A Variational Encoder-Decoder Architecture for Charge and Energy Transport in Disordered Materials

Fatemehsadat Mousavi and Mehdi Ansari-Rad*

Faculty of Physics, Shahrood University of Technology, Shahrood, Iran

Sergei D. Baranovskii

Department of Physics, Philipps-Universität Marburg, 35032 Marburg, Germany and

Department für Chemie, Universität zu Köln, 50939 Cologne, Germany

(Dated: January 4, 2025)

Modeling charge and energy transport in energetically and spatially disordered materials has been an active research area, with various theoretical frameworks proposed and applied. In this study, we present a supervised machine learning algorithm that offers valuable insights into transport mechanisms within these systems. By employing a variational encoder-decoder architecture, we train a deep neural network that generates accurate statistics of transport properties. The network's effectiveness is demonstrated through its successful verification of the widely-accepted transport energy concept. The network exhibits an effective dimensionality reduction, encoding raw input data of hopping transport into a minimal yet comprehensive representation necessary for describing the transport mechanism. This attribute offers a promising tool for future deep learning applications aimed at acquiring deeper physical insights into transport mechanisms in disordered materials.

I. INTRODUCTION

Understanding the dynamics of charge and exciton transport processes in disordered semiconductors has been an ongoing subject of intense interest for many years, owing to its importance from both theoretical and practical perspectives [1–4]. In particular, these processes have significant implications for the field of organic electronics, including applications such as organic solar cells (OSCs), organic light-emitting diodes

* ansari.rad@shahroodut.ac.ir

(OLEDs), and other relevant devices. Examples of transport phenomena in disordered systems include exciton transport in heterojunction layers of OSCs toward donor-acceptor interfaces, excited state transport in emission layers of OLEDs, and charge transport in amorphous molecular films and polymer layers.

Within the context of transport phenomena, the primary characteristic of these disordered materials, now firmly established, is the presence of a Gaussian distribution of localized energy states within the system, characterized by a width of σ as

$$g(\varepsilon) = \frac{\rho}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) \tag{1}$$

where $\rho \sim 1\ \mathrm{nm}^{-3}$ is the number density of the localized states in the material [5–7]. Transport processes in these systems occur through successive hopping of charge or energy carriers between the localized states that are randomly distributed in the space. The rate of hopping from an initial site with energy ε_o to a neighboring state ε_d at a distance of r is assumed to follow the Miller-Abrahams (MA) expression (the subscripts o and d denote origin and destination, respectively) [8]. This expression is widely used in the study of charge transport in disordered semiconductors and is given by

$$\nu_\mathrm{od} \propto \exp\left(-2\frac{r}{\alpha}\right) \times \exp\left(-\frac{\varepsilon_\mathrm{d} - \varepsilon_\mathrm{o} + \mid \varepsilon_\mathrm{o} - \varepsilon_\mathrm{d} \mid}{2k_\mathrm{B}T}\right) \tag{2}$$

where $k_\mathrm{B}T$ represents the thermal energy in the system and α is the localization length of the carriers. This expression demonstrates the tunneling nature and thermally activated behavior of the hopping process while satisfying the detailed balance condition as $\nu_\mathrm{od}/\nu_\mathrm{do} = \exp[-(\varepsilon_\mathrm{d} - \varepsilon_\mathrm{o})/k_\mathrm{B}T]$ (but it should be noted that this is not the only physically acceptable form for the hopping rate that satisfies this condition). One important feature of the MA expression is that for downward jumps, where $\varepsilon_\mathrm{d} < \varepsilon_\mathrm{o}$, the hopping rate becomes independent of the energy. This reflects the assumption that downward jumps are always possible, via phonon emission. Consequently, these jumps are completely random, meaning the energy of the destination site is irrelevant.

Although a single hopping event is clearly described by the MA expression, the parameters that quantify the transport properties of the system, such as the diffusion coefficient D or the mobility μ, are typically determined by a large number of successive hopping events over an extended period t. For example, for the

mobility we have

$$\mu \simeq \frac{e}{kT}\frac{\langle \Delta \mathbf{R}^2 \rangle}{\langle \tau \rangle} \tag{3}$$

in which $\langle . \rangle$ represents the statistical averaging and $\Delta \mathbf{R} = \sum_{\mathrm{s}} \mathbf{r}_{\mathrm{s}}$ is the net displacement and $\tau = \sum_{\mathrm{s}} t_{\mathrm{s}}$ is the total time resulting from the successive steps s. Consequently, to obtain mathematical expressions for the macroscopic parameters like μ, one needs to devise models and approaches that provide a framework for describing the overall dynamics of the carriers. Nearest-neighbor and variable-range hopping mechanisms [9] are the well-established theoretical grounds for describing transport dynamics in disordered systems. Various tools and approaches have been developed to investigate these mechanisms, for systems approaching or at thermal equilibrium, including kinetic Monte-Carlo (kMC) simulations [10], methods based on solving a master equation [11], percolation theory [12], and the transport energy concept [13].

The transport phenomena in inorganic semiconductors with a broad distribution of localized states in the system are successfully described by the so-called multiple-trapping model [14–16]. In this model, instead of direct transitions of carriers between localized states, hopping events occur mainly as jumps from localized states to the mobility edge in the conduction band. The destination (the mobility edge), thus, remains the same for all jumps, irrespective of the energy of the origin site, thereby making the subsequent theoretical considerations tractable. Interestingly, despite the lack of a mobility edge in disordered organic materials, it was shown first for the exponential [17] and then for the Gaussian density of states (DOS) [13] that for hopping transport based on the MA expression in these systems, there is a particular energy level that plays the same role as the mobility edge in the multiple-trapping model. Mathematically, this energy level, the transport energy level $\varepsilon_{\mathrm{tr}}$, is defined as the energy that provides the fastest *upward* hop from a given origin site with energy ε_{o} and is determined by maximizing the upward hopping rate, as [13, 18]

$$\left(\frac{\partial \nu_{\mathrm{od}}}{\partial \varepsilon_{\mathrm{d}}}\right) = 0 \tag{4}$$

This results in an energy level that is independent of the origin state as $\varepsilon_{\mathrm{tr}} = \sigma x$, where, for the case of sufficiently low carrier densities, $x = x(\rho^{\frac{1}{3}}\alpha\sigma/kT)$ is the solution to the following equation:

$$\exp(x^2/2)\left[\int_{-\infty}^{x/\sqrt{2}} \exp(-y^2)\,\mathrm{d}y\right]^{4/3} = \frac{k_{\mathrm{B}}T}{\sigma}\left[9\sqrt{2\pi}\rho\alpha^3\right]^{-1/3} \tag{5}$$

According to the picture provided in this framework, upward hops occur in the vicinity of $\varepsilon_{\rm tr}$, while for $\varepsilon_{\rm o} > \varepsilon_{\rm tr}$, carriers undergo successive downward hops until they reach $\varepsilon_{\rm tr}$. Although both upward and downward hops contribute equally to $\langle \Delta \mathbf{R}^2 \rangle$, the most effective hopping events (those crucial in determining the transport properties of the system) are primarily the upward hops due to their dominant contribution in determining $\langle \tau \rangle$ in Eq.(3).

One well-known observation for the temperature dependence of mobility in organic materials at low carrier density, which can be reproduced using the tools provided by the concept of transport energy, is a dependency expressed as

$$\mu(T) \propto \left[-(C\sigma/k_{\rm B}T)^2\right] \tag{6}$$

with the factor $C \simeq 0.6 - 0.7$ being slightly dependent on the parameter $\rho\alpha^3$ [18–22]. Despite the successful utilization of the transport energy concept for describing transport phenomena, its existence and exact position have been difficult to confirm in direct kMC simulations. The difficulty mainly arises from the fact that oscillatory jumps (which have been discussed as being irrelevant in determining the long-range transport properties) dominate the distribution of the destination energy of the hopping events [23, 24]. As a result, by monitoring the destination sites in the kMC simulations, one finds the most frequently visited energy level as $\varepsilon_{\rm fv} \approx -\sigma^2/2k_{\rm B}T$ [25]. In contrast, Eq.(5) provides an energy level that also depends on $\rho\alpha^3$, with its position being significantly above $\varepsilon_{\rm fv}$. Therefore, even the existence of the transport energy level cannot be easily inferred from the kMC simulations. As the only direct evidence, Oelerich et al. validated the concept of transport energy using an interesting approach in which changes in carrier mobility were monitored by modifying the DOS [25]. Their results demonstrated that the absence of a specific energy range within the DOS considerably reduces mobility, implying the existence of an energy level that determines the transport properties. The position of this energy level was found to be close to, but not exactly the same as, the one predicted by Eq.(5).

In recent years, machine learning algorithms have established a new paradigm in both scientific research and technological innovation [26, 27]. Despite critiques regarding the non-interpretable nature of these algorithms' predictions, sometimes referred to as the *black box* issue, substantial advancements have been made in developing interpretable machine learning algorithms [28–30]. A notable example is probabilistic

encoder-decoder networks, a class of neural network architectures capable of extracting relevant physical parameters, or features, from raw input data [31]. Like traditional encoder-decoders, these architectures enable the encoding of input data into a substantially reduced number of features; a process called dimensionality reduction. However, probabilistic encoder-decoder networks possess an additional capability, as they learn the probability distribution of features rather than a deterministic representation. Consequently, these networks are categorized as generative models, as the learned distributions can be employed to generate probabilistic outputs without requiring any input.

While neural networks and other supervised data-driven methods are effective for making predictions, they require substantial amounts of labeled data to train successfully. As a result, the process of generating data is crucial, not only for choosing the appropriate algorithm but also for ensuring effective training. This process ultimately influences the accuracy and dependability of the predictions. A growing number of recent studies have leveraged machine learning to investigate transport properties and other characteristics of disordered systems [32–35].These studies often rely on experimental data for phenomenological modeling and prediction, or data derived from quantum mechanical computations at molecular and atomic scales [36–38]. In contrast to these approaches, our work utilizes the hopping events generated by kMC simulations as labeled data, which is fed into the network, as will be discussed in more detail later. KMC simulations have long been a key tool for modeling experimental observations and assessing the validity of theoretical models of transport processes [39, 40]. In contrast to phenomenological approaches, kMC modeling provides valuable insights into the inner workings and underlying principles of the system. Compared to quantum-level calculations, kMC has the advantage of exploring a wider range of time and length scales, enabling a comprehensive investigation of long-range transport properties. In this work, we demonstrate that by feeding raw hopping event data, sampled from systems with varying energy disorder σ and localization lengths α, into a neural network with an architecture similar to probabilistic encoder-decoders, it can predict the statistics of the transport process in these systems. Additionally, the network predicts the presence of energy levels in the systems that act as preferential destinations for upward hops. The predicted positions of these energy levels are found to be in agreement with the energy range obtained from the transport energy concept.

In the following, we first present the neural network architecture designed to predict the statistics of carriers' hopping events. Section II A introduces the architecture, while Section II B discusses its results for

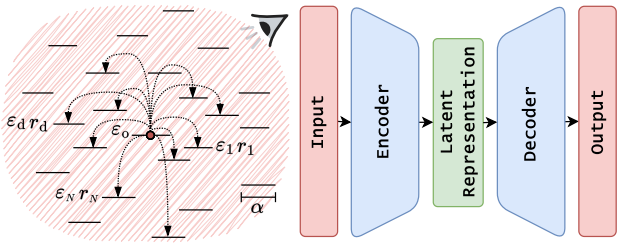

FIG. 1. Encoder-decoder neural network architecture for learning hopping transport statistics in disordered materials. The left panel depicts a schematic illustration of a hopping event scene. The available information is fed into the network as input, consisting of the origin energy, possible destination energies, corresponding hopping distances, and the localization length, represented schematically by the extent of the sites. During the training process, the network learns to encode the input data into a minimal representation in the latent space. Using this compressed data, the decoder generates the desired outputs, i.e., the destination energy and hopping distance.

the distribution of energies and hopping distances. Section II C examines the interpretability of our network, offering insights into its inner workings. Finally, concluding remarks and suggestions for future work are provided in Section III.

II. VARIATIONAL ENCODER-DECODER FOR HOOPING TRANSPORT

A. Neural Network Architecture

We design and implement a neural network architecture to capture the statistical characteristics of hopping events. The training process is formulated as a supervised machine learning problem, where the input to the network incorporates the information available in a *hopping scene*, as illustrated in Fig.1. In the context of machine learning terminology, a single hopping event is an individual instance in the dataset, often referred to as a sample. Each input sample consists of the origin energy ε_o, potential destination energies in the origin neighborhood $\boldsymbol{\varepsilon} = [\varepsilon_1, ..., \varepsilon_d, ..., \varepsilon_N]$, and their corresponding hopping distances $\boldsymbol{r} = [r_1, ..., r_d, ..., r_N]$. The true destination site and hopping distance, denoted by ε_d and r_d, are derived from a simulated Monte-Carlo step for the scene and serve as the target data for the training process of the network. The network predictions, ε'_d and r'_d, are compared to the target values through a cost function that is subsequently minimized, as will be further discussed in detail. The training data supplied to the network comprises

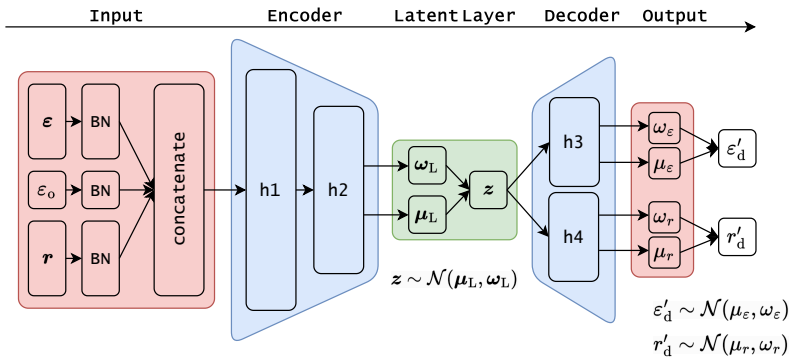

FIG. 2. Details of the β-variational encoder-decoder network (β-VED). The input data, comprising the origin energy ε_o, neighboring site energies $\boldsymbol{\varepsilon}$, and corresponding hopping distances $\boldsymbol{r}$, are connected to batchnormalization layers (BN) and concatenated to feed the encoder. The encoder consists of two hidden layers h1 and h2. The outputs of the encoder are the means $\boldsymbol{\mu}_\mathrm{L}$ and variances $\boldsymbol{\omega}_\mathrm{L}$, each of dimension ℓ. The latent representation, $\boldsymbol{z}$, of the same dimension, is sampled as random Gaussian variables and then provided to the decoder. The decoder comprises two parallel hidden layers h3 and h4, which ultimately outputs means μ_ε and μ_r, and variances ω_ε and ω_r. These outputs are then used to sample the destination energy and hopping distance, ε'_d and r'_d.

individual hopping events (not a sequence of consecutive hops) derived from scenes with varying σ and α values, generated randomly within the relevant ranges (see the end of this subsection for more details).

Within the context of the hopping transport model, an essential question we aim to address is the number of independent parameters required to predict the destination site. According to the transport energy concept, the destination site, essentially represented by the transport energy level, as suggested by Eq. 5, is dependent on the quantities σ, $\sigma/k_\mathrm{B}T$, and $\rho\alpha^3$, while being independent of the origin energy ε_o. Consequently, for problems with fixed temperature and density, understanding two parameters, namely σ and α, would be sufficient to answer this question within the framework of the transport energy concept. In the neural network framework, we aim to examine this problem by exploring the minimal number of neurons necessary to encode the input data into them while maintaining a sufficiently small training cost function. These neurons are schematically depicted in Fig.1 as a single layer, referred to as the latent representation or the coding layer. The number of neurons in the latent layer (i.e. the dimension of the compressed data) will be denoted by ℓ in the following discussion.

Although we formulate our problem as a supervised learning task, it is worth noting that applying neural

networks for dimensionality reduction often involves an unsupervised learning approach. In this context, the target and input data are the same, leading to the so-called autoencoder networks [41, 42]. These networks are trained to reproduce the input data in the output, after a data compression, by encoding them in a latent representation. An important extension of these networks is the probabilistic autoencoders, known as variational autoencoders (VAEs), which were introduced in the work of Kingma and Welling [31]. In a VAE, the loss function can be interpreted as consisting of two components. The first is the latent loss, which is expressed as the Kullback-Leibler (KL) divergence between the distribution of the output of the latent layer, learned by the network, and a target distribution, usually assumed to be a multivariate normal distribution. The KL divergence is, in fact, a measure of the dissimilarity between these two distributions. The KL divergence for the VAE network can be expressed as

$$\mathcal{L}_{\mathrm{L}} = -\frac{1}{2}\sum_{i=1}^{\ell}\left[1 + \log(\omega_i) - \omega_i - \mu_i^2\right] \tag{7}$$

where the means $\{\mu_1, \ldots, \mu_\ell\} = \boldsymbol{\mu}_{\mathrm{L}}$ and variances $\{\omega_1, \ldots, \omega_\ell\} = \boldsymbol{\omega}_{\mathrm{L}}$ are the outputs of the encoder [42]. A network similar to VAE, which we design and utilize in our work, is illustrated in Fig.2. As depicted in the figure, the encoder receives the input data (after being normalized through the so-called batch normalization layers for technical reasons [43, 44]) and generates the encoded representation, or codings, $\boldsymbol{\mu}_{\mathrm{L}}$ and $\boldsymbol{\omega}_{\mathrm{L}}$. Using these codings, the latent representations $\boldsymbol{z} = \{z_1, \ldots, z_\ell\}$ are sampled as Gaussian random variables. In summary, the loss function $\mathcal{L}_{\mathrm{L}}$ in Eq.7 is introduced in the total loss to push the network to learn the *distribution* of the data in the latent layer, rather than finding a deterministic representation of them.

The second component of the loss function is the reconstruction loss, which its minimization drives the decoder to generate outputs similar to the target. As mentioned earlier and illustrated in Fig.2, our training process is formulated as a supervised problem, where the targets (the true destination energy and hopping distance, ε_{d} and r_{d}) differ from the inputs. Therefore, instead of using the term VAE, we refer to our architecture as a variational encoder-decoder (VED). The reconstruction loss is typically expressed as the mean squared error between the target and the output. In our case, for each sample, this can be expressed as contributions to the loss in the form of $(\varepsilon_{\mathrm{d}} - \varepsilon'_{\mathrm{d}})^2$ and $(r_{\mathrm{d}} - r'_{\mathrm{d}})^2$. However, in order to capture the uncertainty (probabilistic nature) of the hopping events, our decoder is designed to output a pair (a mean and a variance) for each of the destination energy and hopping distance, as μ_ε, ω_ε, and μ_r, ω_r, as shown in

Fig.2. The corresponding loss is expressed as

$$\mathcal{L}_R = \frac{1}{2}\left[\log(2\pi\omega_\varepsilon) + \frac{(\varepsilon_\mathrm{d} - \mu_\varepsilon)^2}{\omega_\varepsilon} + \log(2\pi\omega_r) + \frac{(r_\mathrm{d} - \mu_r)^2}{\omega_r}\right] \tag{8}$$

This mathematical form (which essentially represents a negative log-likelihood) ensures that smaller variances yield lower loss values as the predicted mean values approach the targets [45]. Consequently, minimizing the loss drives the network to output values that then can be used to generate ε'_d and r'_d (as depicted in Fig.2). We ultimately compare the distribution of ε'_d and r'_d, predicted for a test dataset, with those of ε_d and r_d to evaluate the performance of our network.

As an important extension, Higgins et al. [46] demonstrated that using a total loss as

$$\mathcal{L} = \mathcal{L}_\mathrm{R} + \beta\mathcal{L}_\mathrm{L} \tag{9}$$

can benefit from both an unentangled representation of the data in the latent space and good reconstruction accuracy [47, 48]. Here, β serves as a hyperparameter (a regularization parameter) to balance the strength of the latent and reconstruction losses. Our network in Fig.2 employs the loss function defined in Eq.9 and, consequently, we refer to the corresponding network as a β-variational encoder-decoder, or β-VED. The β-VED architecture comprises 250, 80, $\ell = 5$, 2×80, and 4 neurons for the encoder, latent space, decoder, and output layers, respectively. The ReLU activation function is used for the hidden layers of the network. We employ the Nadam optimization technique (with a learning rate of 0.001) to minimize the loss function and update the trainable parameters of the network. For the value of β in the total loss, Eq.9, we use the value 0.1. The impact of different choices for this parameter will be discussed in section II C. The dataset is fed to our β-VED network in batches of size 256 and the training process is repeated until the different losses reach saturation, which occurs after approximately 3000 epochs.

The training process employs a dataset consisting of 60000 samples. Each sample represents an independent single hopping event generated for systems with randomly chosen $\sigma/k_\mathrm{B}T \in [1, 7]$ and $\alpha\rho^{1/3} \in [0.1, 0.6]$. As a result, the training dataset is *mixed* and not limited to a fixed σ or α. In a typical kMC simulation, it is worth noting that both σ and α are fixed. With these parameters set, a sequence of consecutive hops, i.e., a trajectory, among neighboring sites is simulated. This process begins by placing a particle at a random site within a simulation box. The positions of the sites inside the simulation box are random and their

energies are assigned using a Gaussian distribution of width σ. A hopping event is carried out by selecting a site from the neighboring sites with probabilities weighted by the corresponding hopping rates, given by Eq.(2). In contrast to performing a full kMC simulation, we generate a sample in the dataset by first randomly selecting a pair of σ and α from the specified ranges. Then, a single hopping event is simulated using the kMC technique described above. The initial (origin) energy, neighboring sites positions, and their energies are used as input (which we referred to earlier as the information available in the hopping scene), while the hopping distance and the destination energy, obtained by performing a kMC step, are used as the target for the network. This straightforward approach enables the network to be trained on individual hopping events snapshotted from various systems, significantly simplifying data generation compared to a traditional full kMC simulation. Once the network is trained, there is no longer a need for kMC simulations. Instead, by providing a hopping scene as input, the network can directly produce a probabilistic energy and hopping distance. Practically, for each sample, generated for a specific σ and α, the origin and neighboring energies, ε_{o} and $\boldsymbol{\varepsilon}$, are drawn from a Gaussian distribution with zero mean and width σ. With the origin site positioned at $(0, 0, 0)$, the neighboring sites are distributed randomly in space with a density of ρ. The first 64 neighbors of the origin site are considered potential destination sites. Consequently, each sample has a dimension (number of features) of $64 + 64 + 1 = 129$; see the input layer in Fig.2. To account for the localization length information in the input data, the scaled distances $\boldsymbol{r}/\alpha$ are fed to the network instead of $\boldsymbol{r}$. We use $k_{\mathrm{B}}T = 0.025$ eV and $\rho = 1$ nm^{-3} throughout the paper.

B. Results and Discussion

Following the training phase, the β-VED network is employed to generate predictions on a test dataset. Fig.3 illustrates the predictions made by the trained network for a test dataset comprising hopping scenes with mixed characteristics similar to those in the training data. It is important to note that these results pertain to instances where the origin energy is sampled from Gaussian distributions with a mean of zero. The figure demonstrates a good agreement between the network's predicted statistics and those of the test dataset for both destination energies and hopping distances. Consequently, it can be concluded that the trained network accurately predicts true distributions for the hopping parameters when the β-VED is

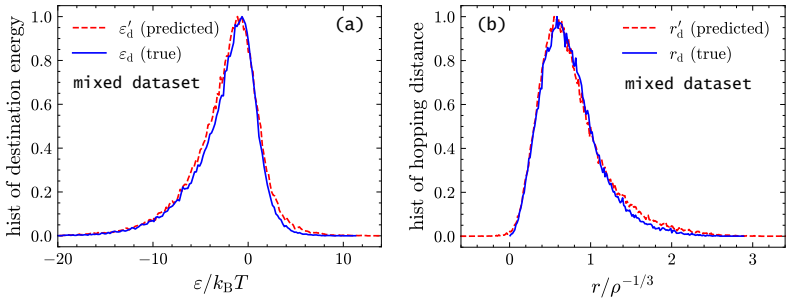

FIG. 3. Histograms illustrating (a): the distribution of destination energy and (b): hopping distance, obtained by processing a test dataset through the network. In each panel, the solid line represents the target (true) values, while the dashed line represents the predicted values outputted by the network. Like the training dataset, the test dataset is mixed, consisting of samples with different disorders and localization lengths, as $\sigma/k_BT \in [1, 7]$ and $\alpha\rho^{1/3} \in [0.1, 0.6]$.

provided with the data sharing statistical similarities with the data used during network training. All results reported in this subsection are derived from this trained network.

It is worth emphasizing that while one could have trained the network using samples with the same σ and α, or, more generally, with the same $\mathrm{A} = \sigma/k_BT$ and $\mathrm{B} = \alpha\rho^{1/3}$, this network would have been applicable and relevant only for a system with that specific (A, B) combination. Consequently, a new network would need to be trained for each set of (A, B). But, fortunately, it was found that a single network could be utilized for any (A, B) provided that a mixed dataset was employed during the training process. This approach ensures that a single network can learn about different (A, B) simultaneously. In other words, a single network can handle various (A, B) combinations, providing a versatile solution instead of requiring a unique network for each individual (A, B) pair. Fig.3 demonstrates that this network has been trained effectively, as it generates predictions comparable to the true values. In the following subsections, we apply this network to pure datasets, where all samples in a dataset share the same (A, B) values. Thus, the prediction of the network for transport energy in a specific system can be investigated.

To evaluate the network's ability to generalize to previously unseen statistics, we will now assess its generalization performance. We will consider three aspects in the following. First, we provide the network with a *pure* dataset, where all samples are generated using the same α and σ (i.e., hopping scenes within a specific system), as discussed above. This differs from the mixed dataset used during the training phase. Second, we examine the network's convergence behavior by iteratively using its predictions. In this process,

the predicted destination energy is repeatedly employed as the new origin energy. The goal is to determine if the network output converges to a specific energy, potentially the same $\varepsilon_{\rm tr}$. Third, we explore whether both downward and upward jumps occur towards the specific energy identified in the convergence analysis.

Pure dataset

Fig.4(a) presents the network's prediction for the destination energy when the network is provided with a pure dataset. The dataset consists of samples generated using a fixed disorder, $k_{\rm B}T/\sigma = 0.3$, and a localization length, $\alpha\rho^{1/3} = 0.3$. As can be inferred from the figure, there is a good agreement between the network's prediction and the true histogram for the destination energy. This implies that β-VED, despite being trained on a mixed dataset, can accurately predict the correct histogram for destination energies when provided with a pure dataset. (Similar agreements were observed for various energy disorder parameters σ and localization lengths α, with the results for hopping distances also showing comparable consistency.) Unlike the mixed dataset, the distribution of the destination energy for a pure dataset exhibits symmetry and can be effectively approximated using a Gaussian function. Two vertical lines in Fig.4(a) emphasize the shift in the peak position of the energy, denoted by $\varepsilon'_{\rm pk}$, after a single pass through the network, from the initial distribution with a zero mean to the destination distribution peaked at a deeper energy on the axis. (The distribution of the hopping distance, on the other hand, remains stretched. We, therefore, report the mean values, $r'_{\rm mean}$, instead of the peak.) It should finally be pointed out that both upward and downward transitions are predicted by β-VED, although the distribution of the destination energy lies at a deeper level compared to the initial input, therefore describing a transition downwards on average.

Convergence

To examine the existence of a transport energy level, the primary focus of this paper, we employ the distribution of the destination energy as the new input for the network. The process is then iterated, as illustrated schematically in Fig.4(b), to determine whether the network converges to a stationary distribution. Under these conditions, the network should operate as an identity network, replicating the input distribution with-

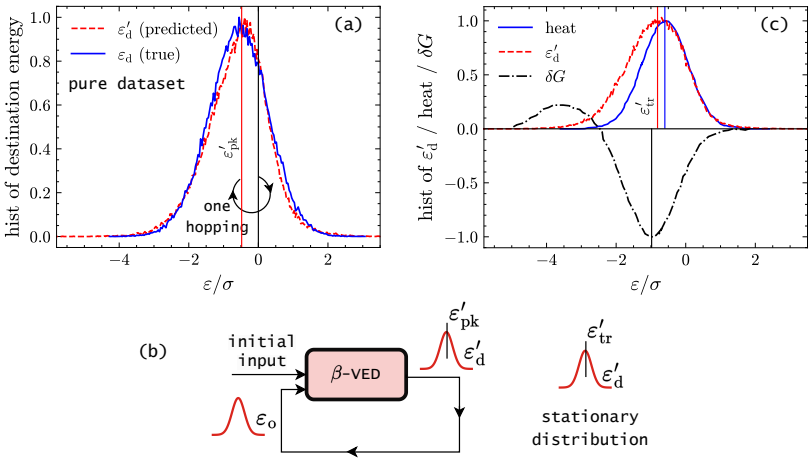

FIG. 4. Histograms illustrating the distribution of destination energy ε'_d obtained for a pure test dataset with $k_BT/\sigma = 0.3$ and $\alpha\rho^{1/3} = 0.3$, for (a): after a single pass through the network, i.e. after one hopping, and (c): after iteratively using the output of the network as the new input and reaching to a stationary distribution. The iterative process is depicted schematically in (b). The vertical line in part (a) highlights the position of the mean energy of the initial input fed into the network. The peak position of the destination energy distribution, ε'_p, and the peak of the stationary distribution, ε'_{tr}, are also marked as vertical lines in the panels. The results in part (c), for the heat distribution and the change in conductance δG (together with the vertical lines indicating the position of the transport energy estimated by each distribution), are adopted from reference [49]. Further details can be found in the main text.

out further shifts in the output. Our findings demonstrated that β-VED generates a stationary distribution in the output for both the destination energy and hopping distance following a few iterations (i.e. hopping events). Fig.4(c) presents the stationary distribution for the system with $k_BT/\sigma = 0.3$ and $\alpha\rho^{1/3} = 0.3$, achieved after approximately 6 iterations, or hoppings. In line with the concept of the transport energy level, we define the peak energy position of this distribution, denoted by ε'_{tr}, as the transport energy level predicted by β-VED. To compare the energy position of ε'_{tr} with those reported in previous studies, the results obtained by Nenashev et al. are also presented in Fig.4(c) [49]. By considering the system as a network of resistors [8], they calculated the distribution of the heat produced in the system, $h(\varepsilon)$, and the change in the conductance, δG, of the system by cutting out a small interval of the energy within the DOS $g(\varepsilon)$. (The result for the change in conductance has been in complete agreement with the direct kMC simulation in a modified DOS in Ref.[25].) As the guiding lines in Fig.4(c) demonstrate, there is good agreement between

$\varepsilon'_{\rm tr}$ predicted by the β-VED and those obtained using the resistor network approach.

Relaxation and activation

In the context of the transport energy level, which is defined as the energy level where hopping events above it are primarily downward in energy and those below it are directed toward it, we analyze the predictions of β-VED for the relaxation and activation process of carriers in the system. We specifically focus on the hopping destination energies when the initial origin energies are chosen to be significantly above or below the center of the DOS, 3σ and $\varepsilon_\infty = -\sigma^2/k_{\rm B}T$ (the equilibrium energy in a Gaussian DOS), respectively. (It should be noted that such inputs were not seen by the network during the training phase.) Fig.5 presents the results for the peak positions $\varepsilon'_{\rm pk}$ and the mean hopping distances $r'_{\rm mean}$ when the output of the network is iteratively used as its input, i.e., the same procedure depicted in Fig.4(b).

As can be seen in Fig.5(a) and (b), for different disorder and localization lengths, β-VED predicts results consistent with the picture provided by the concept of transport energy, as i) transitions from high energy states are downwards and those of low energy states are upwards, and ii) after a few hopping events, stationary conditions are established, and both upwards and downwards hops reach the same energy, $\varepsilon'_{\rm tr}$. Results in this figure interestingly reveal how a higher disorder and localization length hinder the relaxation and activation of carriers in a disordered system. The transition toward the transport energy level indeed happens gradually, within some successive hopping steps, and not through a single hopping event. In addition, although the upwards and downwards branches seem to be symmetric in comparison to each other, as seen in the figure, it should be noted that the corresponding hoppings take place at different time scales. In Fig. 5(c), we observe that under stationary conditions, the mean hopping distance, $r'_{\rm mean}$, exhibits an increasing trend with the amount of disorder. This can be attributed to the fact that, in order to minimize the energy barrier for hopping transitions, carriers tend to hop, on average, to longer distances where a destination state with a suitable energy can be found. This is the essence of the so-called variable-range hopping transport mechanism. Finally, regarding the results in Fig.5(d), although the presented results might at first glance appear counterintuitive, note that the mean distances are displayed in relative terms as $r'_{\rm mean}/\alpha$. In fact, the higher localization lengths correspond to larger $r'_{\rm mean}$ values.

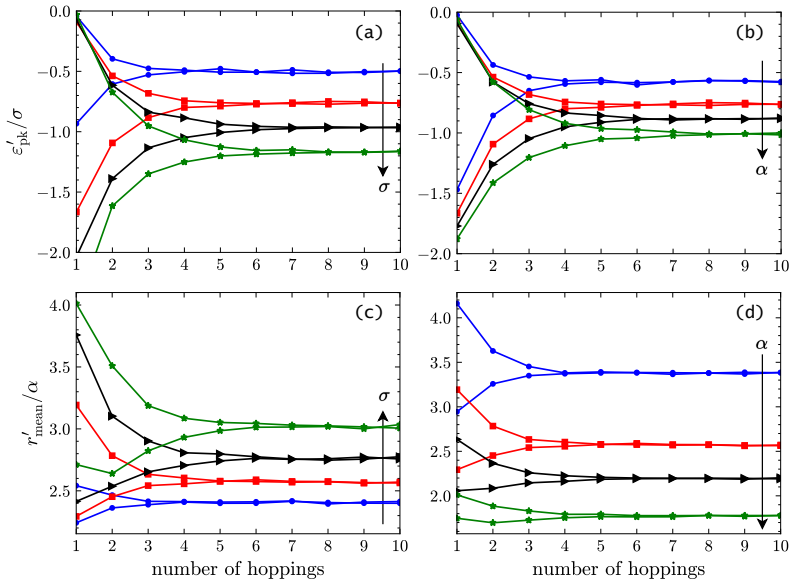

FIG. 5. (a) and (b): The evolution of the peak position of the destination energy when the output of the network is iteratively used as the new input (i.e., new origin energy). For each system with a specified σ and α, the initial input is chosen to be either 3σ or $-\sigma^2/k_{\rm B}T$, creating two branches of evolutions for each system. Results are for different amounts of disorder and localization length values, with $\sigma/k_{\rm B}T = 2, 3, 4, 6$ and $\alpha\rho^{1/3} = 0.3$ in part (a), and $\alpha\rho^{1/3} = 0.2, 0.3, 0.4, 0.6$ and $\sigma/k_{\rm B}T = 3$ in part (b). Parts (c) and (d) show the corresponding results for the mean values of the hopping distances extracted from the distribution of the hopping distances outputted by the network.

C. interpretability of the Network

Although the results in the previous section are promising, demonstrating the potential of exploiting neural networks to gain insights into the statistics of carrier transport in disordered systems, a potential drawback is the black-box nature of the neural network. In comparison to the mathematical description of a physical system, a neural network appears to be unable to provide any insight into the physics underlying the system under study. One initial attempt to address this concern is to explore the internal structure of the network in order to unravel how the network produces its output. Here, considering β-VED, let us take a deeper look at the weights connecting the latent layer to the decoder layer, as illustrated in Fig.6(a), to better understand the transformation and processing of information within the neural network. Among the entire network, these specific connections are chosen due to the bottleneck of the network in the latent layer,

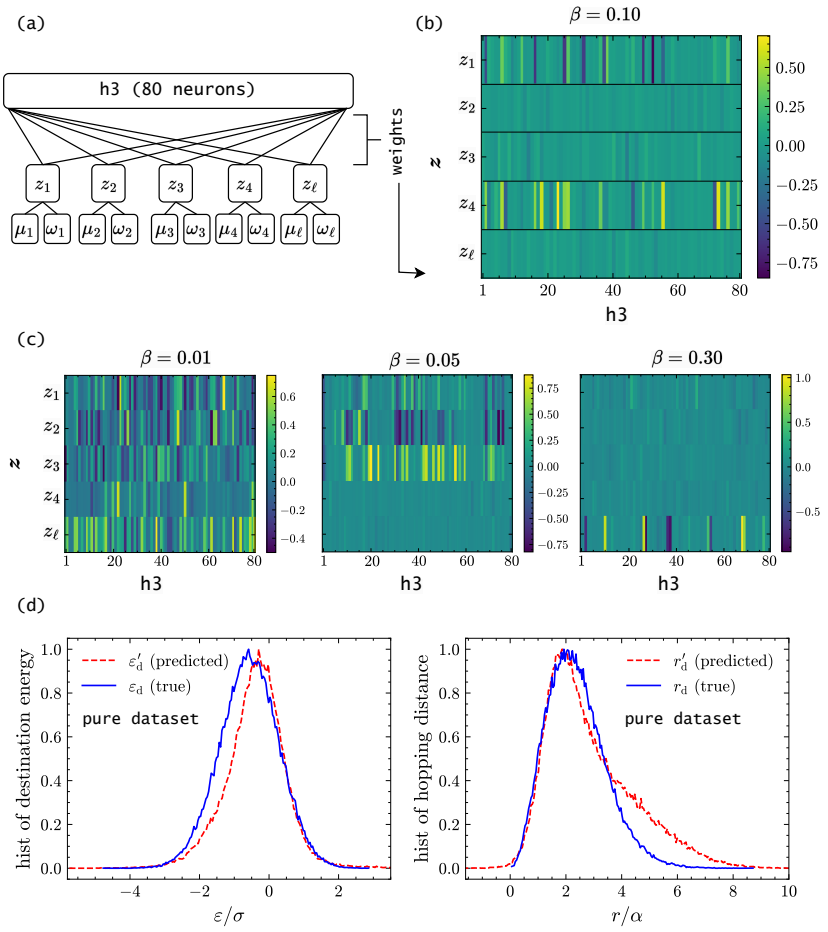

FIG. 6. (a): Schematic depiction of the weights connecting the latent representation (with five neurons z_1 to z_ℓ, where $\ell = 5$) to the decoder layer (h3 for the destination energy, as shown in Fig.2, with 80 neurons). (b): The corresponding weights (a total of 5×80) for the network trained with $\beta = 0.1$. All results in previous figures are for this network. (c): Results for the weights connecting the latent layer to the decoder for the networks separately trained with $\beta = 0.01, 0.05$ and 0.3. To clarify, in each panel, only the distinct, bright colors against the uniform background have nonzero values. (d): Histograms illustrating the distribution of destination energy and hopping distance, obtained by processing a pure test dataset with $\alpha\rho^{1/3} = 0.3$ and $k_{\rm B}T/\sigma = 0.3$, comparable to those in Fig.4(a) but using a network trained with $\beta = 0.3$.

making it reasonable to expect that the final output is essentially generated from the flow of information from this bottleneck to the decoder. It should be noted that the weights in a neural network, after the training process is finished, are constant and not influenced by the input. As can be seen in Fig.6(b), the results for the weight values connecting the latent layer of size 5 to the decoder layer of size 80, i.e. $\boldsymbol{z} \to$ h3,

clearly show that only two sets of connections are nonzero, implying that only two neurons in the latent layer are essential for determining the output of the network. Considering the architecture of β-VED in Fig.2, it should be pointed out that, although only the result for $\boldsymbol{z} \rightarrow$ h3 has been presented in Fig.6(b), the result for $\boldsymbol{z} \rightarrow$ h4 also exhibits similar behavior, suggesting that for both the destination energy and hopping distance, only two neurons in the latent layer are sufficient. In other words, the network achieves accurate results by effectively compressing the input data into just two dimensions.

The results discussed in the previous section and those from Fig.6(b) were all obtained using a network trained with a hyperparameter β value of 0.1 in the loss function of Eq.9. As we discussed earlier, this hyperparameter is introduced to control the contribution of the latent loss in the total loss. To examine the effect of different values for β, we trained additional networks with varying values of this parameter. The results in Fig.6(c) demonstrate the connecting weights, $\boldsymbol{z} \rightarrow$ h3, in these new networks. As seen in this figure, when β is increased, the network is forced to compress the data into a smaller number of neurons in the latent layer due to the increased contribution of the latent loss. On the other hand, because of a decrease in the contribution of the reconstruction loss, the network may not be able to produce sufficiently accurate output anymore. This point is demonstrated in Fig.6(d), in which one can see that for $\beta = 0.3$ (for which only one neuron is active in the latent layer), the reconstruction accuracy is not satisfactory. On the other hand, for lower values of β, although the desired accuracy can be achieved, the dimensionality reduction is not complete, and the network is not forced to extract only the necessary information from the data. (Note that the order of these active neurons is random.)

Returning to the case of $\beta = 0.1$ (and similar nearby values, whose results are not shown here) which ensures a necessary and sufficient number of neurons for describing the statistics of the system, let us explore what information is stored in the two active neurons. Fig.7(a) and (b) show the results at the stationary conditions for the information stored in the mean values of the latent layers when different pure datasets are fed into the network. The results clearly show that it is the information about the strength of disorder and localization length which is stored in the latent layer. Furthermore, no information is stored in the other three neurons, consistent with the behavior observed for the connecting weights. In fact, β-VED has been capable of automatically extracting hidden relevant information from the data on the one hand, and preserving just this information, almost *linearly*, in the latent neurons. (Because of this nearly linear combination,

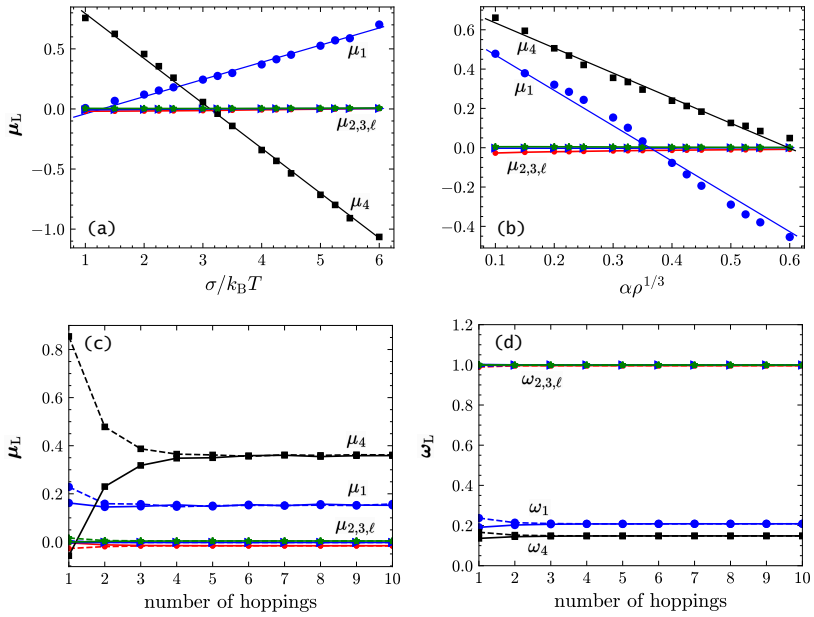

FIG. 7. (a) and (b): stationary mean values in the latent layer, $\boldsymbol{\mu}_{\rm L}$, as a function of the disorder and localization length. Results are for different amounts of disorder and localization length values, with $\alpha\rho^{1/3} = 0.3$ in part (a), and $\sigma/k_{\rm B}T = 2$ in part (b). In each part, the solid lines are guides to eyes. (c) and (d): The evolution of the mean and variance values in the latent layer of β-VED for a system with with $\sigma/k_{\rm B}T = 2$ and $\alpha\rho^{1/3} = 0.3$.

which is also reported in other problems [30], further disentanglement can be achieved by applying a linear transformation.) The decoder then uses this abstract information to generate its output. The importance of the results presented here lies in the fact that even without prior knowledge of the transport properties, the neural network can extract the necessary parameters and provide insights into the transport mechanism, automatically, from raw data.

Based on the concept of the transport energy, this energy level plays a similar role as that of the mobility edge in the multiple trapping model, implying that the hopping event to this universal energy level is essentially independent of the origin energy. The results presented in Fig.5, however, suggest that this energy level is, statistically, reached after a few steps if the starting energy is much higher or lower than the transport energy level, meaning that before reaching stationary conditions, the destination is dependent on

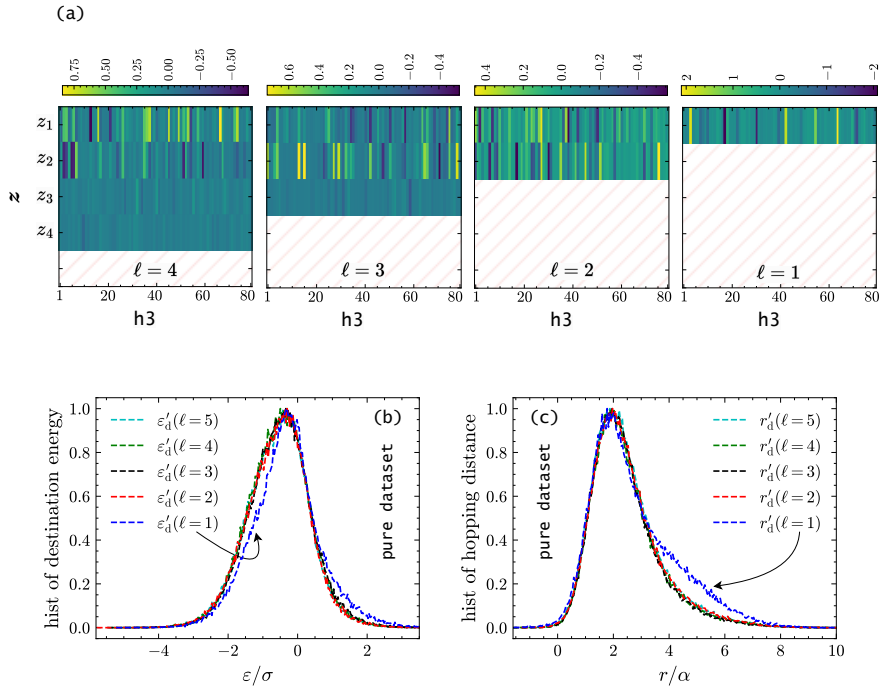

FIG. 8. (a): Depiction of the weights connecting the latent representation (with the dimensionality $\ell = 4, 3, 2,$ and 1) to the decoder layer (h3, with 80 neurons), for networks trained with $\beta = 0.1$. (b) and (c): Histograms showing the distribution of destination energies and hopping distances, obtained by processing a pure test dataset with $\alpha\rho^{1/3} = 0.3$ and $k_{\rm B}T/\sigma = 0.3$, for the networks of part (a). For comparison, the results for the case $\ell = 5$ are also presented in each panel.

the origin state. We may expect, therefore, that the information regarding this time evolution is also stored in the latent layer because the network generates this time evolution. Despite the results in Fig.7(a) and (b), which depict the values of the mean latent neurons at stationary conditions, we present the results for the time evolution towards stationary conditions in parts (c) for a specific system. As seen in this figure, the time evolution is also stored in the latent layer. The network, however, does not allocate an additional neuron to this information, as the same two active neurons encode time evolution along with disorder strength and localization length information. Meanwhile,the variances in the latent layer, Fig.7(d), suggest a value of 1 for inactive neurons. Alongside their mean of zero, this signifies that inactive neurons constitute Gaussian (normal) noise, lacking relevant information.

To provide further evidence of the self-learning capability of β-VED network, let us examine the effect

of the latent layer's dimensionality, ℓ, on the network's behavior in the bottleneck layer and its ability to reconstruct the true transport statistics in the output layer. As a reminder, the results presented in the previous figures were all obtained with $\ell = 5$. It should be noted that for an unknown problem in physics, one generally does not know the number of relevant parameters required to describe a system accurately. Identifying those parameters is a task for β-VED network during the training process. Therefore, in general, the latent layer dimension should be set high enough to allow the network to learn the relevant parameters. Otherwise, it cannot be guaranteed whether the network has found the necessary number of parameters for representing the input data in the latent layer. To address this point, we trained additional networks with varying latent space dimensions, specifically $\ell = 4, 3, 2$ and 1. Fig.8(a) presents the results for the weights connecting the latent representation to the decoder layer for these networks. As seen in the figure, for $\ell = 4, 3$, and 2, the number of active neurons in the latent layer (neurons with non-zero connecting weights) remains two, as in the case of $\ell = 5$. On the other hand, for $\ell = 1$, the network inevitably reduces the data to the only available neuron in the latent layer. However, as shown in part b and c of the figure, although the reconstruction capability of the networks remains the same for $\ell = 4, 3$, and 2, the network with $\ell = 1$ can no longer reconstruct the true statistics (the results for the case $\ell = 5$ are also shown in each panel, for comparison). This indicates that a single neuron is not sufficient for data representation, and necessarily, one needs $\ell > 1$ for this purpose. On the other hand, the network with $\ell = 2$ exhibits a reconstruction capability comparable to that of networks with $\ell = 5, 4$, or 3, implying that two is the sufficient number of latent neurons in the latent layer. In summary, these results demonstrate that, irrespective of the latent space dimension, the network is able to reduce the input data to the necessary and sufficient number of neurons (in this problem, two neurons), provided there are enough neurons in the latent layer.

III. CONCLUDING REMARKS AND SUMMARY

Data-driven approaches are increasingly being adopted as a novel perspective for addressing various physics problems, including those with established theoretical frameworks (e.g., employing machine learning techniques for the renowned Ising model [50]). Since this new perspective is still in its early stages, it is crucial to investigate its potential and possible limitations in examining physical systems. Our study contributes to

this endeavor by introducing a specific machine learning method to explore transport properties in disordered systems. Various elements of our methodology, such as data generation, neural network architecture, probabilistic output generation, network interpretation, and others, can be useful resources for further research within the scientific community. Therefore, apart from the results presented in our work, the methodology itself offers a rich area of research for harnessing the potential of machine learning perspectives in studying transport properties in disordered systems.

As pointed out in Introduction, VAE-based neural networks have earned their reputation as generative models owing to their capacity to yield new outputs without needing new input data post-training. Likewise, β-VED network introduced in this work can produce a series of artificial hopping events (much like specialized VAE networks employed for generating artificial human faces) using its iterative applications on an initial input, illustrated in Fig.4(b). This capability allows β-VED network to create a sequence of hopping events comparable to traditional kMC simulations, establishing it as a prospective alternative to kMC simulations. kMC simulations are essential for validating theories and modeling experimental observations in organic optoelectronics, with continuous efforts to improve their performance. However, a single β-VED network can generate high-quality artificial hopping events for various disorder and localization lengths while avoiding common challenges in traditional kMC simulations. These challenges include setting up a simulation box, managing periodic boundary conditions, identifying possible hopping events, and running numerous realizations to obtain sufficient trajectories for statistical averaging. The network's potential as a substitute for kMC simulations is currently being explored, with findings to be presented in future studies.

β-VED architecture presented in this work has demonstrated its ability to generalize its performance to inputs not seen during training, such as pure datasets, iterative applications on the initial input, and convergence to a universal energy level. One of the most significant predictions made by β-VED is the existence of a transport energy level, which plays a crucial role in the hopping events within the energy-disordered systems. However, it is important to note that the primary output of the network is the distribution of hopping distances and destination energies of hopping events. (In fact, it is from these outputs that one can employ β-VED network to generate artificial hopping events without implementing a full kMC simulation, as pointed out above.) The position of the transport energy is, therefore, a byproduct of these distributions and not the sole advantage of the network. Insightful information can be gleaned from these distributions

concerning the time evolution of transport properties (as explored in Fig.5). In contrast to the evolution of the carrier distribution in the density of states, the evolution of the hopping events toward the transport energy level have not been reported before. This is due to either the predominant focus on equilibrium conditions in existing approaches, like resistor networks or percolation theory, or the practical impossibility of extracting the effective jumps distribution from kMC simulations because of the so-called oscillatory hoppings.

The transport energy concept has proven to be an effective framework for describing transport properties in systems governed by the Miller-Abrahams hopping mechanism. However, when polaronic effects play a significant role in the transport properties of organic systems, it becomes necessary to account for the Marcus expression of the hopping rate. It is essential to recognize that the theoretical approach based on maximizing the hopping rate to determine the transport energy level (discussed in the Introduction section) is exclusively applicable to the Miller-Abrahams hopping rate and not the Marcus rate. In contrast, β-VED network can be employed to explore the possibility of a transport energy level in the context of Marcus transfer rate. Notably, the effective dimensionality reduction of β-VED network facilitates the parameterization of transport properties under the influence of polaronic effects. Consequently, one of the primary objectives of this work is to introduce a novel tool for studying transport properties in systems with varying transport mechanisms, thereby expanding the existing toolkit available to researchers in this field.

The interpretability of neural networks remains an essential issue receiving significant attention. We explored the interpretability of β-VED network, revealing its capability to extract essential parameters for describing transport properties in disordered systems. This capability holds promise for studying transport properties in various other systems, such as exciton transport in host-guest systems, annihilation of excitons in light-emitting diodes, and transport properties in an exponential DOS and correlated disorder. The ability to identify necessary and sufficient parameters for describing a system lies at the core of the current paradigm of science. In line with previous reports, β-VED demonstrates that this task can be assigned to neural networks, underscoring the potential for further research in this domain. Future efforts could, therefore, also focus on devising new network architectures that can provide a more accurate description of the system while maintaining interpretability.

ACKNOWLEDGMENTS

S.D.B. acknowledges financial support by the Deutsche Forschungsgemeinschaft (Research Training Group "TIDE", RTG2591).

---

[1] S. Baranovskii, Theoretical description of charge transport in disordered organic semiconductors, Phys. Stat. Sol. B **251**, 487 (2014).

[2] A. Köhler and H. Bässler, *Electronic processes in organic semiconductors: An introduction* (John Wiley & Sons, 2015).

[3] S. Baranovski, ed., *Charge transport in disordered solids with applications in electronics* (John Wiley & Sons, 2006).

[4] O. V. Mikhnenko, P. W. Blom, and T.-Q. Nguyen, Exciton diffusion in organic semiconductors, Energ. Environ. Sci. **8**, 1867 (2015).

[5] H. Bässler, Charge transport in disordered organic photoconductors. a monte carlo simulation study, Phys. Stat. Sol. B **175** (1993).

[6] J. Oelerich, D. Huemmer, and S. Baranovskii, How to find out the density of states in disordered organic semiconductors, Phys. Rev. Lett. **108**, 226403 (2012).

[7] T. Upreti, Y. Wang, H. Zhang, D. Scheunemann, F. Gao, and M. Kemerink, Experimentally validated hopping-transport model for energetically disordered organic semiconductors, Phys. Rev. Appl. **12**, 064039 (2019).

[8] B. I. Shklovskii and A. L. Efros, *Electronic properties of doped semiconductors*, Vol. 45 (Springer, Berlin, 1984).

[9] N. Mott, Charge transport in non-crystalline semiconductors, in *Festkörper Probleme IX* (Elsevier, 1969) pp. 22–45.

[10] G. Schönherr, H. Bässler, and M. Silver, Simulation of carrier transport and energy relaxation in a macroscopic hopping system of sites with a gaussian energy distribution, Philosophical Magazine B **44**, 369 (1981).

[11] W. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P. Bobbert, P. Blom, . f. D. de Leeuw, and M. Michels, Unified description of charge-carrier mobilities in disordered semiconducting polymers, Phys. Rev. Lett. **94**, 206601 (2005).

[12] A. Nenashev, F. Jansson, J. Oelerich, D. Huemmer, A. Dvurechenskii, F. Gebhard, and S. Baranovskii, Advanced percolation solution for hopping conductivity, Phys. Rev. B **87**, 235204 (2013).

[13] S. Baranovskii, T. Faber, F. Hensel, and P. Thomas, The applicability of the transport-energy concept to various disordered materials, J. Phys.: Condens. Matter **9**, 2699 (1997).

[14] J. Orenstein and M. Kastner, Thermalization and recombination in amorphous semiconductors, Solid State Commun. **40**, 85 (1981).

[15] S. Baranovskii, M. Zhu, T. Faber, F. Hensel, P. Thomas, M. Von Der Linden, and W. Van der Weg, Thermally stimulated conductivity in disordered semiconductors at low temperatures, Phys. Rev. B **55**, 16226 (1997).

[16] M. Ansari-Rad, J. A. Anta, and J. Bisquert, Interpretation of diffusion and recombination in nanostructured and energy-disordered materials by stochastic quasiequilibrium simulation, J. Phys. Chem. C **117**, 16275 (2013).

[17] D. Monroe, Hopping in exponential band tails, Phys. Rev. Lett. **54**, 146 (1985).

[18] S. Baranovskii, H. Cordes, F. Hensel, and G. Leising, Charge-carrier transport in disordered organic solids, Phys. Rev. B **62**, 7934 (2000).

[19] S. S. Lemus and J. Hirsch, Hole transport in isopropyl carbazole—polycarbonate mixtures, Philos. Mag. B **53**, 25 (1986).

[20] P. Borsenberger, L. Pautmeier, and H. Bässler, Charge transport in disordered molecular solids, J. Chem. Phys. **94**, 5447 (1991).

[21] R. Schmechel, Gaussian disorder model for high carrier densities: Theoretical aspects and application to experiments, Phys. Rev. B **66**, 235206 (2002).

[22] S. Baranovskii, A. Nenashev, D. Hertel, K. Meerholz, and F. Gebhard, Parametrization of the charge-carrier mobility in organic disordered semiconductors, Phys. Rev. Appl. **22**, 014019 (2024).

[23] J. Gonzalez-Vazquez, J. A. Anta, and J. Bisquert, Random walk numerical simulation for hopping transport at finite carrier concentrations: diffusion coefficient and transport energy concept, Phys. Chem. Chem. Phys. **11**, 10359 (2009).

[24] D. Mendels and N. Tessler, The topology of hopping in the energy domain of systems with rapidly decaying density of states, J. Phys. Chem. C **117**, 24740 (2013).

[25] J. Oelerich, F. Jansson, A. Nenashev, F. Gebhard, and S. Baranovskii, Energy position of the transport path in disordered organic semiconductors, J. Phys.: Condens. Matter **26**, 255801 (2014).

[26] R. Kitchin, Big data, new epistemologies and paradigm shifts, Big data and society **1**, 2053951714528481 (2014).

[27] 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).

[28] M. L. Minsky, Logical versus analogical or symbolic versus connectionist or neat versus scruffy, AI Mag. **12**, 34 (1991).

[29] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, Theory-guided data science: A new paradigm for scientific discovery from data, IEEE Trans. Knowl. Data Eng..

[30] R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, Discovering physical concepts with neural networks, Phys. Rev. Lett. **124**, 010508 (2020).

[31] D. P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013); An introduction to variational autoencoders, Found. Trends Mach. Learn. **12**, 307 (2019).

[32] M. Lakshminarayanan, R. Dutta, D. M. Repaka, S. Jayavelu, W. L. Leong, and K. Hippalgaonkar, Comparing data driven and physics inspired models for hopping transport in organic field effect transistors, Sci. Rep. **11**, 23621 (2021).

[33] C. Lortaraprasert and J. Shiomi, Robust combined modeling of crystalline and amorphous silicon grain boundary conductance by machine learning, Npj Comput. Mater. **8**, 219 (2022).

[34] M. Cheng, C. Wang, C. Qin, Y. Zhang, Q. Zhang, H. Li, and J. Chen, Predicting macroscopic properties of amorphous monolayer carbon via pair correlation function, arXiv preprint arXiv:2410.03116 (2024).

[35] Y. Dou, K. Shimizu, J. Carrete, H. Fujioka, and S. Watanabe, Machine-learning potential for phonon transport in aln with defects in multiple charge states, arXiv preprint arXiv:2409.16039 (2024).

[36] J. Lederer, W. Kaiser, A. Mattoni, and A. Gagliardi, Machine learning–based charge transport computation for pentacene, Adv. Theory Simul. **2**, 1800136 (2019).

[37] T. Tan and D. Wang, Machine learning based charge mobility prediction for organic semiconductors, J. Chem. Phys. **158** (2023).

[38] C. Wechwithayakhlung, G. R. Weal, Y. Kaneko, P. A. Hume, J. M. Hodgkiss, and D. M. Packwood, Exciton diffusion in amorphous organic semiconductors: Reducing simulation overheads with machine learning, J. Chem. Phys. **158** (2023).

[39] H. Bässler, Localized states and electronic transport in single component organic solids with diagonal disorder, Phys. Stat. Sol. B **107**, 9 (1981).

[40] D. Beljonne and J. Cornil, *Multiscale modelling of organic and hybrid photovoltaics* (Springer, 2014).

[41] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (" O'Reilly Media, Inc.", 2022).

[42] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rev. **810**, 1 (2019).

[43] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167 (2015).

[44] F. Chollet *et al.*, Keras, https://keras.io (2015).

[45] B. Lakshminarayanan, A. Pritzel, and C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, Adv. Neural. Inf. Process. Syst. **30** (2017).

[46] I. Higgins, L. Matthey, A. Pal, C. P. Burgess, X. Glorot, M. M. Botvinick, S. Mohamed, and A. Lerchner, beta-vae: Learning basic visual concepts with a constrained variational framework., ICLR (Poster) **3** (2017).

[47] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, Understanding disentangling in beta-vae, arXiv preprint arXiv:1804.03599 (2018).

[48] G. Fernández-Fernández, C. Manzo, M. Lewenstein, A. Dauphin, and G. Muñoz-Gil, Learning minimal representations of stochastic processes with variational autoencoders, Phys. Rev. E **110**, L012102 (2024).

[49] A. Nenashev, J. Oelerich, and S. Baranovskii, Theoretical tools for the description of charge transport in disordered organic semiconductors, J. Phys.: Condens. Matter **27**, 093201 (2015).

[50] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, Phys. Rev. E **96**, 022140 (2017).