Letter

Learning minimal representations of stochastic processes with variational autoencoders

Gabriel Fernández-Fernández,[1] Carlo Manzo,[2,3] Maciej Lewenstein,[1,4] Alexandre Dauphin,[1] and Gorka Muñoz-Gil[5,*]

[1]*ICFO–Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain*

[2]*Facultat de Ciències, Tecnologia i Enginyeries, Universitat de Vic–Universitat Central de Catalunya (UVic-UCC), C. de la Laura, 13, 08500 Vic, Spain*

[3]*Bioinformatics and Bioimaging, Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), 08500 Vic, Barcelona, Spain*

[4]*ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain*

[5]*Institute for Theoretical Physics, University of Innsbruck, Technikerstr. 21a, A-6020 Innsbruck, Austria*

(Received 31 July 2023; revised 19 March 2024; accepted 29 May 2024; published 18 July 2024)

Stochastic processes have found numerous applications in science, as they are broadly used to model a variety of natural phenomena. Due to their intrinsic randomness and uncertainty, they are, however, difficult to characterize. Here, we introduce an unsupervised machine learning approach to determine the minimal set of parameters required to effectively describe the dynamics of a stochastic process. Our method builds upon an extended β-variational autoencoder architecture. By means of simulated data sets corresponding to paradigmatic diffusion models, we showcase its effectiveness in extracting the minimal relevant parameters that accurately describe these dynamics. Furthermore, the method enables the generation of new trajectories that faithfully replicate the expected stochastic behavior. Overall, our approach enables the autonomous discovery of unknown parameters describing stochastic processes, hence enhancing our comprehension of complex phenomena across various fields.

DOI: 10.1103/PhysRevE.110.L012102

Introduction. The recent advances in machine learning (ML) have not only impacted everyday life but also the development of science. In physics, the predictive power of ML has been used to get insights from theoretical and experimental physical systems with unprecedented accuracy [1,2]. Indeed, ML can easily extract knowledge from a plethora of data types with no prior information about its source.

It has broadly been argued that, if a machine can make predictions over a given physical process, the properties of the latter must be encoded in the internal representation of the machine [3]. Therefore, beyond its predictive nature, ML can also be helpful for scientific discovery. Several examples in biology [4], quantum matter [5,6], quantum information [7], lattice field theory [8], mathematics [9], or experiment design [10,11] show that deep neural networks (NN), despite being often considered as black boxes, can guide scientists to understand complex phenomena or to design involved experiments.

Various techniques exploit the information encoded in the trained model, e.g., by defining a notion of similarity between the different training examples and the test examples [12,13]. Alternatively, the study of the internal representation of the NN allows mapping the statistics of the training and test examples onto a vectorial space. An example of such embedding is the encoding produced at the bottleneck layer of an autoencoder (AE) [14], NN architectures trained to compress and decompress data to and from a given vectorial space. The abstract representation obtained at this level is very useful for unsupervised applications and machine interpretability. For instance, AE representations, and in particular those of variational autoencoders (VAE) [15,16], are currently crucial for the efficient training of some of the most powerful ML models, from image generation with diffusion models [17] to reinforcement learning [18]. Beyond such downstream applications, focused on enhancing the power of ML methods, VAEs can be used to discover hidden factors of variation in an unsupervised way [19–21]. These models create disentangled representations and isolate the generating factors of input data sets. Achieving this objective is not trivial and great effort is currently invested into improving such representations [22,23]. This application is particularly interesting in physical systems, as the generating factors translate to the relevant physical parameters of the system. Their utility has been extensively proven in a variety of scenarios, in particular in the analysis of dynamical systems [24,25]. These seminal works mainly focused on deterministic systems, raising questions about their applicability to a wider range of real-world scenarios, especially those involving stochastic processes. Similar approaches have also been proposed for stochastic processes (see, e.g., Refs. [26,27]), but they rely on the preprocessing or averaging of the data, such that the input to the ML model is effectively a deterministic signal. Therefore, the efficiency of the representation achieved depends not only on the accuracy of the ML model but also on the statistical relevance of its inputs. Thus, it would be beneficial to develop models that can treat raw stochastic data. This kind of approach has been explored by several ML methods with great success. However,

*Contact author: gorka.munoz-gil@uibk.ac.at

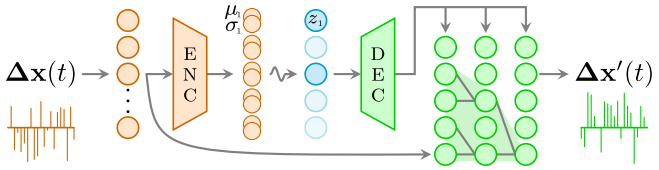

FIG. 1. Interpretable autoregressive β-VAE. Given the displacements $\mathbf{\Delta x}(t)$ of a diffusion trajectory, the encoder (orange) compresses them into an interpretable latent space (blue), in which few neurons (dark blue) represent physical features of the input data while others are noised out (light blue). An autoregressive decoder (green) generates from this latent representation the displacements $\Delta\mathbf{x}'(t)$ of a new trajectory recursively, considering a certain receptive field RF (light green cone).

these methods rely on prior information about the system such as, e.g., the underlying physical model [28,29] or a basis of preselected functions [30,31]. Similarly, symbolic regression approaches can find the governing equations of stochastic processes [32] but necessitate a meaningful set of initial terms to build a proposed expression. This further impedes their application to systems involving parameters that cannot be expressed in closed form.

In this work, we aim at determining whether a machine can extract, in an unsupervised way, the minimal parametric representation of a stochastic process from trajectories without any prior knowledge of the system. Extending previous works on unsupervised learning approaches to diffusion [33,34], we train a β-variational autoencoder (β-VAE) [19] to generate trajectories with the same properties as the ones used as inputs. The architecture presents an information bottleneck constructed to represent conditionally independent factors of variation. Using an adaption of the original β-VAE [35], we successfully train the architecture with various sets of data corresponding to diffusion processes with different characteristics. Our results show that only the minimal necessary properties describing the motion of the particles arise in the bottleneck and can be directly related to the known theories describing these models. Moreover, the training provides a generative model that can produce new trajectories with the same properties as the training data set, thus allowing for an in-depth study of their statistical properties. Besides its fundamental value, this work offers a valuable tool for the study of molecular diffusion from individual trajectories, such as those obtained with single-molecule imaging techniques [36,37], for which extensive ML methods have been developed [38]. In contrast to the latter, rather than predicting known parameters with increasing accuracy, we aim here at solving a more fundamental question: learning the most efficient description of a given stochastic process.

Interpretable generative model. We aim to construct a machine learning (ML) architecture capable of (i) extracting interpretable physical variables from stochastic time series, and (ii) modeling the probability distribution function of the input data. To this end, we consider a β-variational autoencoder (β-VAE) architecture [19], schematically depicted in Fig. 1 (see Supplemental Material (SM) [39] and Ref. [40] for further details). In this architecture, an encoder (depicted in orange in Fig. 1) compresses displacements from an input trajectory $\mathbf{x}$ into a latent space $\mathbf{z}$ (shown in blue), for which each neuron is parameterized via a normal distribution $\mathcal{N}(\mu_{z_i}, \sigma_{z_i})$. Throughout this work, we consider $|\mathbf{z}| = 6$ latent neurons. A sample is then drawn from the latent space and fed into the decoder (depicted in green), which generates a distribution function from which the displacements of new trajectories $\mathbf{x}'$ can be sampled. The training of this architecture is based on a loss function that consists of two terms: a reconstruction loss that compares the model's inputs and outputs, and a second loss term that measures the dissimilarity between the distribution of the latent variables and their prior. For the prior, a standardized normal distribution $\mathcal{N}(0, 1)$ is typically considered. A parameter β is used to control the relative weight of the two loss components and, through an *ad hoc* annealing schedule, can be tuned in such a way that only the minimum number of latent neurons remains informative (i.e., $\sigma_{z_i} \ll 1$). In a physical context, only the pertinent properties governing the process will manifest in the latent space and serve to reproduce the input [24].

Traditional approaches for representation learning typically focus on deterministic decoders and use a reconstruction error to directly compare the input and output of the autoencoder [24,41]. However, when dealing with stochastic signals, any form of compression inevitably leads to significant information loss in the reconstructed trajectory. Therefore, we adopt a distinct approach based on a probabilistic decoder to model the distribution of displacements $p(\mathbf{\Delta x})$ through $p_\theta(\mathbf{\Delta x}|\mathbf{z})$, where θ represents the trainable parameters from which the individual displacements Δx_i of the trajectories are sampled. We thus use the maximum likelihood estimation to compare the resulting p_θ with the samples of the training data set, assumed to be representative of the distribution p. Notably, the stochastic signal that corresponds to the input may exhibit various types of correlations, which play a crucial role in modeling important physical processes. To ensure that these properties are preserved at the autoencoder output, we follow the architecture proposed in Ref. [35] and construct the decoder using an autoregressive (AR) convolutional network known as WaveNet [42]. WaveNet models the output distribution according to the following recursive conditional probability

$$p_\theta(\mathbf{\Delta x}|\mathbf{z}) = \prod_{t=1}^{T} p_\theta(\Delta x_t|\Delta x_{t-1}, \ldots, \Delta x_{t-\mathrm{RF}}; \mathbf{z}), \qquad (1)$$

where T is the length of the trajectory and RF is the receptive field, i.e., the number of past displacements used to predict the forthcoming one (light green cone in Fig. 1).

Extracting physical variables from stochastic data. To test the ability of the architecture to extract relevant physical variables from stochastic data, we train it on four data sets, constructed with three paradigmatic models of diffusion. First, we consider Brownian motion (BM) [43,44], used to describe the stochastic motion of a particle suspended in a fluid. The diffusion of a Brownian particle is characterized by a single parameter, the diffusion coefficient D, hence serving as an initial benchmark for our study. More precisely, BM can be expressed as a Langevin equation of the form $\dot{\mathbf{x}}(t) = \xi(t)$, where $\xi(t)$ is a Gaussian noise with autocorrelation function

$$\langle \xi(t)\xi(t')\rangle = 2D\delta(t - t'). \qquad (2)$$

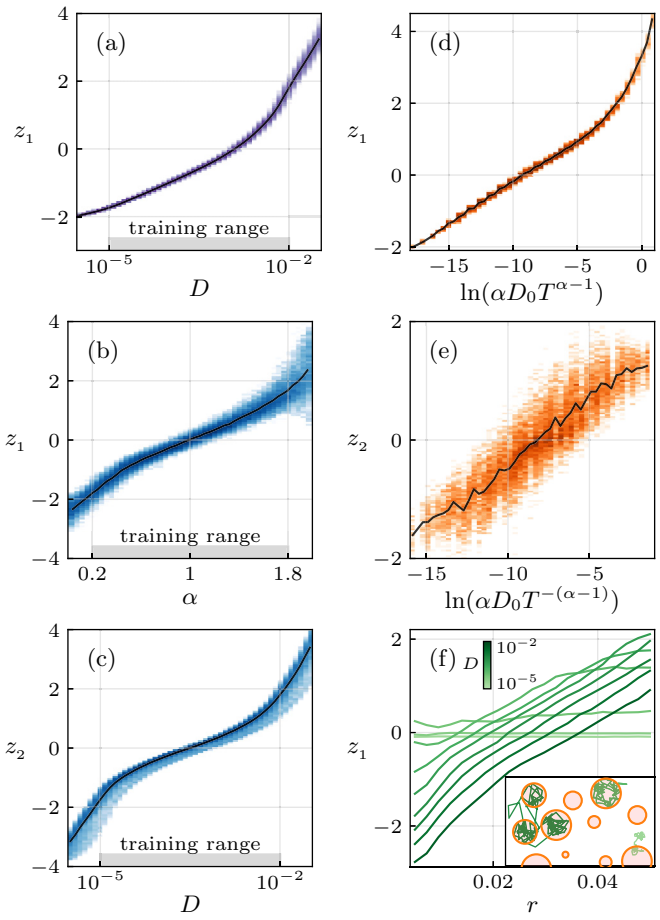

FIG. 2. Interpretation of the latent space. Distribution of latent neuron activations z_i for four data sets: (a) BM; (b), (c) FBM; (d), (e) SBM; and (f) BM with confinement. Only surviving neurons are shown (i.e., $\sigma_{z_i} \ll 1$). For all data sets, the number of surviving neurons agrees with their respective number of degrees of freedom.

To train the autoencoder, we generate a data set of trajectories with $D \in [10^{-5}, 10^{-2}]$ using the *andi_datasets* library [45]. As training proceeds, the model improves its generative capabilities while minimizing the KL divergence of the latent neurons with respect to their prior $\mathcal{N}(0, 1)$. After training, a single neuron of the six available survives, i.e., differs drastically from its purely noisy prior. Figure 2(a) shows a direct relation between such neuron and D, highlighting that the autoencoder has learned that the only information needed by the decoder to generate a new trajectory is its diffusion coefficient. Furthermore, the correlation between z_1 and D extends beyond the training set range (gray shaded area), indicating the model's ability to generalize the representation of diffusive parameters beyond the specified training range.

Next, we consider two extensions of BM, namely fractional Brownian motion (FBM), and scaled Brownian motion (SBM). These are paradigmatic models of anomalous diffusion, i.e., diffusion that deviates from the typical Brownian behavior. These models have found extensive application in describing motion in different biological scenarios at various scales [46–49] and thus constitute a valuable benchmark to demonstrate the method's utility in experimental settings. Both models are characterized by only two parameters: the diffusion coefficient and the anomalous diffusion exponent α. However, the source of anomalous diffusion is different in each model.

FBM can be derived from the Langevin equation and expressed as $\dot{\mathbf{x}}(t) = \xi_{fGn}(t)$, where $\xi_{fGn}(t)$ represents fractional Gaussian noise with the autocorrelation function

$$\langle \xi_{fGn}(t)\xi_{fGn}(t')\rangle = \alpha(\alpha-1)D|t-t'|^{\alpha-2}, \tag{3}$$

where D is here referred to as a generalized diffusion coefficient with dimensions $[l]^2[t]^{-\alpha}$. Importantly, Eq. (3) implies that FBM displacements are correlated. This feature provides an interesting benchmark for the autoregressive properties of the decoder, as we will discuss in the following section. We train an autoencoder with a data set consisting of FBM trajectories with $\alpha \in [0.2, 1.8]$ and $D \in [10^{-5}, 10^{-2}]$. In Figs. 2(b) and 2(c) we show the only two surviving neurons: one (z_1) shows a nearly linear relation with the anomalous diffusion exponent α, whereas the other (z_2) has a monotonic dependence on the $\log(D)$. These results prove the model's ability to only retain minimal information to correctly reproduce FBM trajectories through the probabilistic decoder.

SBM extends Brownian diffusion by considering an aging diffusion coefficient $D(t)$, which is usually considered to scale as $D_\alpha(t) = \alpha D_0 t^{\alpha-1}$, where α is the anomalous diffusion exponent and D_0 is a constant with dimensions $[l]^2[t]^{-\alpha}$. After training, we again observed that only two neurons survived. However, in contrast to earlier cases, these two neurons exhibit a more intricate relationship with α and D_0, as depicted in Figs. 2(d) and 2(e). It must be pointed out that the only constraint imposed by the β-VAE loss in order to obtain these results is that the representation in the latent space is minimal, while still achieving good reconstruction loss. Hence, nothing prevents the network to learn a minimal representation based on combinations of the independent factors (meaningful physical variables in this case) [50,51]. Nonetheless, the number of surviving neurons should never exceed the number of independent factors (or degrees of freedom), a situation that would not correspond to a minimal representation.

In many scenarios, the factors of variation may not be inherently linked to a closed-form equation, unlike the examples previously mentioned. In such cases, phenomenological models are frequently employed to accurately capture the dynamics of the system by defining a few relevant parameters. The methodology we propose is particularly useful for deriving these parameters in situations where no prior information about the physical process exists. As an example, we examine the behavior of a Brownian particle subjected to random confinement, a phenomenon prevalent in numerous biological contexts [52,53]. This scenario involves a particle freely diffusing in a medium containing circular compartments of random sizes, where the particle reflects off the boundaries with a certain probability [Fig. 2(f) inset and SM]. Each particle's behavior can be characterized by two factors: its diffusion coefficient D and, if applicable, the confinement radius r of the compartment it enters. Unlike D and α, identifying r poses a greater challenge. Given the random radius distribution and the boundaries' partial transmittance, it is necessary to isolate the confined segments within the overall trajectory and then compute their confinement radius

[54,55]. Remarkably, the autoencoder is able to overcome such challenge and not only identifies this factor [Fig. 2(f)], but achieves it with remarkable precision. This suggests that the encoder has learned to segment the input trajectory autonomously and derive r without any supervised guidance. For sufficiently large values of D, the surviving neuron correlates with the confinement radius. Expectedly, for smaller D values, the particle's motion within the compartment is too limited, making the accurate determination of r unfeasible.

Generating trajectories from meaningful representations. An essential feature of the presented architecture is its ability to generate trajectories with the same physical properties as the training samples. Moreover, the representational power of the latent space allows one to set the properties of the output trajectories by tuning the value of the latent neurons. Inference is then done directly from the latent space, without any need of the encoder.

As expressed in Eq. (1), the decoder predicts the probability of each displacement Δx_t by means of a conditional probability related to previous displacements and, most importantly, the latent vector $\mathbf{z}$. In practice, by means of the reparameterization trick [15], the decoder outputs the mean μ_t and variance σ_t^2 of a normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$, and we then use the latter to sample each displacement Δx_t. In the case of BM trajectories, the autoencoder correctly learns to set $\mu_t = 0$ and $\sigma_t^2 = 2D\,\forall\, t$, as the displacements of such trajectories are independent and stationary. Hence, the decoder only needs to properly learn the exact transformation from z_1 in Fig. 2(a) to σ_t^2.

Next, we analyze the more complex cases of FBM and SBM. In this sense, a fundamental feature of FBM trajectories is the correlation of displacements, which has a characteristic power-law behavior directly connected to Eq. (3). As commented, the architecture includes an autoregressive decoder to preserve this feature in generated trajectories. In fact, in Fig. 3(a), we show that when generating trajectories for a given α, the power-law correlation is preserved in a range defined by the architecture's receptive field (RF) and then lost, as expected from Eq. (1). Since power-law correlations produce anomalous diffusion in FBM, their loss affects the anomalous diffusion exponent of the generated trajectories, as shown in Fig. 3(b) (see SM [39] for details). While the exponent is correct for $\Delta t < RF$, it rapidly converges to one at longer times. In our experiments, increasing the RF hindered training substantially. A possible solution is to consider a transformer-based decoder [56], where extensive efforts to enlarge context length are currently being pursued [57].

With respect to the SBM data set, the β-VAE must encode into the latent space the time-dependent diffusion coefficient $D_\alpha(t)$ in order to generate trajectories with anomalous diffusion exponent α. We have shown that the latent space obtained for the model trained on SBM trajectories has a complex relationship with the input parameters α and D_0. To simplify the analysis, instead of generating trajectories directly from the latent space as we did with FBM, we feed trajectories with a given ground-truth α and D_0 to the decoder, extract their latent representation $\mathbf{z}$, and use it to generate new trajectories. As shown in Fig. 3(c), the generator is able to correctly reproduce trajectories with the correct exponent for various D_0 and a wide range of α. In Fig. 3(d), we show $D_\alpha(t)$ calculated as the variance of the displacements for different t. The β-VAE perfectly reproduces the expected behavior over all generated times, confirming the generative capabilities of the architecture.

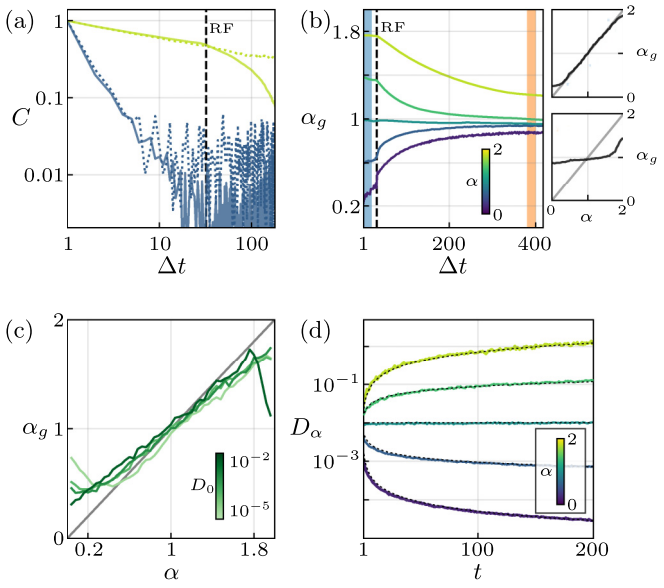

FIG. 3. Statistical properties of generated anomalous diffusion. Top (bottom) row corresponds to the FBM (SBM) data set. (a) Displacements correlations $C = |\langle \Delta x_t \Delta x_{t+\Delta t}\rangle|/\Delta x_0^2$ for the input (dotted) and generated data (solid) with $\alpha = 0.6, 1.8$ (blue and green, respectively). (b) Anomalous exponent α_g of the generated FBM data fitted from the time-averaged mean-squared displacement at different Δt for different input α. Insets show the two-dimensional histograms of the input vs. generated anomalous exponent at the highlighted Δt, before (blue) and after (orange) the receptive field RF. (c) Anomalous exponent α_g of the generated SBM data vs. the input exponent α for various D_0. (d) Evolution of the diffusion coefficient for generated SBM trajectories at various α. Dotted lines show the expected scalings.

Conclusions. In this work, we have explored the application of machine learning (ML) techniques to provide interpretable representations of stochastic processes from time series. We have shown that a method based on a β-variational autoencoder with an autoregressive decoder can retrieve the minimal parametric representation of trajectories corresponding to different processes describing diffusion.

The architecture has been specially developed to account for common features present in stochastic data. First, the output of the network is probabilistic. Due to the stochastic nature of diffusion trajectories, trajectory reconstruction after compression is effectively unfeasible. Hence, instead of reconstructing, as done typically in AE, we aim at generating new trajectories via a parameterized distribution optimized to match the input data distribution. Second, the decoder is autoregressive, a feature introduced in order to model distributions with correlations, as for the case of FBM trajectories.

In contrast to the predominantly employed supervised methods, our study showcases the potential of unsupervised machine learning techniques to uncover the intrinsic structure of stochastic processes and determine the

minimal parametrization required for accurate characterization. As such, it offers a promising avenue for uncovering previously unknown physical degrees of freedom inherent in stochastic physical processes. A significant advantage of this approach is its ability to operate without prior information about the data or the underlying physical process. This makes it particularly well suited for experimental settings. In this sense, a promising avenue is the one related to interventional causal representation learning [50,58], which considers scenarios in which actions (interventions) are applied into a system, changing its properties and facilitating better representations. In this sense, one could leverage such strategies to study the impact of changes in experimental conditions to both understand their influence into the system and better extract the underlying physical model [51,59].

The results of our study also offer practical implications for model simplification and computational efficiency. In phenomenological models, characterized by multiple input parameters, the reduction of the dimensionality of the parameter space can significantly decrease the computational cost associated with the modeling and simulation of stochastic processes, thus enabling more efficient analysis and prediction of their behavior. Thus, the proposed approach offers a promising avenue for advancing the modeling and analysis of stochastic systems, enabling researchers to gain deeper insights into physical processes.

Acknowledgments. G.M.-G. acknowledges funding from the European Union. C.M. acknowledges support through Grant No. RYC-2015-17896 funded by MCIN/AEI/10.13039/501100011033 and "ESF Investing in your future," Grants No. BFU2017-85693-R and No. PID2021-125386NB-I00 funded by MCIN/AEI/10.13039/501100011033/ and "ERDF A way of making Europe," and Grant AGAUR 2017SGR940 funded by the Generalitat de Catalunya. G.F.-F., A.D., and M.L. acknowledge support from: European Research Council AdG NOQIA; MCIN/AEI (PGC2018-0910.13039/501100011033, CEX2019-000910-S/10.13039/501100011033, Plan National FIDEUA PID2019-106901GB-I00, Plan National STAMEENA PID2022-139099NB, I00, project funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR" (PRTR-C17.I1), FPI); QUANTERA MAQS PCI2019-111828-2); QUANTERA DYNAMITE PCI2022-132919, QuantERA II Programme co-funded by European Union's Horizon 2020 program under Grant Agreement No. 101017733); Ministry for Digital Transformation and of Civil Service of the Spanish Government through the QUANTUM ENIA project call - Quantum Spain project, and by the European Union through the Recovery, Transformation and Resilience Plan - NextGenerationEU within the framework of the Digital Spain 2026 Agenda; Fundació Cellex; Fundació Mir-Puig; Generalitat de Catalunya (European Social Fund FEDER and CERCA program, AGAUR Grant No. 2021 SGR 01452, QuantumCAT\U16-011424, co-funded by ERDF Operational Program of Catalonia 2014-2020); Barcelona Supercomputing Center MareNostrum (FI-2023-3-0024); (HORIZON-CL4-2022-QUANTUM-02-SGA PASQuanS2.1, 101113690, EU Horizon 2020 FET-OPEN OPTOlogic, Grant No. 899794), EU Horizon Europe Program (This project has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101080086 NeQST Grant Agreement 101080086 NeQST); ICFO Internal "QuantumGaudi" project; European Union's Horizon 2020 program under the Marie Sklodowska-Curie Grant Agreement No. 847648; "La Caixa" Junior Leaders fellowships, La Caixa" Foundation (ID 100010434): CF/BQ/PR23/11980043. Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union, European Commission, European Climate, Infrastructure and Environment Executive Agency (CINEA), the European Research Executive Agency nor any other granting authority. Neither the European Union nor any granting authority can be held responsible for them.

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[2] A. Dawid, J. Arnold, B. Requena, A. Gresch, M. Płodzień, K. Donatella, K. Nicoli, P. Stornati, R. Koch, M. Büttner, R. Okuła, G. Muñoz-Gil, R. A. Vargas-Hernández, A. Cervera-Lierta, J. Carrasquilla, V. Dunjko, M. Gabrié, P. Huembeli, E. van Nieuwenburg, F. Vicentini *et al.*, arXiv:2204.04198.

[3] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (Leanpub, Victoria, 2018).

[4] C. J. Soelistyo, G. Vallardi, G. Charras, and A. R. Lowe, Learning biophysical determinants of cell fate with deep neural networks, Nat. Mach. Intell. **4**, 636 (2022).

[5] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Correlator convolutional neural networks as an interpretable architecture for image-like quantum matter data, Nat. Commun. **12**, 3905 (2021).

[6] N. Käming, A. Dawid, K. Kottmann, M. Lewenstein, K. Sengstock, A. Dauphin, and C. Weitenberg, Unsupervised machine learning of topological phase transitions from experimental data, Mach. Learn.: Sci. Tech. **2**, 035037 (2021).

[7] A. Pozas-Kerstjens, N. Gisin, and M.-O. Renou, Proofs of network quantum nonlocality in continuous families of distributions, Phys. Rev. Lett. **130**, 090201 (2023).

[8] S. Blücher, L. Kades, J. M. Pawlowski, N. Strodthoff, and J. M. Urban, Towards novel insights in lattice field theory with explainable machine learning, Phys. Rev. D **101**, 094507 (2020).

[9] A. Davies, P. Veličković, L. Buesing, S. Blackwell, D. Zheng, N. Tomašev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhász, M. Lackenby, G. Williamson, D. Hassabis, and P. Kohli, Advancing mathematics by guiding human intuition with AI, Nature (London) **600**, 70 (2021).

[10] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, Active learning

machine learns to create new quantum experiments, Proc. Natl. Acad. Sci. USA **115**, 1221 (2018).
[11] M. Krenn, M. Erhard, and A. Zeilinger, Computer-inspired quantum experiments, Nat. Rev. Phys. **2**, 649 (2020).
[12] P. W. Koh and P. Liang, Understanding black-box predictions via influence functions, in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, 2017), pp. 1885–1894.
[13] A. Dawid, P. Huembeli, M. Tomza, M. Lewenstein, and A. Dauphin, Hessian-based toolbox for reliable and interpretable machine learning in physics, Machine Learning: Science and Technology **3**, 015002 (2022).
[14] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science **313**, 504 (2006).
[15] D. P. Kingma and M. Welling, Auto-encoding variational bayes, in *2nd International Conference on Learning Representations* (ICLR, 2014).
[16] D. J. Rezende, S. Mohamed, and D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, in *Proceedings of the 31st International Conference on Machine Learning* (PMLR, 2014), pp. 1278–1286.
[17] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, High-resolution image synthesis with latent diffusion models, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 10674–10685.
[18] I. Higgins, A. Pal, A. Rusu, L. Matthey, C. Burgess, A. Pritzel, M. Botvinick, C. Blundell, and A. Lerchner, DARLA: Improving zero-shot transfer in reinforcement learning, in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, Sydney, 2017), pp. 1480–1490.
[19] I. Higgins, L. Matthey, A. Pal, C. P. Burgess, X. Glorot, M. M. Botvinick, S. Mohamed, and A. Lerchner, β-VAE: Learning basic visual concepts with a constrained variational framework, in *International Conference on Learning Representations,* (ICLR, 2017), p. 22.
[20] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, Understanding disentangling in β-VAE, in *31st Conference on Neural Information Processing Systems* (NIPS, Long Beach, CA, USA, 2017).
[21] R. T. Q. Chen, X. Li, R. Grosse, and D. Duvenaud, Isolating sources of disentanglement in variational autoencoders, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2018), Vol. 31.
[22] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, Challenging common assumptions in the unsupervised learning of disentangled representations, in *Proceedings of the 36th International Conference on Machine Learning* (PMLR, Long Beach, California, 2019), pp. 4114–4124.
[23] I. Eddahmani, C.-H. Pham, T. Napoléon, I. Badoc, J.-R. Fouefack, and M. El-Bouz, Unsupervised learning of disentangled representation via auto-encoding: A survey, Sensors **23**, 2362 (2023).
[24] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering physical concepts with neural networks, Phys. Rev. Lett. **124**, 010508 (2020).
[25] P. Y. Lu, S. Kim, and M. Soljačić, Extracting interpretable physical parameters from spatiotemporal systems using unsupervised learning, Phys. Rev. X **10**, 031056 (2020).
[26] M. Choi, D. Flam-Shepherd, T. H. Kyaw, and A. Aspuru-Guzik, Learning quantum dynamics with latent neural ordinary differential equations, Phys. Rev. A **105**, 042403 (2022).
[27] J. Im, F. P. J. de Barros, S. Masri, M. Sahimi, and R. M. Ziff, Data-driven discovery of the governing equations for transport in heterogeneous media by symbolic regression and stochastic optimization, Phys. Rev. E **107**, L013301 (2023).
[28] A. Frishman and P. Ronceray, Learning force fields from stochastic trajectories, Phys. Rev. X **10**, 021009 (2020).
[29] D. B. Brückner, P. Ronceray, and C. P. Broedersz, Inferring the dynamics of underdamped stochastic systems, Phys. Rev. Lett. **125**, 058103 (2020).
[30] Y. Huang, Y. Mabrouk, G. Gompper, and B. Sabass, Sparse inference and active learning of stochastic differential equations from data, Sci. Rep. **12**, 21691 (2022).
[31] G.-J. Both, S. Choudhury, P. Sens, and R. Kusters, DeepMoD: Deep learning for model discovery in noisy data, J. Comput. Phys. **428**, 109985 (2021).
[32] G. La Malfa, E. La Malfa, R. Belavkin, P. M. Pardalos, and G. Nicosia, Distilling financial models by symbolic regression, in *Machine Learning, Optimization, and Data Science*, Lecture Notes in Computer Science, edited by G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, G. Jansen, P. M. Pardalos, G. Giuffrida, and R. Umeton (Springer, Cham, 2022), pp. 502–517.
[33] H. Kabbech and I. Smal, Identification of diffusive states in tracking applications using unsupervised deep learning methods, in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (IEEE, Kolkata, 2022), pp. 1–4.
[34] G. Muñoz-Gil, G. Guigó i Corominas, and M. Lewenstein, Unsupervised learning of anomalous diffusion data: An anomaly detection approach, J. Phys. A: Math. Theor. **54**, 504001 (2021).
[35] J. Chorowski, R. J. Weiss, S. Bengio, and A. van den Oord, Unsupervised speech representation learning using WaveNet autoencoders, IEEE/ACM Trans. Audio Speech Lang. Proc. **27**, 2041 (2019).
[36] C. Manzo and M. F. Garcia-Parajo, A review of progress in single particle tracking: From methods to biophysical insights, Rep. Prog. Phys. **78**, 124601 (2015).
[37] E. Barkai, Y. Garini, and R. Metzler, Strange kinetics of single molecules in living cells, Phys. Today **65**(8), 29 (2012).
[38] G. Muñoz-Gil, G. Volpe, M. A. Garcia-March, E. Aghion, A. Argun, C. B. Hong, T. Bland, S. Bo, J. A. Conejero, N. Firbas, Ò. Garibo i Orts, A. Gentili, Z. Huang, J.-H. Jeon, H. Kabbech, Y. Kim, P. Kowalek, D. Krapf, H. Loch-Olszewska, M. A. Lomholt *et al.*, Objective comparison of methods to decode anomalous diffusion, Nat. Commun. **12**, 6253 (2021).
[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.110.L012102 for further details about the machine learning architecture used, along with further explanations on the loss function, data sets, and training, as well as the full latent space and Refs. [60–65].
[40] G. Fernández-Fernández, SPIVAE GitHub repository (2023).
[41] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, Phys. Rev. E **96**, 022140 (2017).
[42] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, arXiv:1609.03499.

[43] A. Einstein, Über die von der molekularkinetischen theorie der Wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen, Annalen der Physik **322**, 549 (1905).

[44] M. von Smoluchowski, Zur kinetischen theorie der Brownschen molekularbewegung und der suspensionen, Annalen der Physik **326**, 756 (1906).

[45] G. Muñoz-Gil, B. Requena, G. Fernández-Fernández, H. Bachimanchi, J. Pineda, and C. Manzo, andi_datasets Python library (2023).

[46] A. Sabri, X. Xu, D. Krapf, and M. Weiss, Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells, Phys. Rev. Lett. **125**, 058101 (2020).

[47] S. C. Lim and S. V. Muniandy, Self-similar Gaussian processes for modeling anomalous diffusion, Phys. Rev. E **66**, 021114 (2002).

[48] G. Muñoz-Gil, C. Romero-Aristizabal, N. Mateos, F. Campelo, L. I. de Llobet Cucalon, M. Beato, M. Lewenstein, M. F. Garcia-Parajo, and J. A. Torreno-Pina, Stochastic particle unbinding modulates growth dynamics and size of transcription factor condensates in living cells, Proc. Natl. Acad. Sci. USA **119**, e2200667119 (2022).

[49] W. Wang, R. Metzler, and A. G. Cherstvy, Anomalous diffusion, aging, and nonergodicity of scaled Brownian motion with fractional Gaussian noise: Overview of related experimental observations and models, Phys. Chem. Chem. Phys. **24**, 18482 (2022).

[50] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, Toward causal representation learning, Proc. IEEE **109**, 612 (2021).

[51] H. P. Nautrup, T. Metger, R. Iten, S. Jerbi, L. M. Trenkwalder, H. Wilming, H. J. Briegel, and R. Renner, Operationally meaningful representations of physical systems in neural networks, Mach. Learn.: Sci. Tech. **3**, 045025 (2022).

[52] M. G. Sugiyama, A. I. Brown, J. Vega-Lugo, J. P. Borges, A. M. Scott, K. Jaqaman, G. D. Fairn, and C. N. Antonescu, Confinement of unliganded EGFR by tetraspanin nanodomains gates EGFR ligand binding and signaling, Nat. Commun. **14**, 2681 (2023).

[53] J. A. Torreno-Pina, B. M. Castro, C. Manzo, S. I. Buschow, A. Cambi, and M. F. Garcia-Parajo, Enhanced receptor–clathrin interactions induced by *N*-glycan–mediated membrane micropatterning, Proc. Natl. Acad. Sci. USA **111**, 11037 (2014).

[54] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, and M. Bathe, Inferring transient particle transport dynamics in live cells, Nat. Methods **12**, 838 (2015).

[55] A. Argun, G. Volpe, and S. Bo, Classification, inference and segmentation of anomalous diffusion with recurrent neural networks, J. Phys. A: Math. Theor. **54**, 294003 (2021).

[56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30.

[57] A. Bulatov, Y. Kuratov, and M. S. Burtsev, Recurrent memory transformer, Adv. Neural Inf. Process Syst. **35**, 11079 (2022).

[58] K. Ahuja, D. Mahajan, Y. Wang, and Y. Bengio, Interventional causal representation learning, in *Proceedings of the 40th International Conference on Machine Learning* (PMLR, Honolulu, Hawaii, USA, 2023), pp. 372–407.

[59] W. F. Whitney, R. Agarwal, K. Cho, and A. Gupta, Dynamics-aware embeddings, in *8th International Conference on Learning Representations* (ICLR, 2020).

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2019), Vol. 32.

[61] J. Howard and S. Gugger, Fastai: A Layered API for deep learning, Information **11**, 108 (2020).

[62] D. P. Kingma and J. Ba, Adam: A Method for stochastic optimization, in *3rd International Conference on Learning Representations*, edited by Y. Bengio and Y. LeCun (ICLR, San Diego, CA, USA, 2015).

[63] L. N. Smith, A disciplined approach to neural network hyper-parameters: Part 1 - Learning rate, batch size, momentum, and weight decay, CoRR abs/1803.09820 (2018).

[64] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, in *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1026–1034.

[65] S. R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio, Generating sentences from a continuous space, in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, edited by S. Riezler and Y. Goldberg (Association for Computational Linguistics, Berlin, Germany, 2016), pp. 10–21.