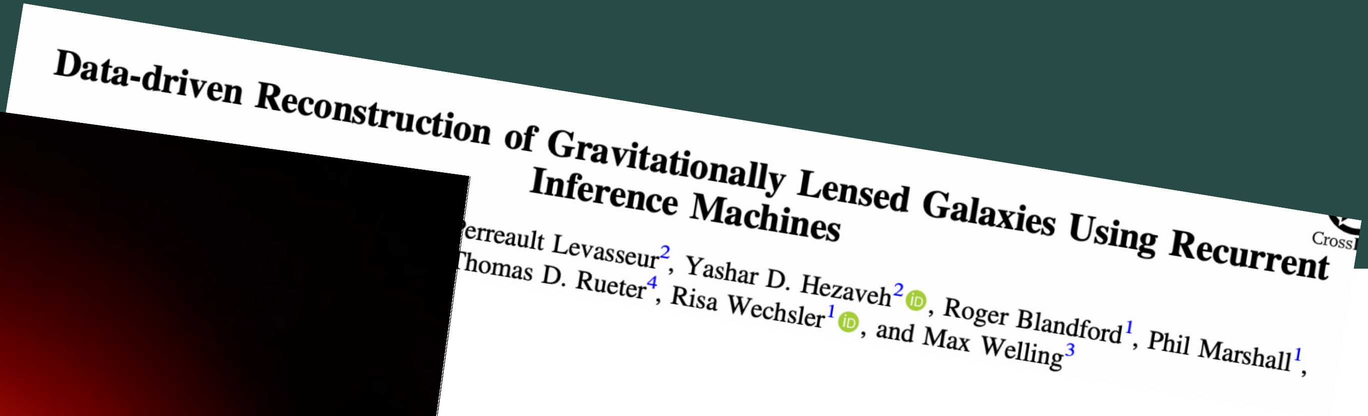
→ F-R Grad. Flow on $E[\psi]$ + Projection

Idea: Use Wasserstein Gradient Flow Instead



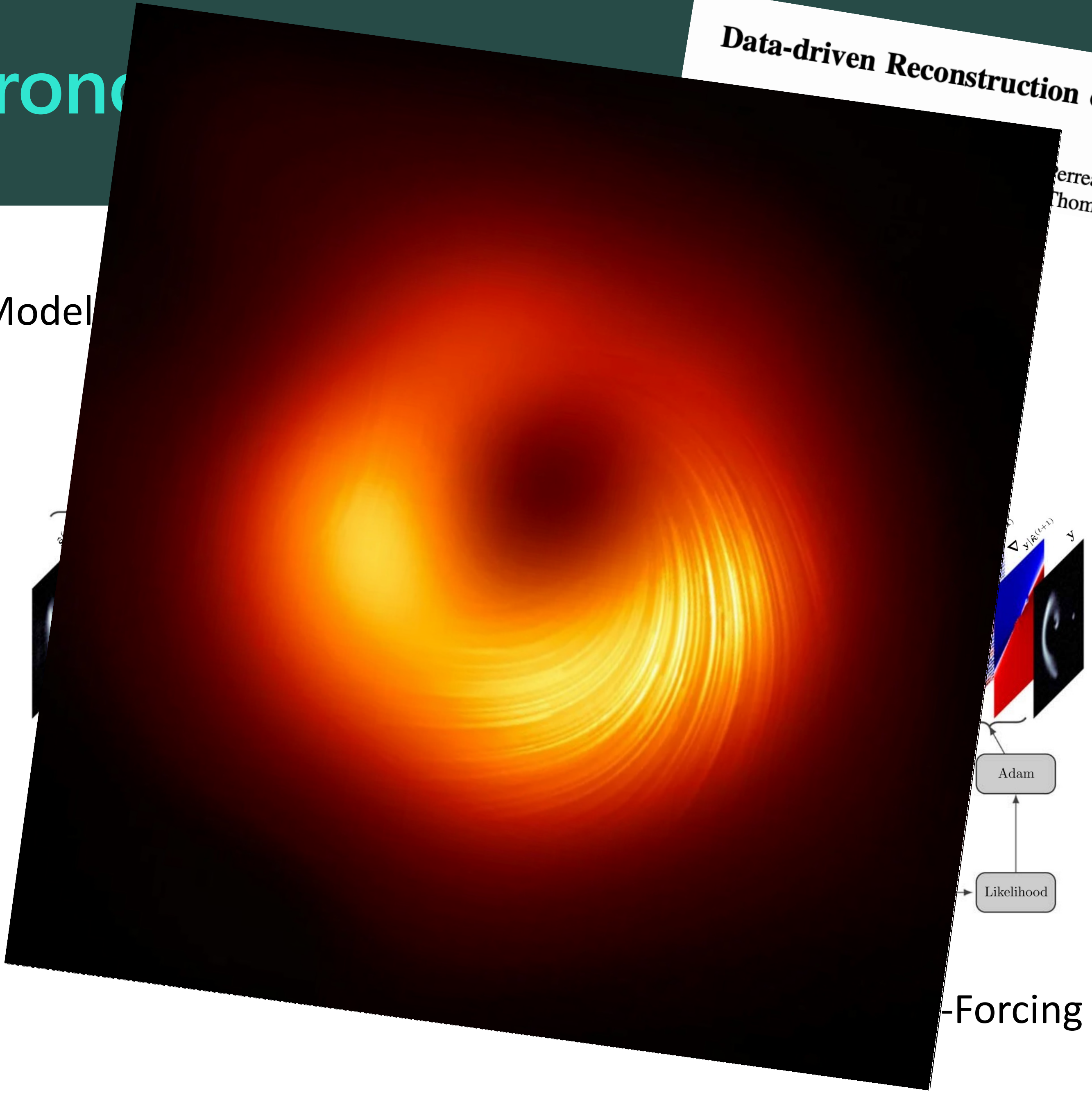
- Estimator has better variance properties.
- Through geometrizing QVMC we could improve the method.

Astrono

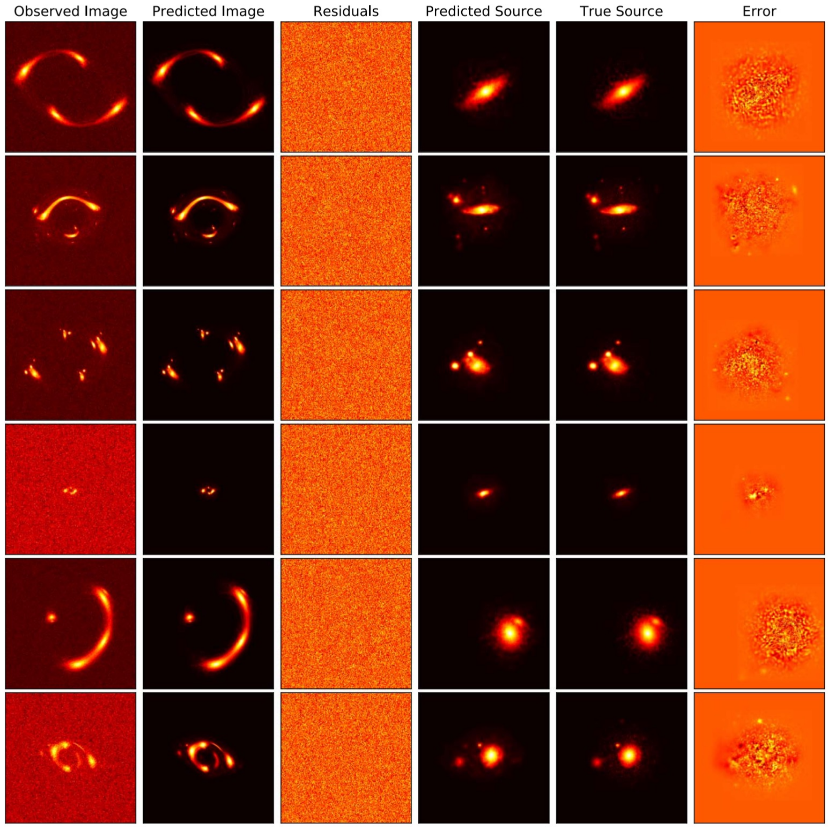

Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines

erreault Levasseur[2], Yashar D. Hezaveh[2], Roger Blandford[1], Phil Marshall[1],
Thomas D. Rueter[4], Risa Wechsler[1], and Max Welling[3]

Model

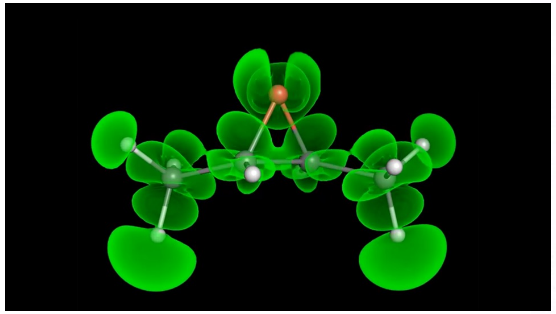


-Forcing



How can AI help the Sciences?

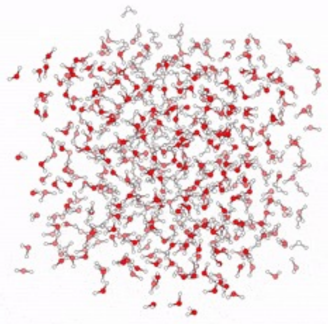
Schrodinger eqn.



$$i\hbar\frac{d}{dt}\psi(t) = H(t)\psi(t)\,.$$

(DFT+ML,
QVMC: Fermi & Pauli Net)

Langevin eqn.

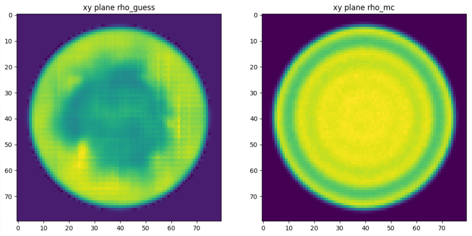


$$\dot{x} = v$$

$$\dot{v} = -\gamma v - \frac{1}{m}\frac{dU}{dx} + \sqrt{2B}\xi(t)$$

(MD+ML=ML Force Fields)

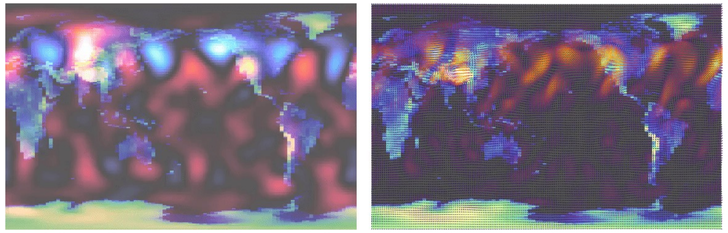
Fokker-Planck eqn.



$$\partial_t \rho(x,t) = \gamma \nabla_x \cdot \left(\rho(x,t)\nabla_x \frac{\delta\Omega[\rho]}{\delta\rho(x,t)}\right)$$

(Classical DFT + ML)

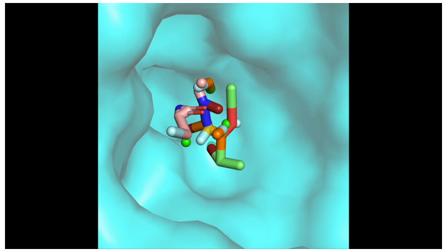
Navier Stokes eqn.



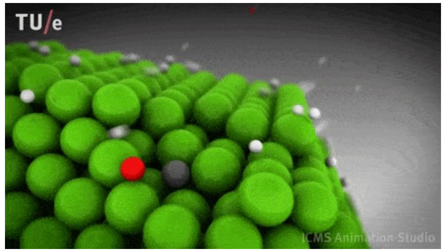
$$\partial \boldsymbol{u}/\partial t = \nu\,\Delta \boldsymbol{u} - (\boldsymbol{u}\cdot\boldsymbol{\nabla})\,\boldsymbol{u} + \boldsymbol{\nabla} p,$$
$$\boldsymbol{\nabla}\cdot\boldsymbol{u} = 0,$$

(PDE Surrogates: FourcastNet,
GraphcastNet)

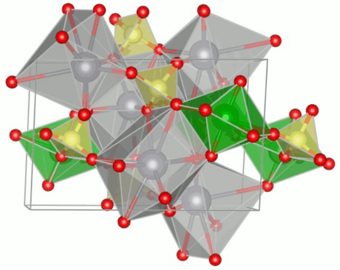
Drug Discovery:
Generating Molecules



Catalysis:
Accelerate reactions.



Materials Science:
Generating materials



A Golden Age of Materials

Time

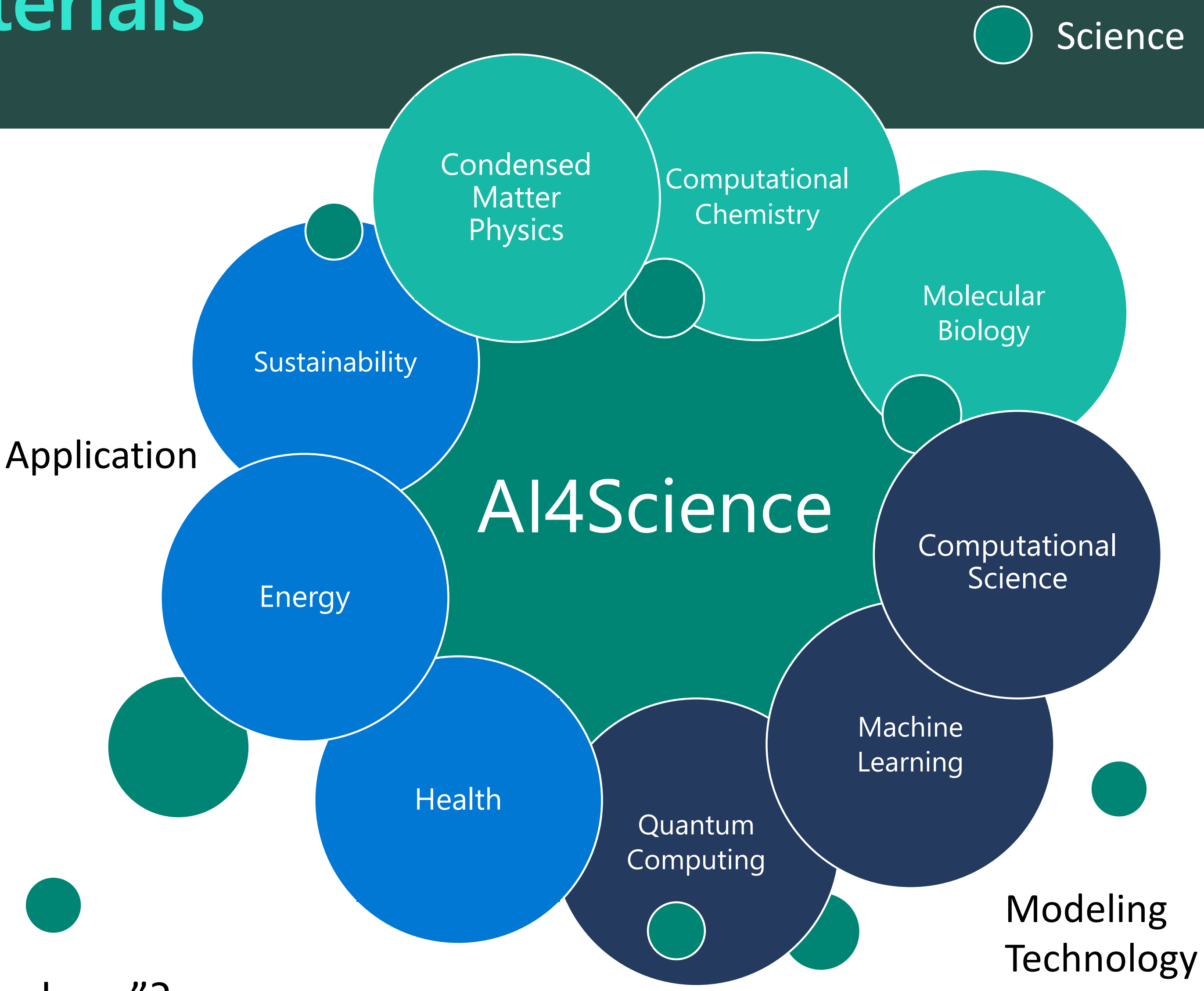
Stone age

Bronze age

Iron age

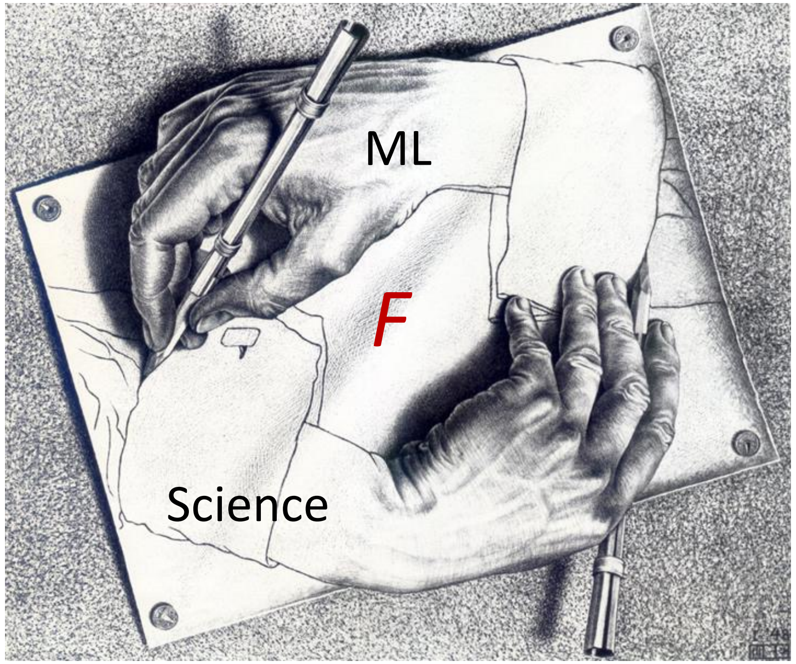
Plastic age

“Material-on-demand age”?



Conclusions

Deep connection between AI and natural sciences
→Free Energy is the bridge



ML is the new hammer for computational scientist.

Applications in health & sustainability: new drugs & materials.