

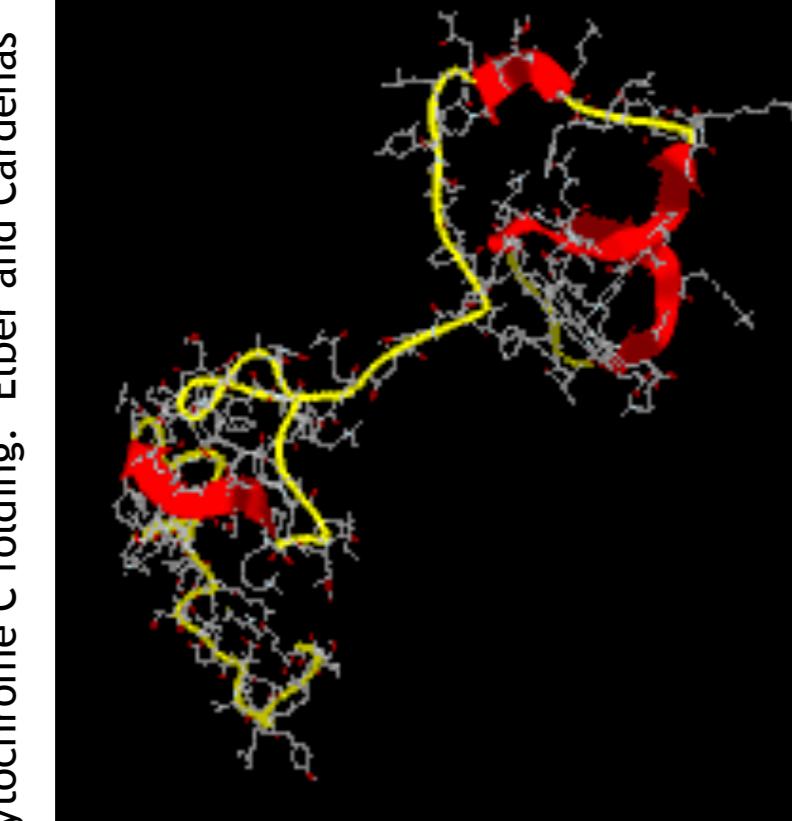
# A random walk from molecular dynamics to data science

**Benedict Leimkuhler**

University of Edinburgh  
The Maxwell Institute for Mathematical Sciences  
and The Alan Turing Institute

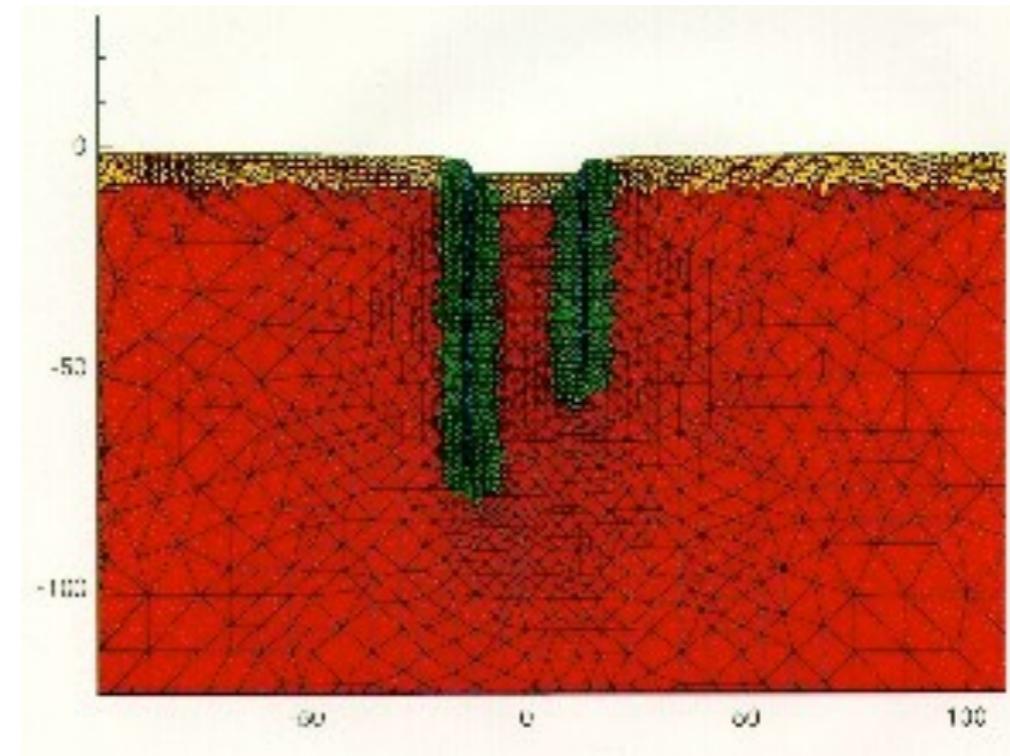
# Molecular Dynamics (MD)

Biomolecules



cytochrome C folding: Elber and Cardenas

Materials



Quasi-continuum method: molecular dynamics with coarse-graining for indentation studies.  
Ortiz, Phillips and Tadmor

# Molecular Dynamics

2017: article in *Drug Discovery Today*

## “Molecular dynamics-driven drug discovery: *leaping forward with confidence*”

“developing a market-approved drug costs a staggering \$2.6 billion”

“Given that the dynamics of the covalent bonds involving hydrogen atoms are not crucial in biological problems, they are usually constrained using integration algorithms, such as SHAKE, RATTLE, ... Hence, a time-step value in the range of **1.5–2 fs** is possible and has been shown to be suitable for MD simulations of many biological systems.”

Examples of some **2017** simulations out of around **20000 published**

Ligand binding in the HIV-integrase enzyme

Druggability of membrane bound Rab5

Fission fragment damage in nuclear fuel

Patchy particle systems

Nanoparticle cholesterol metabolism therapeutics

Multiscale protein dynamics in antigen presentation

Bottlebrush copolymers

Modification of membrane structure by lithium

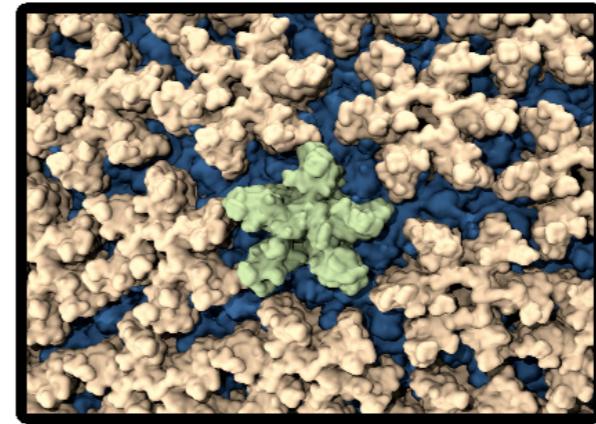
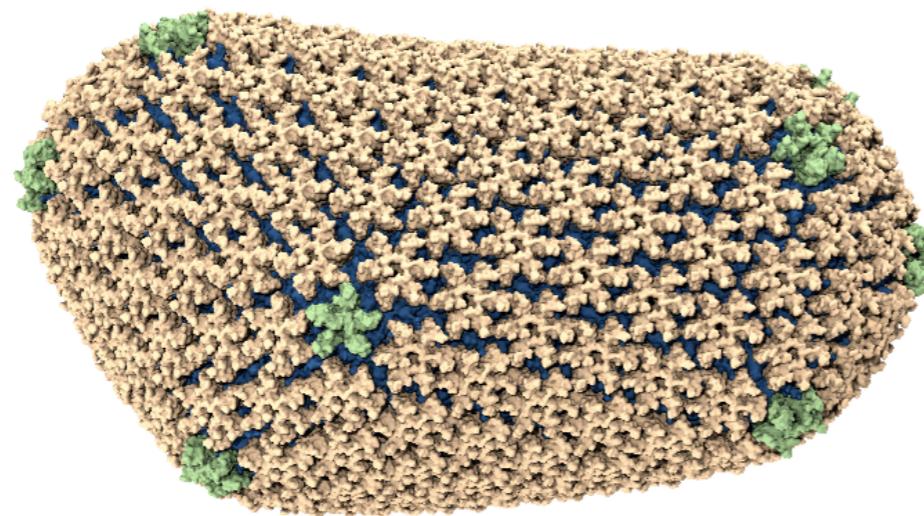
Smectic C semiflexible polymers

Graphene reinforced polymer nanocomposites

Unsaturated Zwitterionic lipids

Identifying the warfarin binding site

# HIV-1 Virus Capsid (protective shield around virus)



Molecular dynamics 64M atoms (including surroundings)  
Described by a potential energy function  $U(x)$

- **Chaotic, nonlinear dynamics, ever expanding scale**
- **Key questions are of a *probabilistic* nature**

# Turning it into maths

Physicists tell us that what they need to do is calculate

$$\int \varphi(x) d\mu(x)$$

observable function  $\varphi(\cdot)$

Gibbs measure  $d\mu(x) \propto e^{-U(x)} dx$

**just integration...but with 180M variables!**

# Sampling and MCMC

$$\mathbb{E}_\mu \varphi(x) = \int_{\mathcal{D}} \varphi(x) d\mu(x) \quad d\mu(x) = Z_\mu^{-1} e^{-U(x)} dx$$

Typical approach: using a Markov Chain with prescribed invariant measure  $\mu$

$$(x_n)_{n=0}^{\infty} \quad \hat{\varphi}_m = \frac{1}{m+1} \sum_{n=0}^m \varphi(x_n)$$

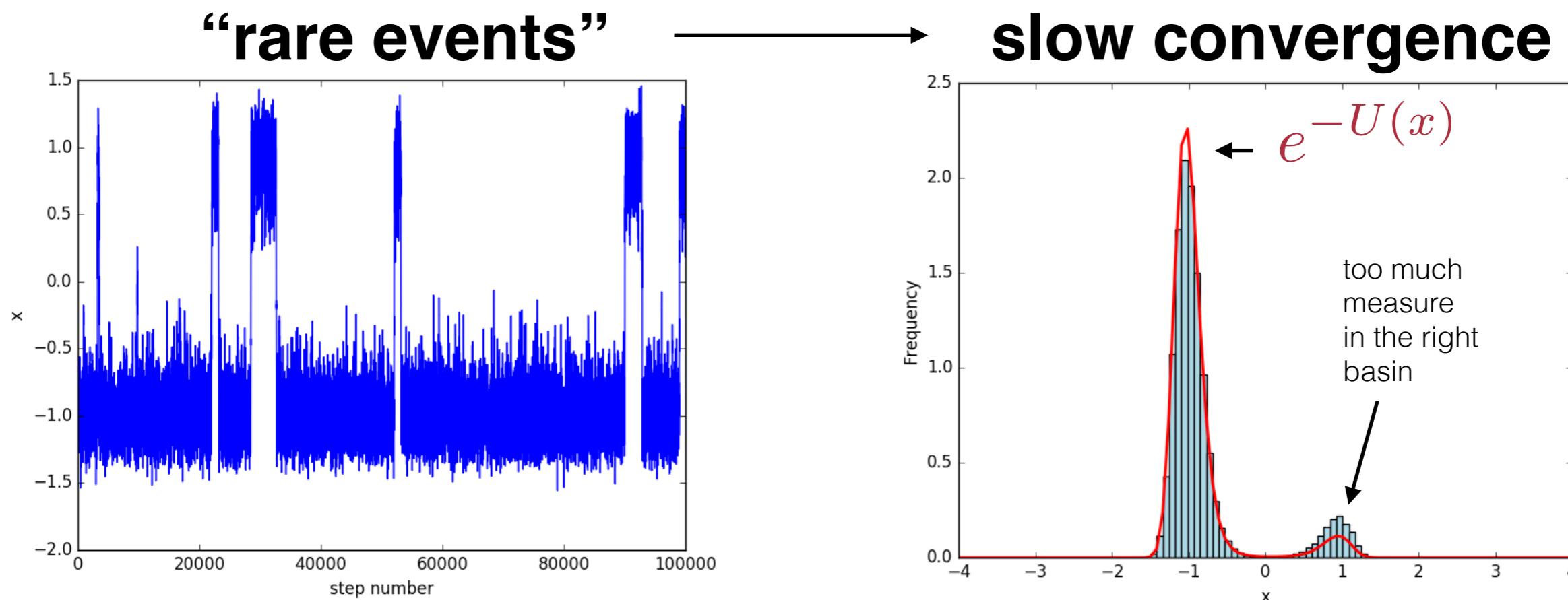
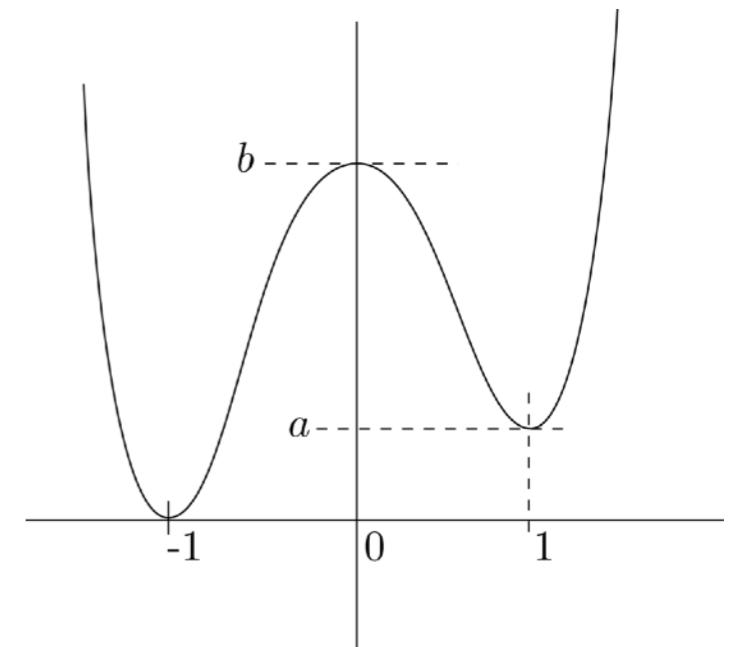
Example method: **Metropolis Monte Carlo**

# Example: Metropolis Monte Carlo for the Uneven Double Well

$$U(x) = (b - a/2)(x^2 - 1)^2 + (a/2)(x + 1)$$

Gaussian prior:

$$g(x|x') = e^{-\frac{(x-x')^2}{2\delta^2}}$$



# Sampling using SDEs

Itō-type Stochastic Differential Equation

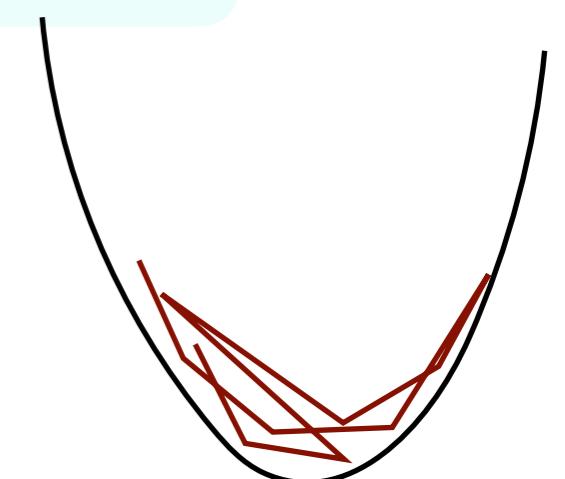
$$dX_t = a(X_t)dt + b(X_t)dW_t$$

Wiener measure

**Brownian Dynamics**

$$dx = -\nabla U(x)dt + \sqrt{2}dW_t$$

*describes a particle diffusing in  
a potential  $U$  at fixed temperature*



# Fokker-Planck Equation

The stochastic paths of an Itō SDE sample the evolving distribution with probability density satisfying

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^\dagger \rho$$

$\mathcal{L}^\dagger$  is the  $L^2$  adjoint of  $\mathcal{L}$  defined by

$$(\mathcal{L}g)(x) = \lim_{\delta t \rightarrow 0} \frac{\mathbb{E}[g(x(\delta t)) | x(0) = x] - g(x)}{\delta t}$$

# Brownian Dynamics

$$dx = -\nabla U(x)dt + \sqrt{2}dW_t$$

$$\mathcal{L}^\dagger e^{-U(x)} = -\nabla \cdot (\nabla U(x)e^{-U(x)}) + \Delta e^{-U(x)} = 0$$

so  $\rho(x) = e^{-U(x)}$  is a stationary density

& **geometric ergodicity** under mild conditions on  $U$

$$\lim_{t \rightarrow \infty} t^{-1} \int_0^t \varphi(x(t))dt = \mathbb{E}_\mu \varphi \quad \textit{Ergodic limit}$$

***Just need a reliable way of computing numerical solutions of SDEs (on long intervals)!***

# SDE Numerics

**What is a suitable definition for “error” for an SDE?**

*Examples: “weak”, “strong”, “ergodic”*

**How to get high accuracy in SDE discretisation?**

**How does discretisation affect convergence to the invariant distribution?**

# Numerical Method

## Euler-Maruyama Method

$$x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_N, \quad Nh = \tau$$

discrete Brownian path

$$\cancel{x_{n+1} = x_n + hF(x_n) + \sqrt{2h}R_n}$$

$R_n \sim \mathcal{N}(0, 1)$

## Leimkuhler-Matthews Method

$$x_{n+1} = x_n + hF(x_n) + \sqrt{h/2}(R_n + R_{n+1})$$

[L. & Matthews, AMRX, 2013]

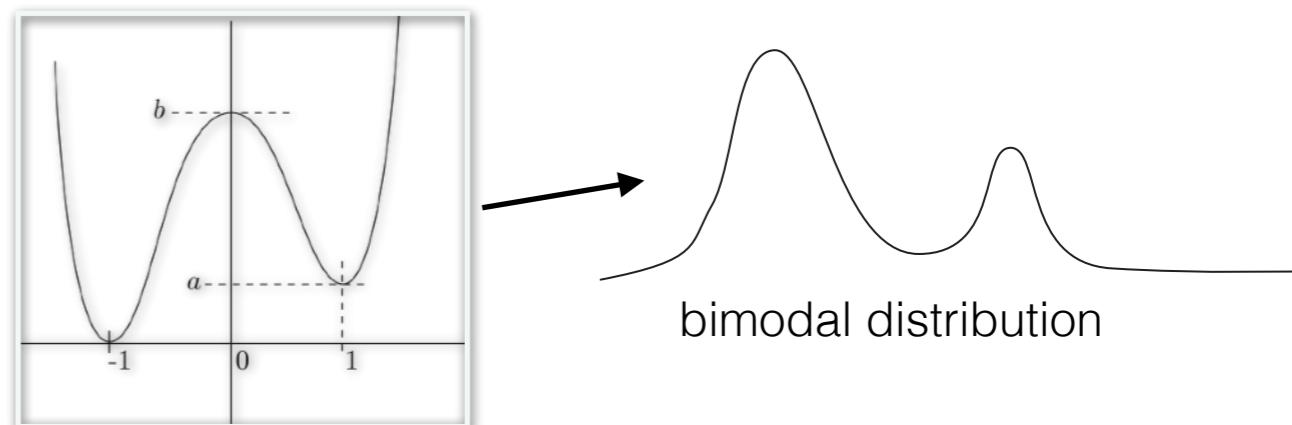
[L., Matthews & Stoltz, IMA J. Num. Anal., 2015]

[L., Matthews & Tretyakov, Proc Roy Soc A, 2014]

# Uneven Double Well

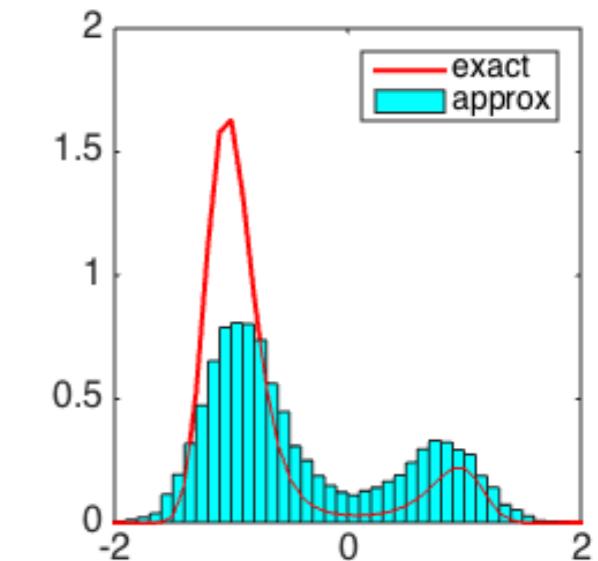
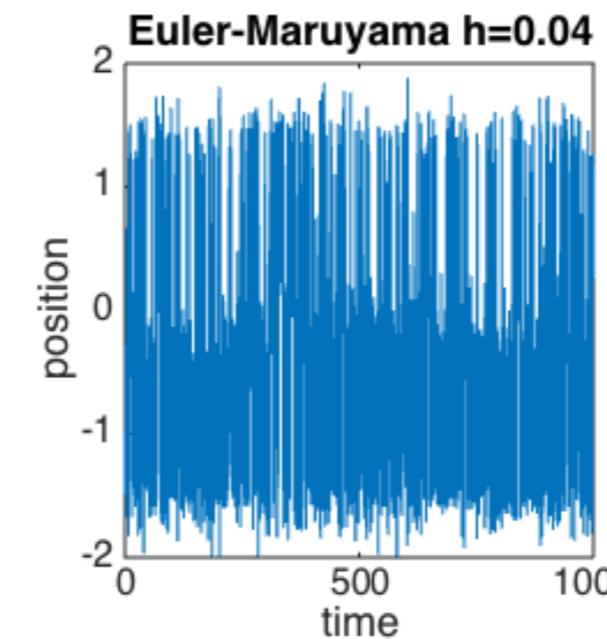
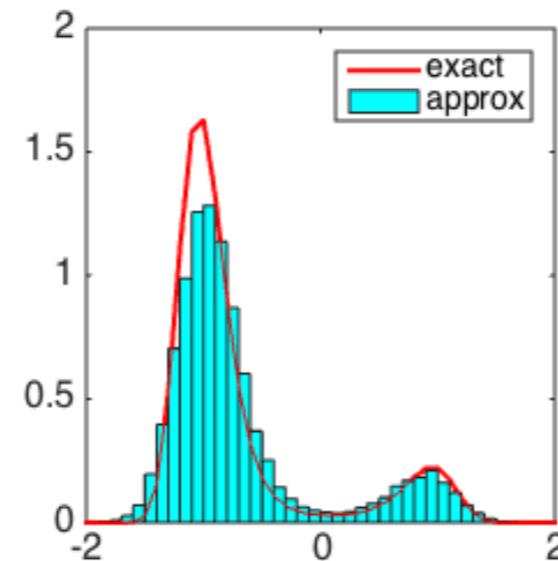
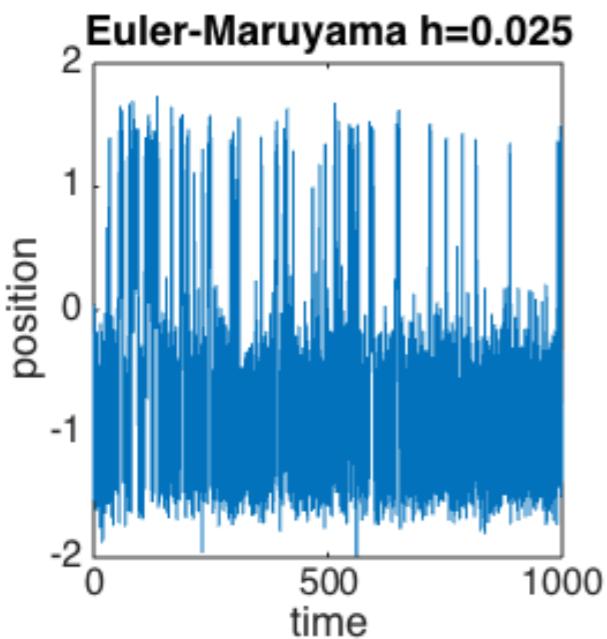
**“ergodic error”**

small stepsize

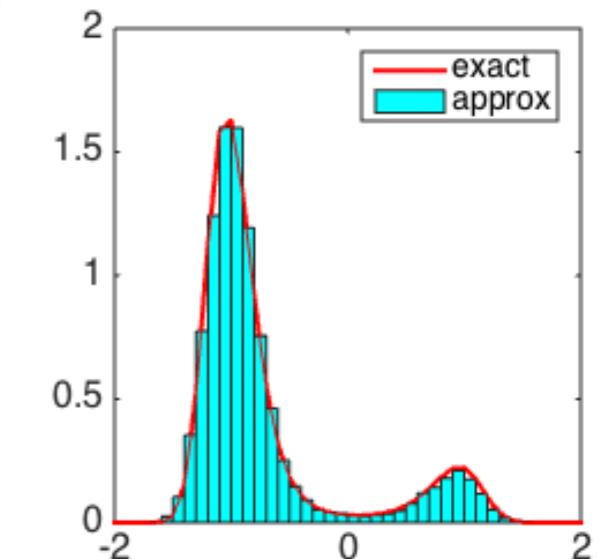
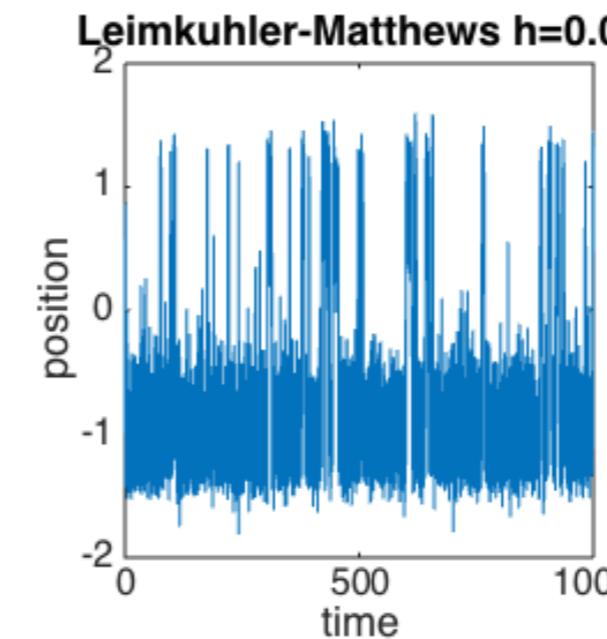
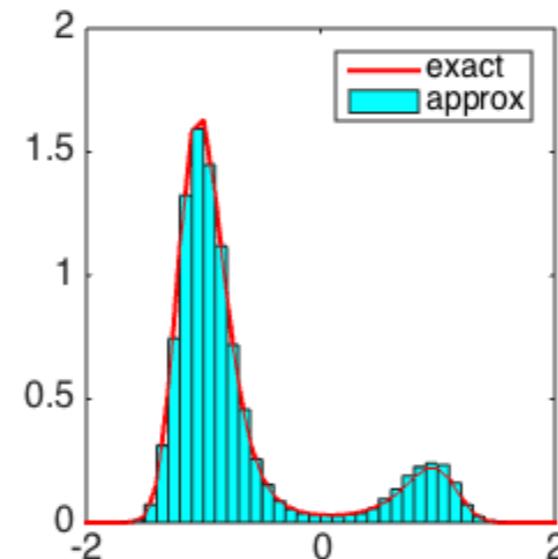
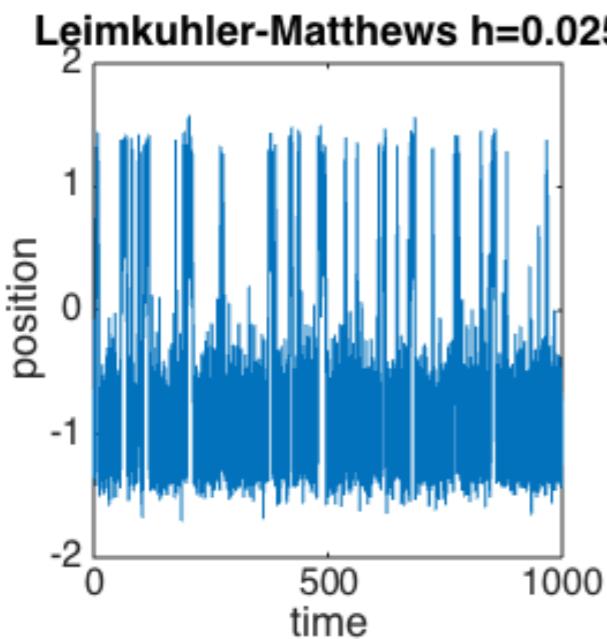


large stepsize

**E-M**



**L-M**





**Charlie Matthews**  
(UoE PhD 2012-2016)

# Invariant measures of numerical methods Sampling Applications

5 papers  
+ a book  
and over 300 citations

B. Leimkuher and C. Matthews (2013) "Stochastic Numerical Methods for Molecular Sampling," Applied Mathematics Research Express, Vol. 2013, No. 1, pp. 34–56  
doi:10.1093/amrx/abs010

Rational Construction of Stochastic Numerical Sampling  
Benedict Leimkuher and Charles Matthews

School of Mathematics and Maxwell Institute of Mathematics  
James Clerk Maxwell Building, Kings Buildings, Edinburgh, Edinburgh EH9 3JZ, UK

Correspondence to be sent to: e-mail: b.leimkuher@ed.ac.uk

In this article, we focus on the sampling of the configuration distribution, that is, the calculation of averages of functions of a molecular  $N$ -body system modeled at constant temperature by a Campbell–Hausdorff expansion, we compare Langevin dynamics with their invariant distributions and demonstrate a superconvergent method in the high friction limit; this method, moreover, simple modification of the Euler–Maruyama method for Brownian non-Markovian (coloured noise) random process. In fully resolved dynamics simulations, for our favored method, we observe up to no evident reduction in the configuration sampling accuracy and a reduction in the size of the largest usable timestep ( $\Delta t$ ) compared to the standard Euler–Maruyama method.

## Abstract

We consider numerical methods for computing sequences of points i.e., computing sequences of points of Gibbs–Boltzmann distribution, using overdamped Langevin dynamics (B) variety of numerical methods for Langevin dynamics simulations, for our favored method, we observe up to no evident reduction in the size of the largest usable timestep ( $\Delta t$ ) propagated exactly in the sense of deterministic alternatives.

The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics  
Benedict Leimkuher, Charles Matthews, Gabriel Stoltz

IMA Journal of Numerical Analysis, Volume 36, Issue 1, 1 January 2016,  
Pages 13–79, <https://doi.org/10.1093/imanum/drv030>  
Published: 29 January 2015 Article

Cite  
Permissions  
Share

“Cite  
Permissions  
Share

THE JOURNAL OF CHEMICAL PHYSICS 138, 174102 (2013)  
Robust and efficient configurational molecular sampling  
via Langevin dynamics

Benedict Leimkuher and Charles Matthews<sup>a)</sup>  
<sup>a</sup>School of Mathematics and Maxwell Institute of Mathematical Sciences, James Clerk Maxwell Building, King's Buildings, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom  
(Received 9 January 2013; accepted 12 April 2013; published online 1 May 2013)

A wide variety of numerical methods are evaluated and compared for deterministic impulses, drifts, and Brownian motions. The method of Hausdorff expansion is used to study sampling accuracy in some oscillators, configuration averaging is exact for certain schemes, while in the modeling of biomolecules where bond stretches performances of the alanine dipeptide reported here (both solvated and systems, an optimal method can be identified that has very low bias). The simulation of the conformal exploration rate (the effective diffusion) is obtained without loss of computational efficiency, while allowing improvement of the conformal exploration rate (the effective diffusion). The optimal scheme is a uniformly better performing algorithm for simulations of 25% or more in practical timestep size. The efficiency improvements of 25% or more in practical timestep size reductions in the error of configurational averages of a factor of 10 are obtained at large timestep. © 2013 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4803034]

Received: 23 February 2016  
Accepted: 8 April 2016  
For timescale restricted to  
Subject Areas:  
Keywords:

applied mathematics, computational chemistry

Author for correspondence:  
Charles Matthews  
e-mail: c.matthews@chicago.edu

Downloaded from <http://rspa.royalsocietypublishing.org>  
PROCEEDINGS A  
rspa.royalsocietypublishing.org

Cite this article: Leimkuher B, Matthews C.  
2013 Efficient molecular dynamics using geodesic integration and solvent-solute splitting  
Proc. R. Soc. A 472: 20120334.  
<http://dx.doi.org/10.1098/rspa.2012.0334>

Research Article

CrossMark

Received: 23 February 2016  
Accepted: 8 April 2016  
For timescale restricted to  
Subject Areas:  
Keywords:

applied mathematics, computational chemistry

Author for correspondence:  
Charles Matthews  
e-mail: c.matthews@chicago.edu

Interdisciplinary Applied Mathematics 39

Ben Leimkuher  
Charles Matthews

Molecular Dynamics  
With Deterministic and Stochastic Numerical Methods

Springer

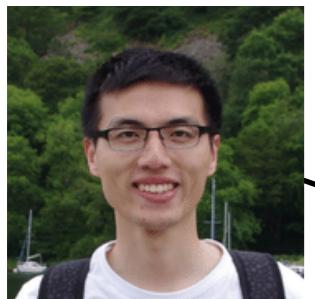
**Charlie  
Matthews**  
**(PhD 2015, now @ Chicago)**



**Matthias Sachs**  
**(PhD 2017, now @ Duke)**



**Xiaocheng Shang**  
**(PhD 2015, now @ETH)**



**Gabriel Stoltz**  
**(ENPC Paris)**

**Michael  
Tretyakov**  
**(Nottingham)**

**Vincent Danos**  
**(ENS Paris)**

Langevin (BAOAB)

Invariant measures

Overdamped Limit

Constrained LD

Langevin/AdL Ergodicity

Adaptive Langevin

# the most recent citations for our 2016 geodesic integrators paper:

## Freezing on a sphere

RE Guerra, CP Kelleher, AD Hollingsworth, PM Chaikin - *Nature*, 2018 - nature.com  
The best understood crystal ordering transition is that of two-dimensional freezing, which proceeds by the rapid eradication of lattice defects as the temperature is lowered below a critical threshold 1, 2, 3, 4. But crystals that assemble on closed surfaces are required by ...  
☆ 99 Cited by 10 Related articles All 6 versions

## Quantifying configuration-sampling error in Langevin simulations of complex molecular systems

J Fass, D Sivak, GE Crooks, KA Beauchamp... - *bioRxiv*, 2018 - bioRxiv.org  
While Langevin integrators are popular in the study of equilibrium properties of complex systems, it is challenging to estimate the timestep-induced discretization error: the degree to which the sampled phase-space or configuration-space probability density departs from the ...  
☆ 99 Cited by 2 Related articles All 4 versions

## Hybrid Monte Carlo methods for sampling probability measures on submani

T Lelièvre, M Roussel, G Stoltz - arXiv preprint arXiv:1807.02356, 2018 - arxiv.org  
Probability measures supported on submanifolds can be sampled by adding an extra momentum variable to the state of the system, and discretizing the associated Hamiltonian dynamics with some stochastic perturbation in the extra variable. In order to avoid biases in ...  
☆ 99 All 3 versions

## Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia

G Minuesa, SK Albanese, A Chow, A Schurer, SM Park... - *bioRxiv*, 2018 - bioRxiv.org  
The MUSASHI family of RNA binding proteins (MSI1 and MSI2) contribute to a wide spectrum of cancers including acute myeloid leukemia. We found that the small molecule Ro 08-2750 (Ro) directly binds to MSI2 and competes for its RNA binding in biochemical ...  
☆ 99 Related articles

## MIST: A Simple and Efficient Molecular Dynamics Abstraction Library for Integrator Development

I Bethune, R Banisch, E Breitmoser, ABK Colla... - arXiv preprint arXiv ..., 2018 - arxiv.org  
We present MIST, the Molecular Integration Simulation Toolkit, a lightweight and efficient software library written in C++ which provides an abstract interface to common molecular dynamics codes, enabling rapid and portable development of new integration schemes for ...  
☆ 99 Related articles All 2 versions

## Quadrature Points via Heat Kernel Repulsion

J Lu, M Sacha, S Steinerberger - arXiv preprint arXiv:1804.02327, 2018 - arxiv.org  
We discuss the classical problem of how to pick  $S$   $N$  weighted points on a  $S$  d-S dimensional manifold so as to obtain a reasonable quadrature rule  $\int_M f(x) d\mu = \sum_{i=1}^N w_i f(x_i)$ . This problem, naturally, has a long ...  
☆ 99 Cited by 1 Related articles All 2 versions

## Systematic derivation of hybrid coarse-grained models

N Di Pasquale, T Hudson, M Icardi - arXiv preprint arXiv:1804.08157, 2018 - arxiv.org  
Significant efforts have been devoted in the last decade towards improving the predictive power of coarse-grained models in molecular dynamics simulations and providing a rigorous justification of their use, through a combination of theoretical studies and data-driven ...  
☆ 99 Cited by 1 Related articles All 2 versions

## Geometric Bayes

A Holbrook - 2018 - escholarship.org  
This dissertation is an investigation into the intersections between differential geometry and Bayesian analysis. The former is the mathematical discipline that underlies our understanding of the spatial structure of the universe; the latter is the unified framework for ...  
☆ 99

## Note on the geodesic Monte Carlo

A Holbrook - arXiv preprint arXiv:1805.05289, 2018 - arxiv.org  
Geodesic Monte Carlo (gMC) comprises a powerful class of algorithms for Bayesian inference on non-Euclidean manifolds. The original gMC algorithm was cleverly derived in terms of its progenitor, the Riemannian manifold Hamiltonian Monte Carlo (RMHMC). Here ...  
☆ 99 Cited by 1 Related articles All 2 versions

physical chemistry

biomolecules

data science

biomolecules

molecular software

data science

physics

stats & data science

stats & data science

*Why are statisticians reading molecular dynamics papers?*

# Wind Turbines

(with industry partner)



**Zofia Trstanova**  
(postdoc)



**Anton Martinsson** (PhD student)



Financial giant Goldman Sachs announces that it plans to invest \$150 billion in clean energy projects and technology like solar and wind farms, energy efficiency upgrades for buildings, and power grid infrastructure.

**Problem:** Goldman Sachs analysts' lack of turbine knowledge

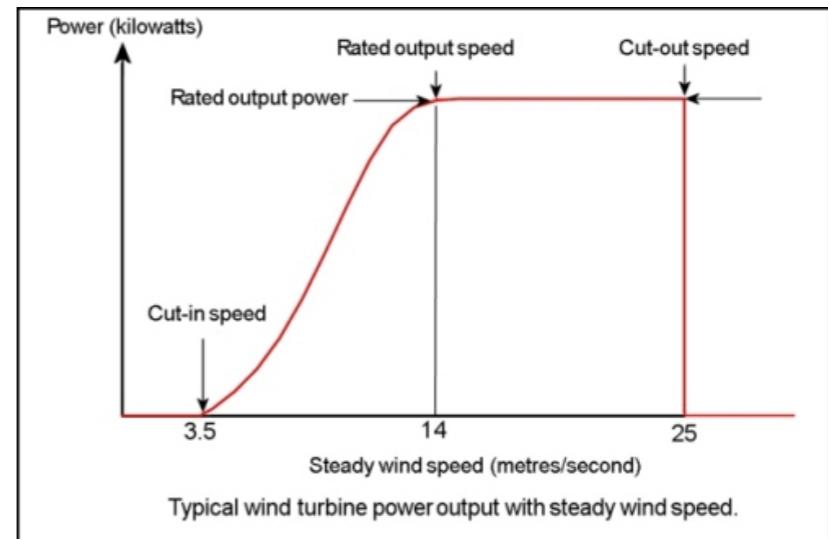
**Solution:** Employ a specialist energy management firm

**Problem:** Each windfarm may have hundreds of turbines each turbine produces data at 10s to 10m intervals on dozens of properties ⇒ **Large data problem [10K turbine years!]**

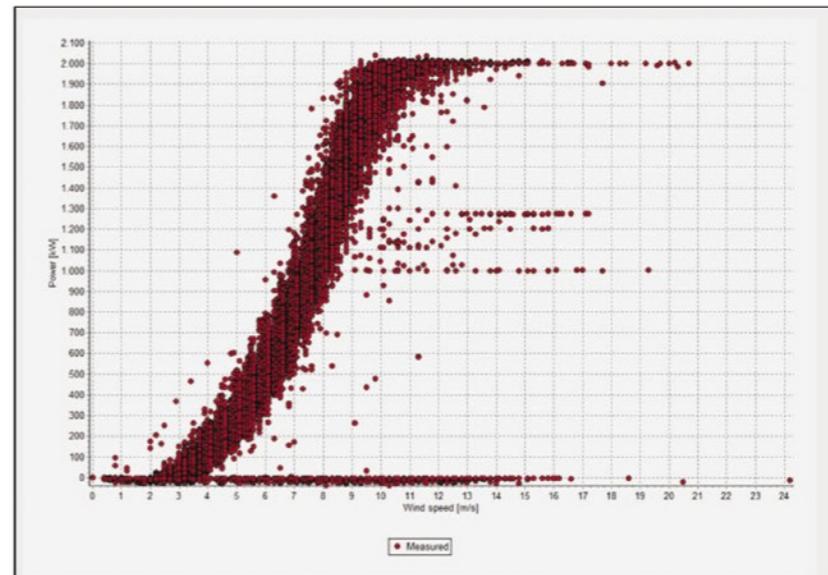
**Solution 1:** Hire engineers to stare at data

**Solution 2:** Get a team of mathematicians to develop a data-driven model to analyse the data.

## Ideal power output to windspeed diagram



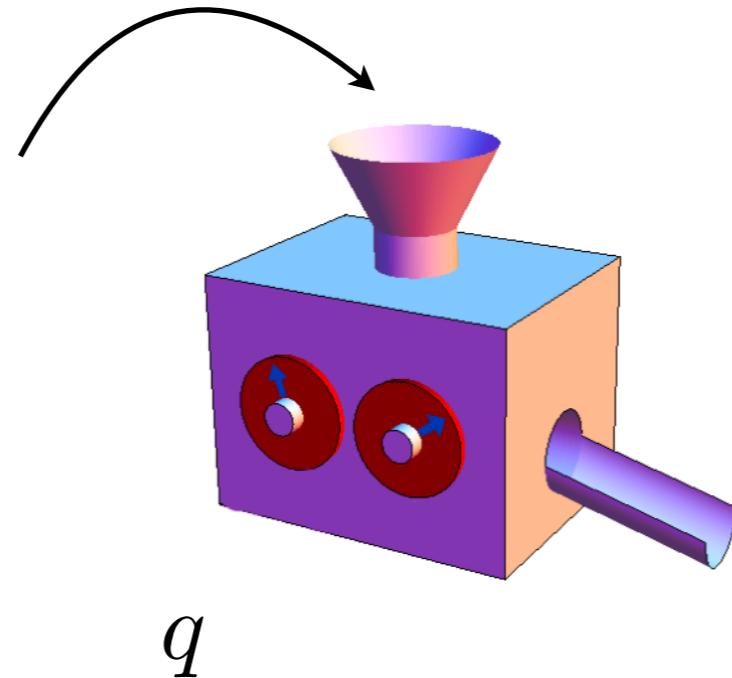
## Example data set



**Learn:** to categorise conditions: “**derating**”, “**icing**”, ...

**Goal:** Automatic analysis of data streams for maintenance, performance analysis, planning

# Model



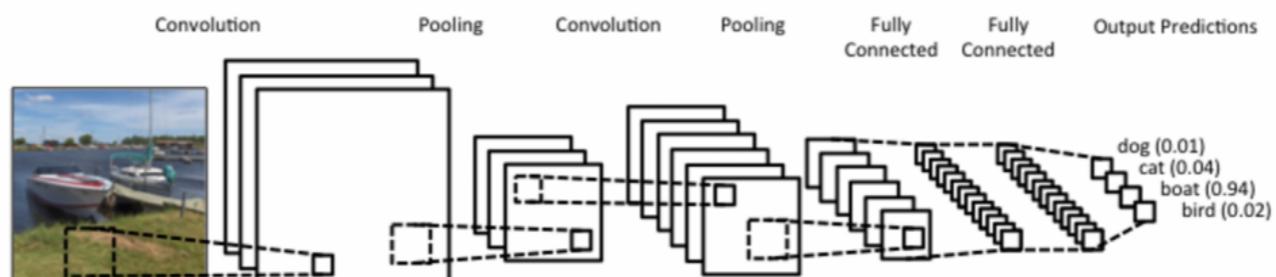
$$P_{\text{cat}} = 0.97$$

$$P_{\text{dog}} = 0.02$$

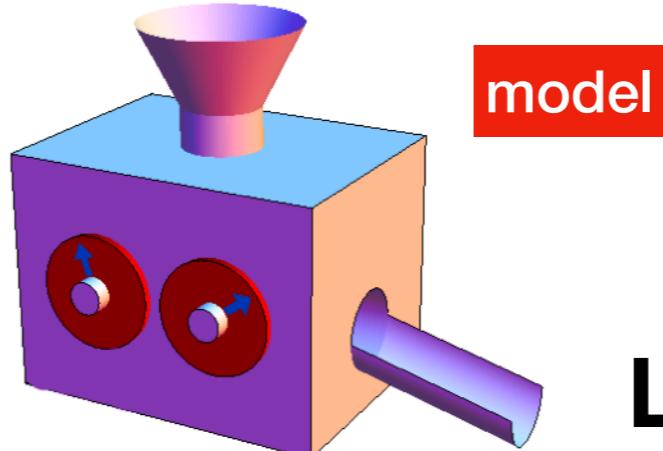
$$P_{\text{other}} = 0.01$$

Models may be artificial or not,  
but the distinction is often artificial!

Ex: convolutional neural network



**DATA**  $X = \{x_1, x_2, \dots, x_N\}$



## PARAMETERS

$$q = (q_1, q_2, \dots, q_d)$$

$$\text{prior } \pi_0(q)$$

**LIKELIHOOD**  $\pi(X|q)$

$$\pi(X|q) = \prod_{i=1}^N \pi(x_i|q)$$

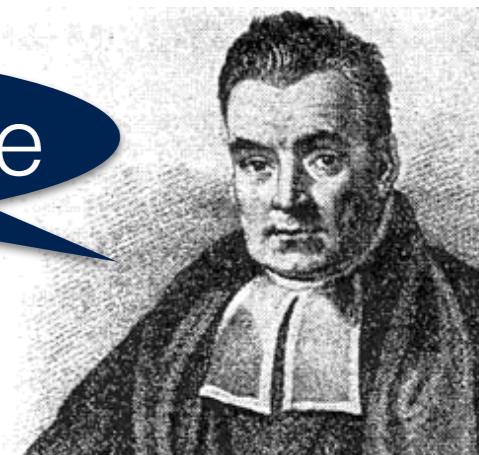
$$\pi(q|X) \propto \pi(X|q)\pi_0(q) \quad \textbf{Bayes' Theorem}$$

$$\pi(q|X) \propto \exp(-U(q)), \quad U(q) = -\log \pi(X|q) - \log \pi_0(q)$$

minimise  $U(q)$  to determine ‘optimal’ parameters

I'm not really me

*Data Scientist Thomas Bayes, U of Edinburgh, Class of 1721*



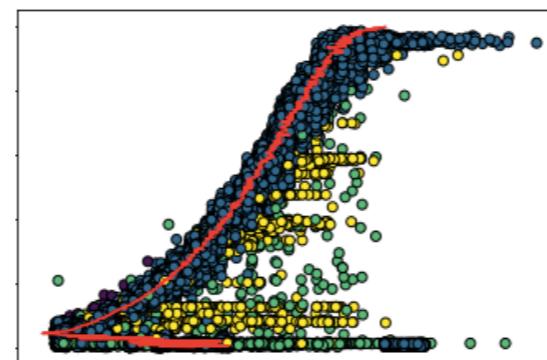
# Wind Turbine Benchmarking

With enough data, it is straightforward to replace the engineer with a convolutional neural network for detecting a specific condition, e.g. icing

2 Convolutional NN (nws, pwr, rotorspd, pitch)		
	icy	icy 2
trained on icy	98% (100%, 93%, 99%, 1%, 3%)	97% (100%, 72%, 98%, 12%, 0%)
trained on icy 2	97% (99%, 87%, 96%, 0%, 0%)	97% (100%, 74%, 97%, 0%, 0%)

## Problems:

1. Mixed data sets
2. Lack of transferability
3. Multimodality

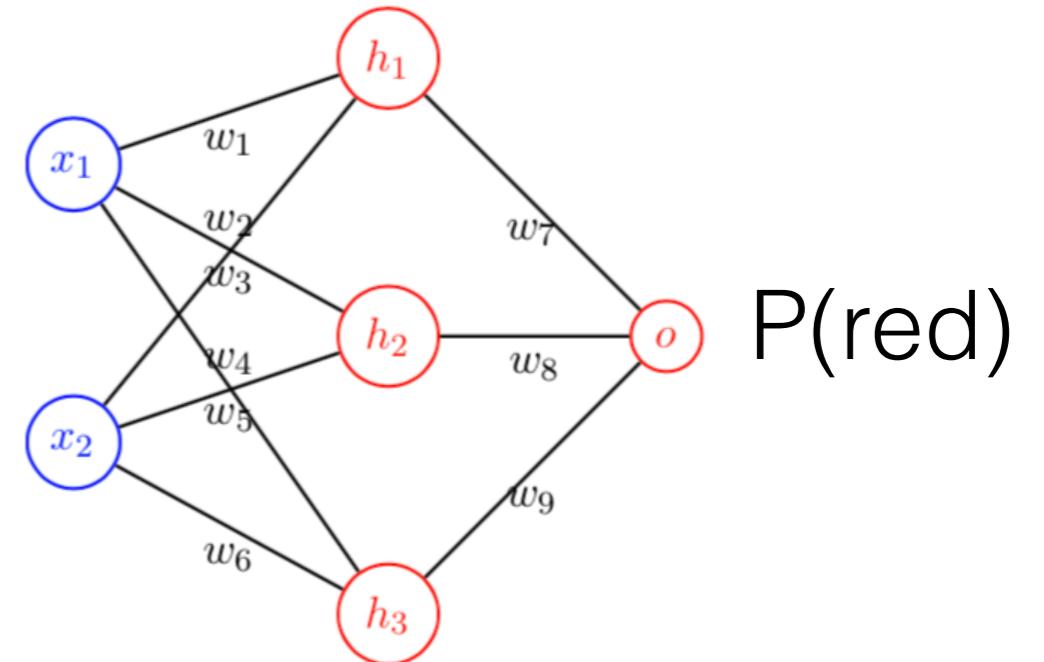
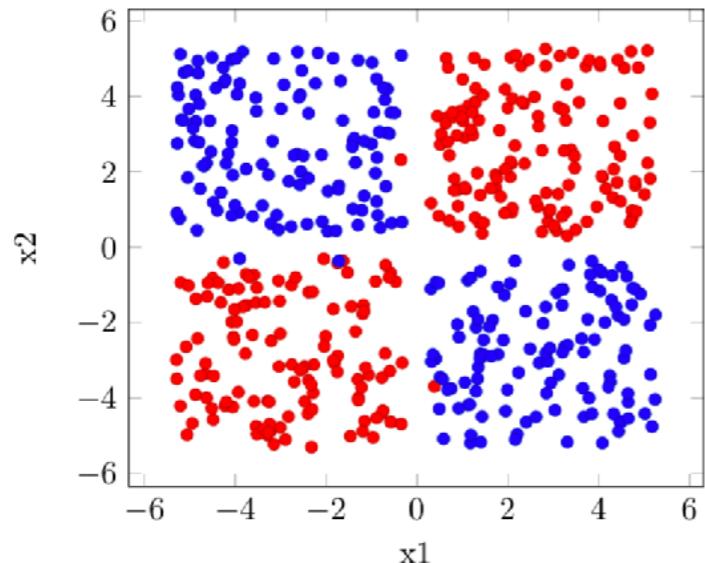


- **Unsupervised learning** to learn classes from data
  - **Generalisation error**
  - **Generative networks**

all require sampling!

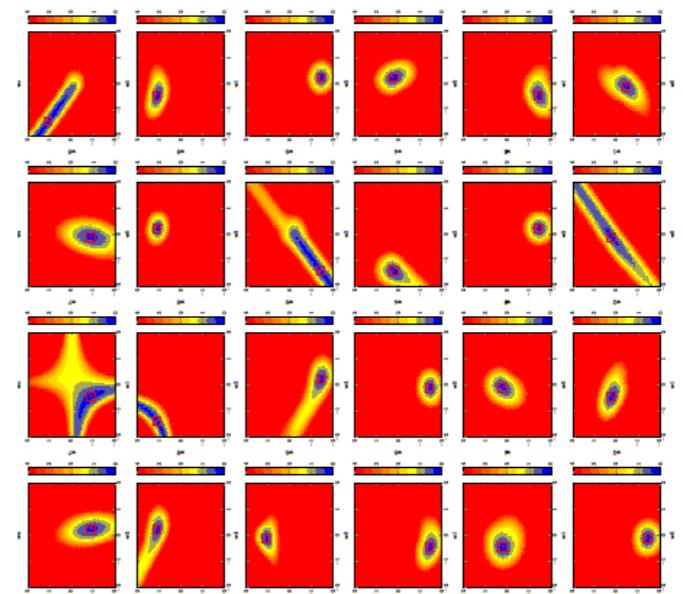
# Multimodality in NNs

With **Zofia Trstanova** and **Frederik Heber** (Turing)

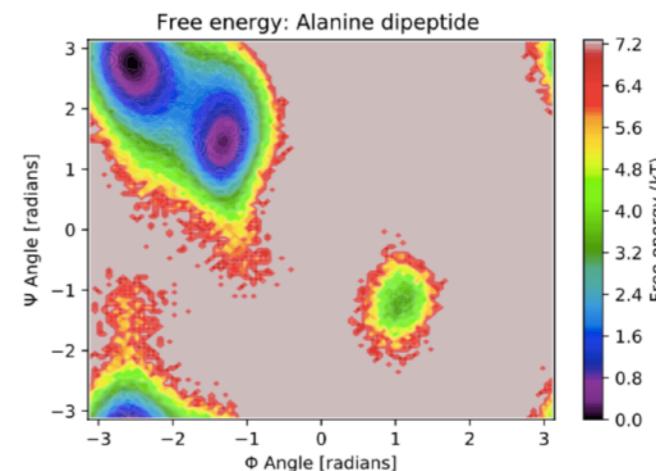
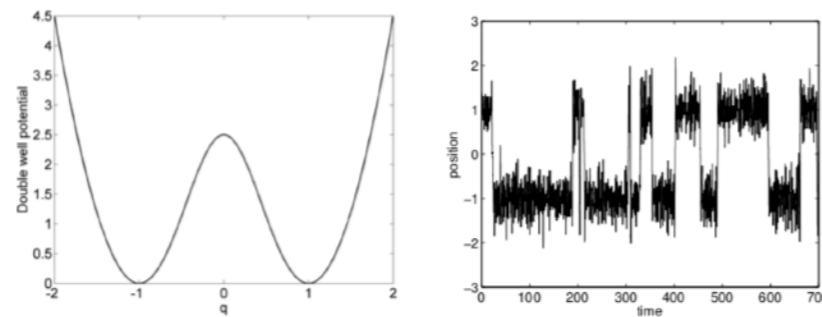


isolated local min

9 parameter model, multimodal loss landscape with pretty severe trapping in upper well.



# Sampling Metastable Systems

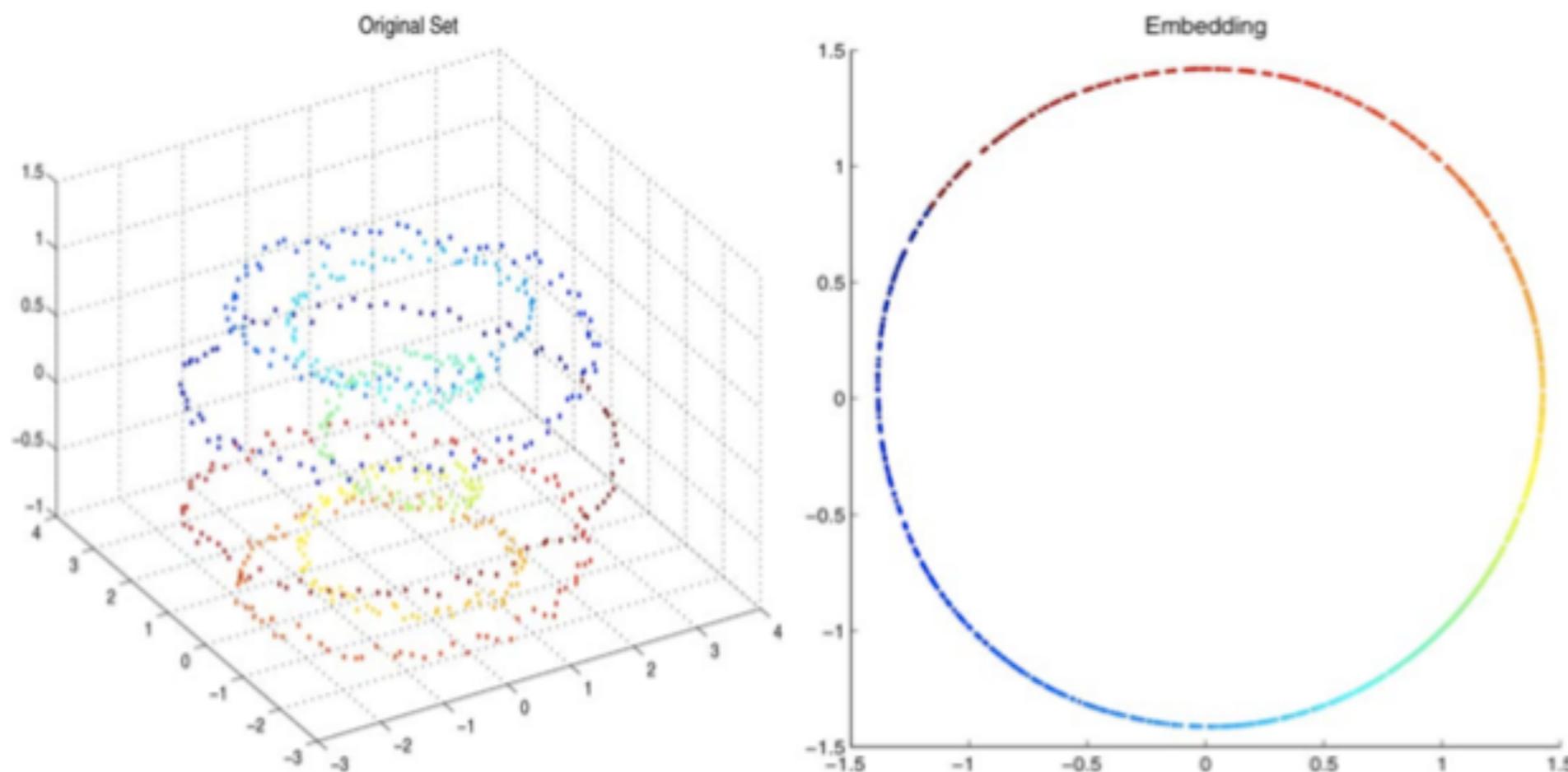


## Ways forward:

1. **EQN.** **New types of diffusions** that more efficiently explore the complex geometry; in particular “**non-reversible diffusions.**” **w. J. Weare & C. Matthews, Chicago.**
2. **ISST.** Use of **temperature** and other “**collective variables**” to enhance and guide sampling. **w. A. Martinsson, J. Lu (Duke) & E. Vanden-Eijnden (NYU).**
3. **DM.** Data-analytic tools based on **diffusion maps** to perform **unsupervised manifold learning** from simulated data in order to enhance exploration. **w. Z. Trstanova, T. Lelievre (Paris) and R. Banisch (Berlin)**

## Manifold learning

- Data points  $\mathbb{D}^{(m)} := \{x_1, x_2, \dots, x_m\} \subset \mathbb{R}^N$  with  $N > 0$
- Compact  $d$ -dimensional differentiable submanifold  $\mathcal{M} \subset \mathbb{R}^N$ , sampled according to a density  $\mu(x)$ .
- Geometric structure of  $\mathcal{M}$  from the data  $\mathbb{D}^{(m)}$  by constructing an  $m \times m$  matrix that approximates a differential operator
- Diffusion maps (Coifman, Lafon, 2006): learn structure on  $\mathcal{M}$  through approximation of a differential operator



# Diffusion maps

Construction:

- Isotropic-kernel  $k_\varepsilon(x, y) = \exp(-(4\varepsilon)^{-1}\|x - y\|^2)$  on  $\mathbb{D}^{(m)}$
- ① Evaluate the kernel to get  $m \times m$  kernel matrix  $K_\varepsilon$
- ② Compute kernel density estimate  $q_i = \sum_{j=1}^N K_{ij}$
- ③ Row normalize  $K_\varepsilon$  using  $\alpha$  power of  $q$  to obtain transition matrix  $P = D^{-1}K_\varepsilon$
- ④ Graph Laplacian matrix  $L_{\varepsilon, \alpha} = \varepsilon^{-1}(P - I)$

## Pointwise convergence to Kolmogorov operator

Let  $f : \mathcal{M} \rightarrow \mathbb{R}$  be smooth,  $\alpha \in [0, 1]$ , then in the limit  $m \rightarrow \infty$  and  $\varepsilon \rightarrow 0$ , for any point  $x_k \in \mathbb{D}^{(m)}$ ,

$$(L_{\varepsilon, \alpha}[f])_k \rightarrow \mathcal{L}f(x_k), \quad \mathcal{L} = \Delta f + (2 - 2\alpha)\nabla f \cdot \frac{\nabla \mu}{\mu}.$$

- Case  $\mu \propto \exp(-\beta V)$  with  $\beta > 0$ , and  $\alpha = 1/2$

$$\mathcal{L}f = \Delta f - \beta \nabla V \cdot \nabla f,$$

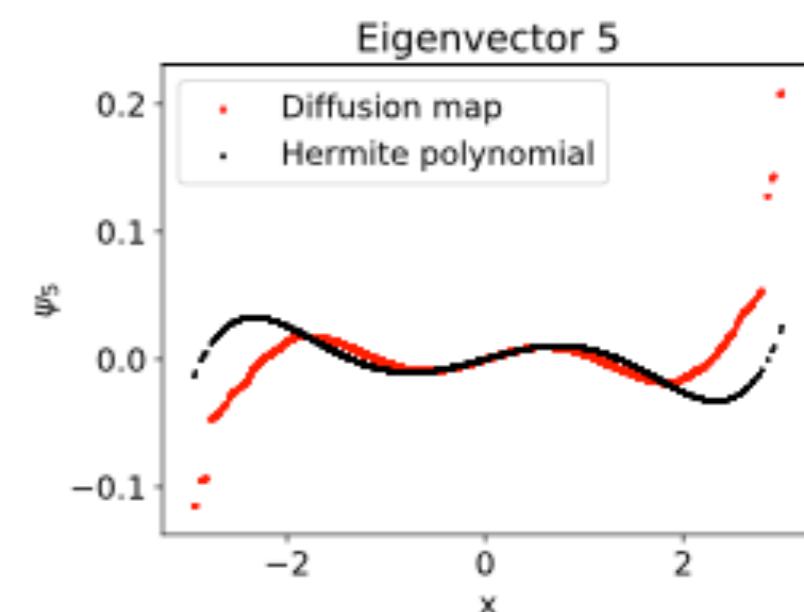
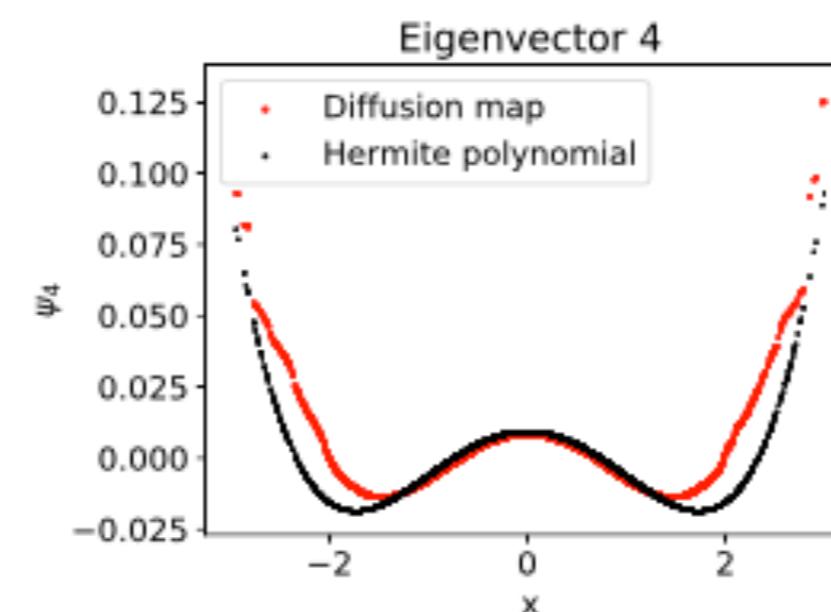
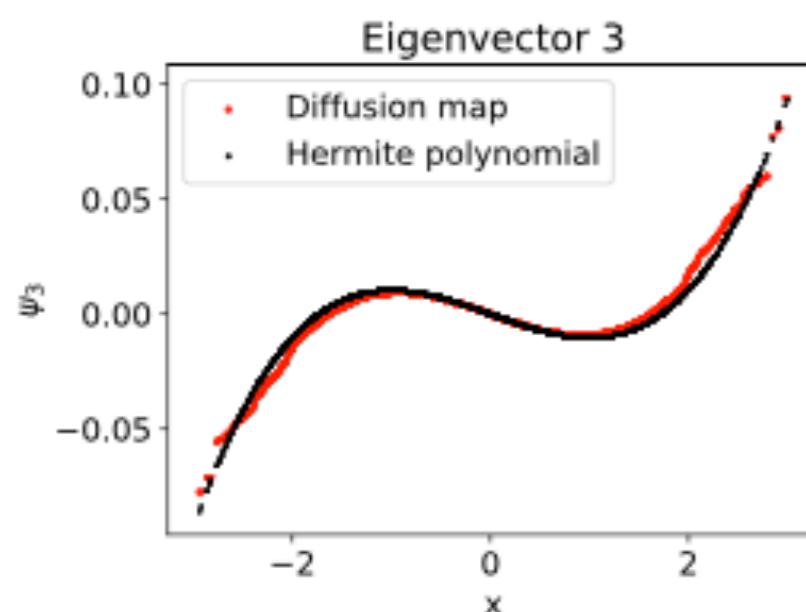
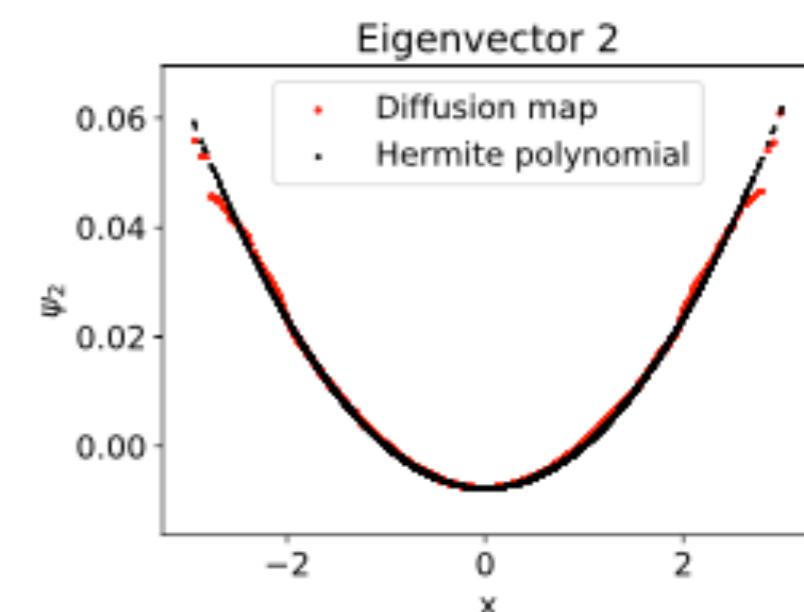
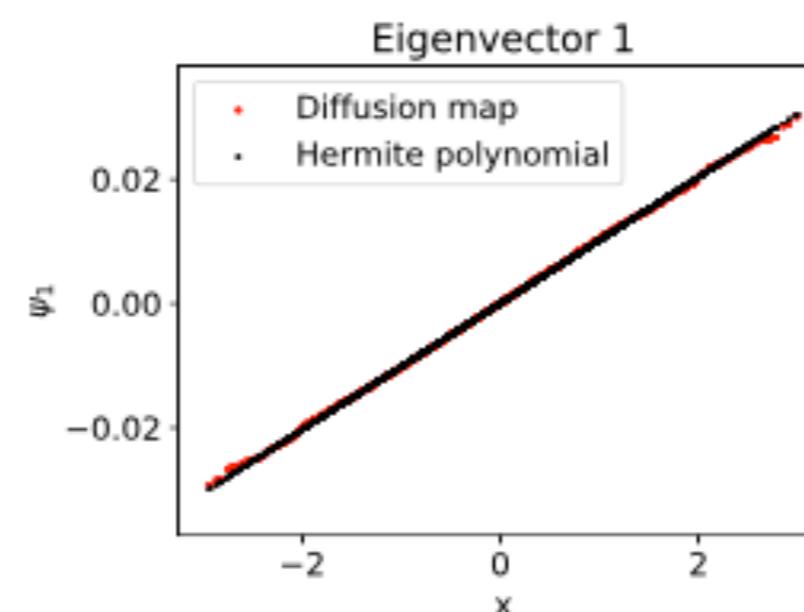
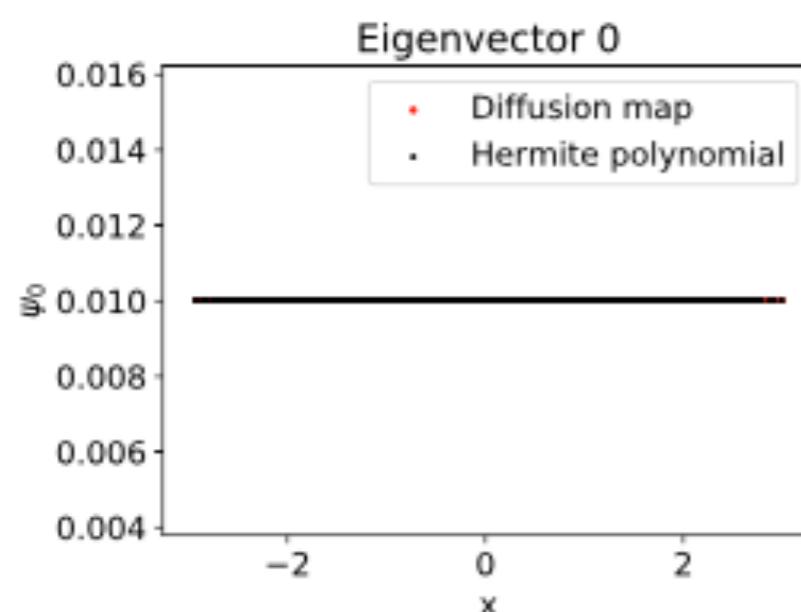
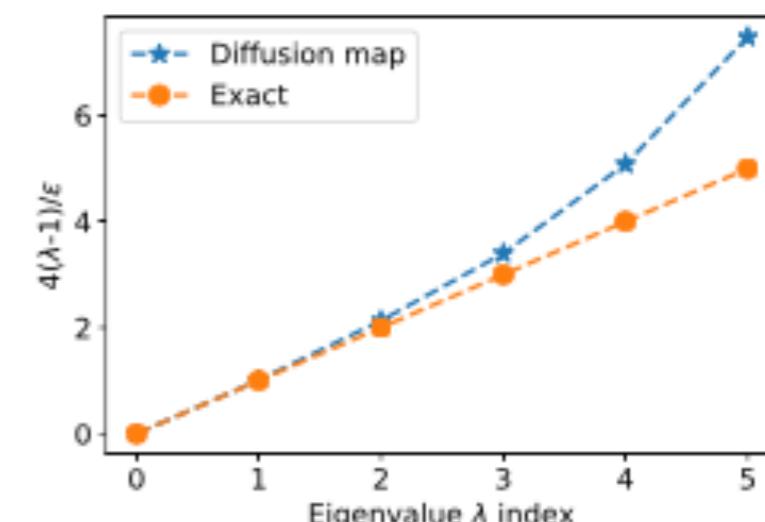
- Approximation error  $|L_{\varepsilon, \alpha}f(x_i) - \bar{L}_{\varepsilon, \alpha}f(x_i)| = O(m^{-1/2}\varepsilon^{-d/4-1/2})^1$

---

<sup>1</sup>A. Singer, 2005, R. R. Coifman and S. Lafon, 2006.

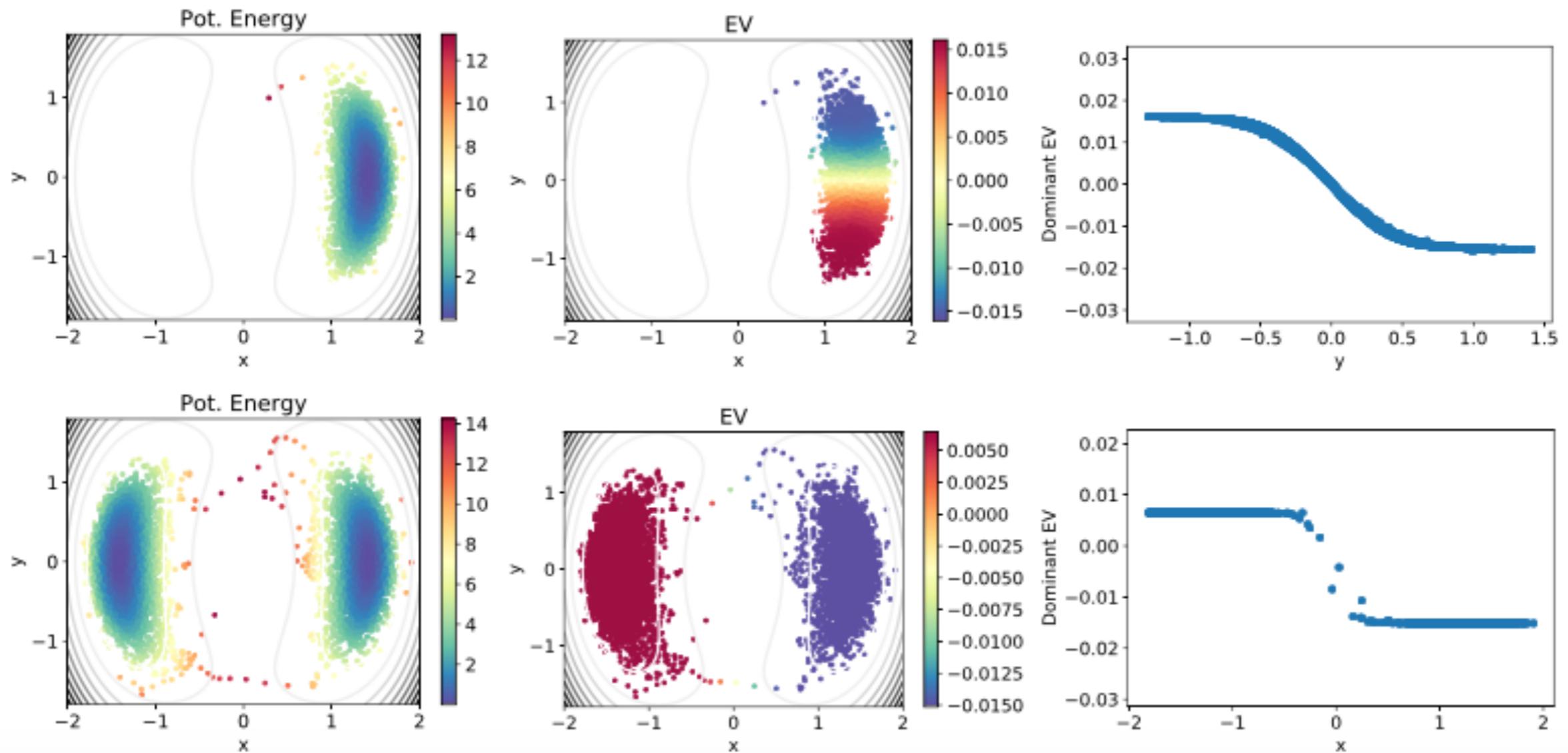
# Eigenfunction approximation in 1D

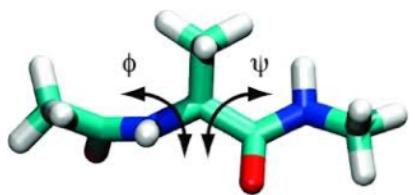
- Quadratic potential  $V(x) = \frac{1}{2}x^2$  on  $\mathbb{R}$
- Sample from  $\mathcal{N}(0, 1)$
- Generator  $\mathcal{L} = -x\partial_x + \partial_x^2$  has Hermite polynomials as eigenfunctions:  $\mathcal{L}H_n = -nH_n, n \geq 0$



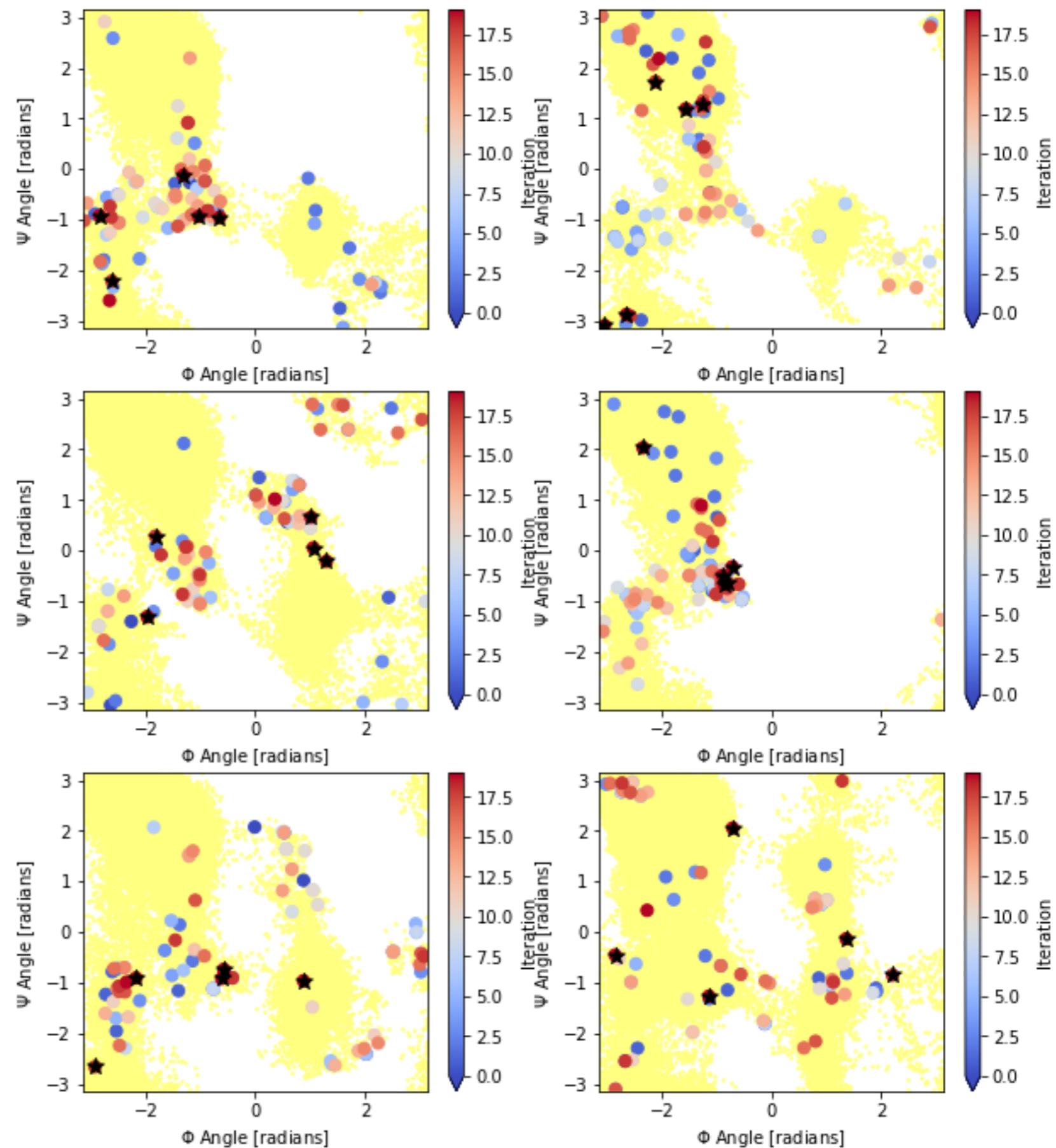
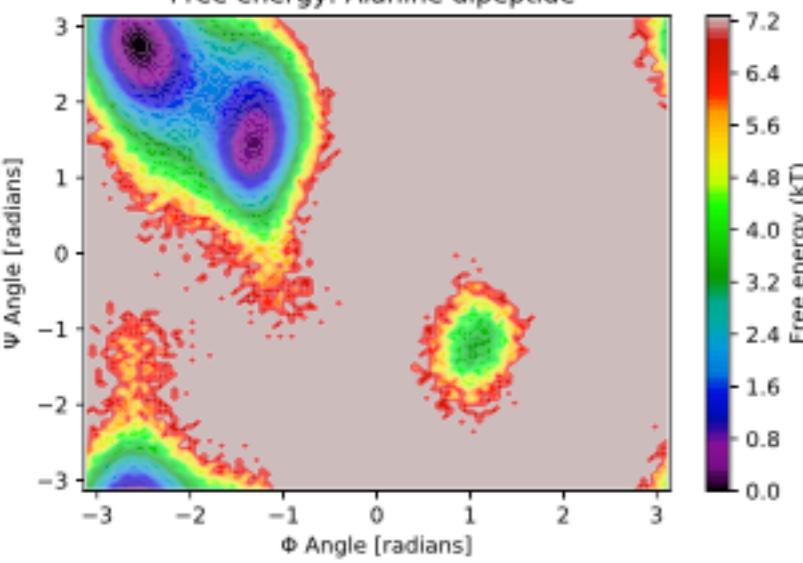
# Geometry-driven sampling (building on work of Clementi, Kevrekides...)

- Enhance sampling by exploration
- Local use of diffusion maps
- Langevin dynamics with force  $\nabla \log \mu = -\nabla V$  has invariant measure  $\mu = e^{-\beta V}$
- Generator  $\mathcal{L} = -\nabla V \cdot \nabla + \frac{1}{\beta} \Delta$
- Eigenfunctions parametrise slowest time scales





Free energy: Alanine dipeptide



# Thermodynamic Analytics Toolkit (TATi)

**Frederik Heber** (Rutherford-Turing Fellow)

**Benedict Leimkuhler**

**Zofia Trstanova** (EPSRC PDRA - Edinburgh)

Vision:

*Exploration tool for loss landscape in ANNs*

*Sampling methods in TensorFlow*

*purposes*

*Develop understanding of NN properties*

*Assist in analysis/redesign of NNs*

# The Edinburgh Environment

## **Maxwell Institute**

Union of mathematicians at Edinburgh and Heriot-Watt Universities. Currently unifying graduate programmes under the banner **Maxwell Institute Graduate School**.

## **International Centre for the Mathematical Sciences**

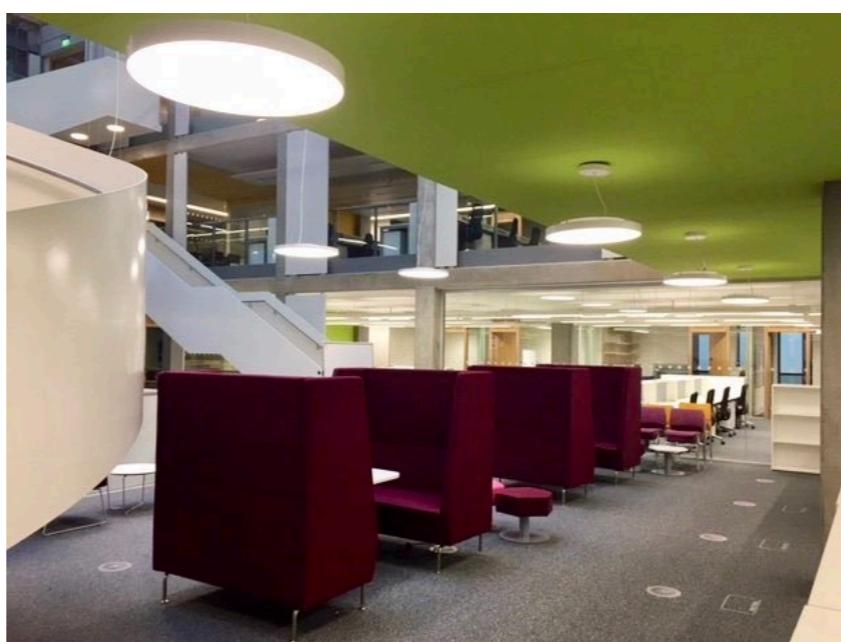
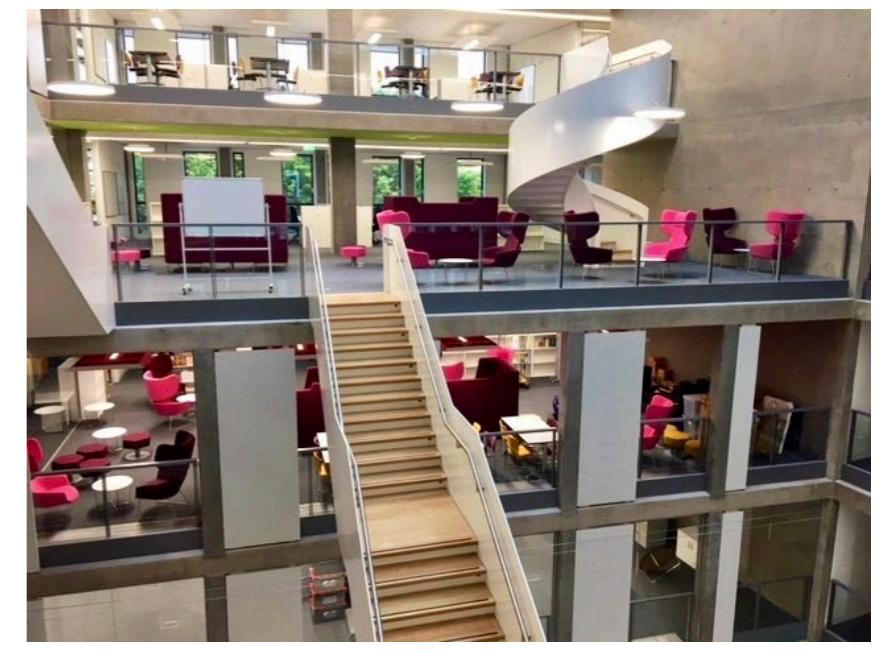
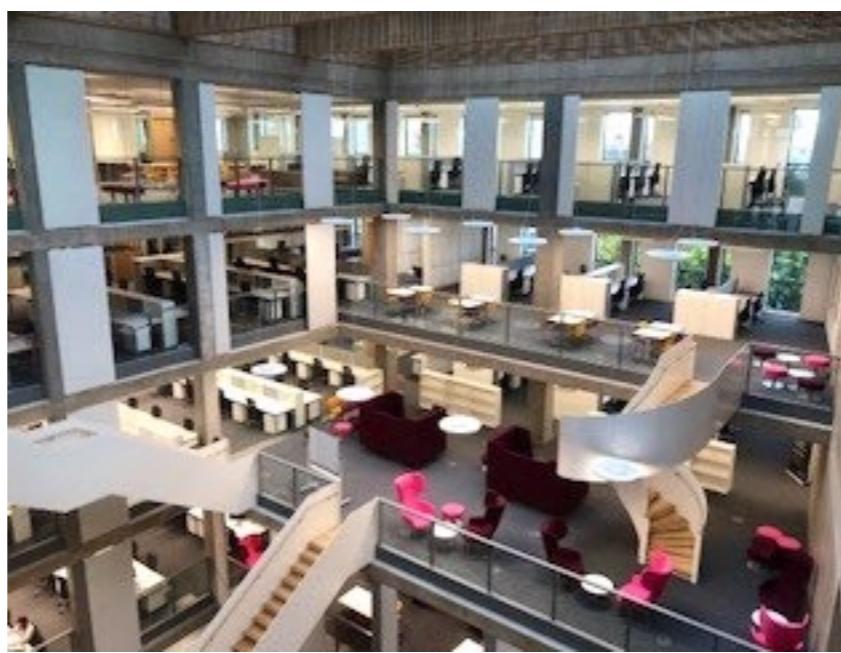
Runs workshops, supports small group research and coordinates knowledge exchange and public events.

## **Centres for Doctoral Training e.g. MIGSAA**

Main mechanism for funding PhD students and coordinating training across the Maxwell Institute

3 proposed in 2018: **MAC-MIGS (Modelling, analysis and computation), GLAMS (Glasgow-Maxwell School in algebraic methods), and Data Science & AI (a joint venture with Informatics)**

# Thomas Bayes Centre



# Why Edinburgh?

amazing skyline:



access to the  
highlands



folk music



fringe  
festival



still part of the EU



**but....**

# The real reason...

Scottish Boletus Harvest (Porcini/Ceps)



**Thank you for listening!**