ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

D4: Distance Diffusion for a Truly Equivariant Molecular Design

Samuel Cognolato[1,2], Davide Rigoni[3✉], Marco Ballarini[1], Luciano Serafini[2], Stefano Moro[3] and Alessandro Sperduti[1,2,4 *]

1- University of Padova - Dept of Mathematics "Tullio Levi-Civita"
Via Trieste, 63, 35121 Padova - Italy

2- Fondazione Bruno Kessler
Via Sommarive, 18, 38123 Trento - Italy

3- University of Padova - Dept of Pharmaceutical and Pharmacological Sciences
Via Marzolo, 5, 35131 Padova - Italy

4- University of Trento - Dept of Information Engineering and Computer Science
Via Sommarive, 9, 38123 Trento - Italy

Abstract. Recent years have witnessed an increase in interest in leveraging generative models for *de novo* molecular design in drug discovery. Many State-of-the-Art (SotA) models incorporate the 3D structural information of the molecule, particularly atomic spatial coordinates. However, such approaches face challenges integrating SE(3) equivariance when trained on coordinates. This work explores the use of the distance matrix for molecular structures, natively SE(3) invariant, avoiding whatever the issue. Experimental evaluation shows that our proposed approach significantly improves upon MiDi, a SotA 3D molecule generator.

1 Introduction

De novo molecular design involves creating new chemical molecules with specific properties, a critical task in drug discovery and material science fields. Given the complexity related to the chemical space, where only a small subset of molecules are both chemically feasible and possess desirable properties, computational methods have become essential. As traditional approaches are slow and costly, recent advancements in artificial intelligence and computational power have led to significant progress in applying generative models to molecular design [1]. With the advent of deep generative models utilizing graph representations, some methods incorporate only structural topology [2, 3], while others include a richer representation in 3D space [4, 5, 6]. When considering 3D coordinates, several challenges arise. One of the primary issues is achieving rotational and translational invariance, also known as SE(3) invariance, as molecular properties remain unchanged under such transformations. Models must effectively encode 3D spatial information while avoiding sensitivity to arbitrary coordinate system choices [7]. Secondly, methods must also consider the natural symmetries

✉Corresponding author: davide.rigoni.1@unipd.it

* The authors acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources, and thank the MMS, University of Padova, for verifying the chemical structures.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

in graph representations. Addressing these issues often involves leveraging techniques like equivariant neural networks or graph-based representations tailored to preserve geometric invariances [8], although they come with some challenges. These models are computationally intensive and require complex operations, like spherical harmonics for encoding equivariance, making them difficult to implement and scale for large datasets. The approach proposed in [9] focuses on the conformation, splitting its generation into finding local structures and torsion angles. This task has also been tackled with Diffusion Models by applying Gaussian noise to 3D coordinates [10, 6]. These models explicitly build in equivariance in the diffusion process using a zero Center of Mass (CoM) Gaussian distribution and using SE(3) equivariant features as input to the denoising model. Still, [10] doesn't include the generation of the topology, while [6] (MiDi) is missing SE(3) equivariance in the training target. More specifically, the Graph Transformer [11] in MiDi is trained to predict the centered coordinates of atoms, making the target translation invariant but not rotation invariant.

This work introduces D4, a Distance and Discrete Denoising Diffusion Model based on MiDi [6], aiming to enhance chemical validity in molecular generation by leveraging atomic distances as a proxy for spatial information. Incorporating distances between atoms is intuitively beneficial, as spatial arrangements critically influence a molecule's physical and chemical properties. Unlike 3D atomic coordinates [10, 6] which require SE(3) equivariant architectures, this work employs the natively invariant distance matrix, simplifying the overall architecture. The distance matrix approach offers additional advantages, such as leveraging bond-dependent atomic distances (e.g., shorter triple bond lengths versus single bonds) to improve conformer accuracy. While coordinate-based models implicitly handle distances, explicitly integrating distance information enhances information flow and improves results. Appendices, as Supplementary Material, are available at: `https://www.math.unipd.it/~drigoni/files/ESANN2025-D4-Supplementary.pdf`.

2 Our Approach: Diffusion on Distances

Motivated by the shortcomings of previous methods, we introduce D4, a Distance and Discrete Denoising Diffusion model based on MiDi [6]. D4 generates the molecular structure and replaces 3D coordinates with pairwise distances $\mathbf{D} \in \mathbb{R}^{n \times n}$, which are natively SE(3) equivariant features. Further motivation derives from the common knowledge in chemistry that bond distances are closely related to bond types. Denoising Diffusion Probabilistic Models (DDPMs) [12] learn a distribution of data by reversing a noise process, which gradually corrupts data points. A diffusion transition for distances is defined as $q(\mathbf{D}_t|\mathbf{D}_{t-1}) = \mathcal{N}(\mathbf{D}_t;\ \alpha_t \mathbf{D}_{t-1},\ \sigma_t^2 \boldsymbol{I})$, with α_t, σ_t being time-dependent parameters from the Cosine Schedule [13], defined as $\bar{\alpha}_t = \cos^2(0.5\pi(t/T + s)/(1 + s))$ with a small s, where $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ and $\bar{\sigma}_t = \sigma_t^2 - \alpha_t^2$. A requirement of DDPMs is to have stationary distributions that are simple and known. In the gaussian case $q(\mathbf{D}_T) = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ for $T \to \infty$. Specifically, the objective of DDPMs is to learn

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

the reverse of the diffusion process through a neural network as $p_\theta(\mathbf{D}_{t-1}|\mathbf{D}_t) = \mathcal{N}(\mathbf{D}_{t-1};\ \mu_t\hat{\mathbf{D}}_\theta + \nu_t\mathbf{D}_t,\ \tilde{\sigma}_t^2\boldsymbol{I})$, with $\hat{\mathbf{D}}_\theta$ being the network prediction, $\mu_t = \bar{\alpha}_t(1 - \alpha_t^2\bar{\sigma}_{t-1}^2/\bar{\sigma}_t^2)$, $\nu_t = \alpha_t\bar{\sigma}_{t-1}^2/\bar{\sigma}_t^2$ and $\tilde{\sigma}_t = \bar{\sigma}_{t-1}^2(1 - \alpha_t^2\bar{\sigma}_{t-1}^2/\bar{\sigma}_t^2)$. Regarding the generation of the graph structure, we start from the diffusion and denoising processes of [6], which are based on Discrete Diffusion [14, 2]. For a molecular graph $\mathcal{G} = (\boldsymbol{X}, \boldsymbol{C}, \mathbf{E})$, atoms, formal charges and bonds are modeled as one-hot encoded node and edge labels $\boldsymbol{X} \in \{0,1\}^{n\times d_x}$, $\boldsymbol{C} \in \{0,1\}^{n\times d_c}$, $\mathbf{E} \in \{0,1\}^{n\times n\times d_e}$, where d_x, d_c, d_e are the number of types of atoms, formal charges, and types of bonds, respectively. The non-existing bond is a bond type indexed as $\mathbf{E}[0]$. These evolve through three Markov Chains of the form: (i) $q(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) = \boldsymbol{X}_{t-1}Q_t^X$; (ii) $q(\boldsymbol{C}_t|\boldsymbol{C}_{t-1}) = \boldsymbol{C}_{t-1}Q_t^C$; (iii) $q(\mathbf{E}_t|\mathbf{E}_{t-1}) = \mathbf{E}_{t-1}Q_t^E$; where $Q_t^X \in [0,1]^{d_x\times d_x}$, $Q_t^C \in [0,1]^{d_c\times d_c}$ and $Q_t^E \in [0,1]^{d_e\times d_e}$ are the three time-dependent transition matrices [14], which collapse labels to follow uniform distributions at $T \to \infty$. Regarding the denoising of edges, differently from other works [2, 6], we split edge prediction into two stages: first, our model predicts the adjacency matrix $\mathbf{A} = 1 - \mathbf{E}[0]$ with probabilities p_θ^A, which signifies the presence of a bond or not; then, only for existing bonds, the probabilities p_θ^E classifies these entries in a single, double and triple bond. We denote classes of existing bonds as the selection $\mathbf{E}[\mathbf{A}]$ of edges where $\mathbf{A}_{ij} = 1$. The motivation lies in balancing the sparsity in the training signal for edges, which is common in molecular graphs. Thus the denoising model predicts the true labels as probabilities $p_\theta^X, p_\theta^C, p_\theta^A, p_\theta^E$. The full diffusion and denoising processes assume the following form:

$$q(\mathcal{G}_t|\mathcal{G}_{t-1}) = \left(\mathcal{N}(\mathbf{D}_t;\ \alpha_t\mathbf{D}_{t-1},\ \sigma_t^2\boldsymbol{I}),\ \boldsymbol{X}_{t-1}Q_t^X,\ \boldsymbol{C}_{t-1}Q_t^C,\ \mathbf{E}_{t-1}Q_t^E\right), \tag{1}$$

$$p_\theta(\mathcal{G}_{t-1}|\mathcal{G}_t) = p_\theta(\boldsymbol{X}_{t-1}|\mathcal{G}_t)p_\theta(\boldsymbol{C}_{t-1}|\mathcal{G}_t)p_\theta(\mathbf{E}_{t-1}|\mathcal{G}_t)p_\theta(\mathbf{D}_{t-1}|\mathcal{G}_t). \tag{2}$$

We share the same architecture as MiDi [6] taking as input the graph $\mathcal{G}_t$, but removing the coordinates $\boldsymbol{R}$ from the model. Using distances as a way to obtain the 3D conformation circumvents completely issues related to SE(3) equivariance. We integrate $\mathbf{D}$ into the Graph Transformer architecture [11], and avoid the special treatment given to coordinates in [10, 6], keeping the model simple. The network, with parameters θ, can be trained through the loss function:

$$\begin{aligned}\mathcal{L}(\mathcal{G}, \hat{\mathcal{G}}_t) = \lambda_x\,\mathrm{CE}(\boldsymbol{X}, p_\theta^X) + \lambda_e\left(\mathrm{BCE}(\mathbf{A}, p_\theta^A) + \mathrm{CE}(\mathbf{E}[\mathbf{A}], p_\theta^E)\right) + \\ +\lambda_c\,\mathrm{CE}(\boldsymbol{C}, p_\theta^C) + \lambda_d\|\hat{\mathbf{D}}_\theta - \mathbf{D}\|^2,\end{aligned} \tag{3}$$

where CE and BCE are, respectively, the Cross-Entropy and the Binary Cross-Entropy, λ variables are losses weights. From the distance matrix $\mathbf{D}$, not necessarily satisfying the triangle inequality, we recover the 3D coordinates $\boldsymbol{R}$ using the Multidimensional Scaling (MDS) algorithm [15]. It computes the eigenvectors of the Gram matrix of $\mathbf{D}$, and returns the three with the highest associated eigenvalues as coordinate matrix $\boldsymbol{R} \in \mathbb{R}^{n\times 3}$.

3 Experimental Assesment

Dataset and Evaluation Metrics All models are evaluated on the QM9 and GDB13 datasets, split into 75% training, 15% validation, and 10% testing. The

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

QM9 dataset [16] is a widely adopted benchmark, containing 134K small organic molecules with up to nine heavy atoms. The GDB13 dataset [17] contains about one million molecules with up to 13 heavy atoms. In this work, we adopt a random subset of 169K molecules. More details are reported in Appendix A.

The final testing uses the best checkpoint based on denoising loss from the validation set, with results from a 10K molecule sample. Our model and MiDi results are shown as averages with 95% confidence intervals across 5 generated example sets. Several well-known metrics [18] in literature are considered to evaluate the models' generation quality: (i) *Validity*: assesses the number of model's generated structures that are chemically valid[1]. (ii) *Uniqueness*: is the ratio of unique generated molecules over all molecules. This metric is essential to verify the occurrence of the model collapse issue; (iii) *Novelty*: is the percentage of generated molecules that are not in the training set, measuring the generalization power of the model. (iv) *Distance Error*: measures the Wasserstein distance among the frequencies of distances in the test set and those generated by the model. More details are reported in Appendix B.

Model Selection and Implementation Details Model selection was performed only on QM9 using a grid search strategy focusing on the learning rate and the loss function weights (Equation 3). Then, the same values are also applied to GDB13. The former was tested for values in the range of 0.001 and 0.01, refining incrementally by 0.002. For loss function weights, we started from the configuration of MiDi, then explored variations with steps of ± 0.5.

All the models considered in this work have been trained with a batch size of 1024 and for 2000 epochs. For reproducibility, the remained of the experimental setup follows the proposal in MiDi. The results from the grid search yielded a value of 0.002 for the learning rate and the following loss weights: $\lambda_x = 1$, $\lambda_c = 1$, $\lambda_e = 2$ and $\lambda_d = 1$. In this work, hydrogens are not considered [6, 7, 2], and all the pre-processing and post-processing stages performed on molecules are done using RDkit. Our proposed model consists of 20,003,660 parameters vs. the 20,124,219 of MiDi. More details are reported in Appendix C. The implementation code is available at: `https://github.com/MrcBalla/D4-code`.

Results and Discussion Table 1 compares our model's performance with the SotA MiDi approach across the QM9 and GDB13 datasets. Notably, both models exhibit very similar performance on these datasets considering the Validity and Uniqueness metrics. However, on QM9 our proposed model shows less Novelty than MiDi. In the literature, there is an emphasis on maximizing Novelty. However, this is not entirely applicable to the QM9 dataset, as noted by the authors of [2], because QM9 is a curated dataset comprising specific molecules that satisfy certain constraints. Within these constraints, QM9 serves as an exhaustive enumeration of molecular types, and thus, having a high Novelty might suggest that the learning architecture is struggling to capture the chemical space adequately. Based on this, overall, we can conclude that our model's generation capability matches that of MiDi regarding the molecule's topology structures.

[1]RDkit is used to verify the molecule's structure: `https://www.rdkit.org/`

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

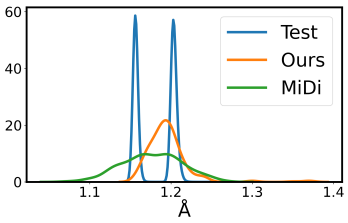

Fig. 1: KDE of triple bond distances.

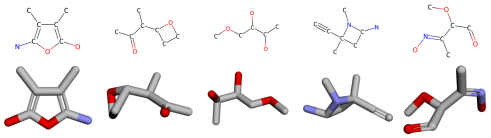

Fig. 2: Examples of molecules generated by our model on the QM9 dataset, shown in 2D and 3D.

Model	QM9				GDB13			
	Validity	Uniqueness	Novelty	Dist. Error	Validity	Uniqueness	Novelty	Dist. Error
MiDi	99.5±0.1	95.8±0.2	49.2±0.0	0.008±0.003	99.7±0.1	99.9±0.1	99.9±0.1	0.026±0.008
Ours	99.5±0.1	95.7±0.1	41.2±0.4	0.005±0.001	99.8±0.1	99.9±0.1	99.9±0.1	0.004±0.003

Table 1: Comparison between our approach and MiDi.

Model	QM9			GDB13		
	Single	Double	Triple	Single	Double	Triple
MiDi	0.017±0.003	0.027±0.007	0.066±0.029	0.066±0.006	0.081±0.028	0.115±0.024
Ours	0.009±0.002	0.011±0.001	0.269±0.009	0.014±0.003	0.048±0.027	0.087±0.031

Table 2: Distance comparison between our approach and MiDi.

Bigger improvements are found for the Distance Error metric where our model outperforms MiDi in both datasets. A closer inspection of Table 2, reporting the Distance Error for each bond type in QM9 and GDB13, highlights that our model outperforms MiDi in all cases except for triple bonds in the QM9 dataset. However, a closer look at Figure 1, displaying the KDE of the distances of each model compared to that in the test set, reveals that both our approach and MiDi perform poorly and are not able to detect the two modalities. One hypothesis is that the model may struggle to learn the triple bond type. However, a detailed analysis of the predicted atom and bond type frequencies, provided in Appendix D, suggests otherwise. Thus we argue that this could be caused by two main facts: (i) triple bonds represent about the 3% (1% in GDB13) of all the bonds in the dataset making the learning task difficult; and (ii) there is in, both double and triple bond distance distribution, a mode at 1.2Å which complicates identifying distinct patterns. More KDE plots are reported in Appendix E.

Figure 2 showcases some example molecules generated by our model trained on QM9 which have been verified by computational chemistry experts. Example molecules generated from the GDB13 dataset can be found in Appendix F.

4 Conclusions and Future Work

Recent advances in deep generative models have improved molecular modelling using 3D coordinates of atoms, although rotational invariance challenges remain.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

This work proposes D4, an SE(3) model that leverages diffusion on distances to reconstruct 3D molecular structures, generating chemically valid molecules surpassing a SotA approach. Future work will condition distances on bond types within the diffusion framework to enhance accuracy, especially for triple bonds, and inserting prior knowledge about distances in the diffusion model.

References

[1] Davide Rigoni, Sachithra Yaddehige, Nicoletta Bianchi, Alessandro Sperduti, Stefano Moro, and Cristian Taccioli. Tumflow: An ai model for predicting new anticancer molecules. *International Journal of Molecular Sciences*, 25(11):6186, 2024.

[2] Clément Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. In *ICLR*, 2023.

[3] Samuel Cognolato, Alessandro Sperduti, and Luciano Serafini. Ifh: a diffusion framework for flexible design of graph generative models. In *ECAI*, pages 3039–3046. 2024.

[4] Davide Rigoni, Nicolò Navarin, and Alessandro Sperduti. Conditional constrained graph variational autoencoders for molecule design. In *IEEE SSCI*, pages 729–736. IEEE, 2020.

[5] Davide Rigoni, Nicolò Navarin, and Alessandro Sperduti. Rgcvae: Relational graph conditioned variational autoencoder for molecule design. *Machine Learning*, 114(2):47, 2025.

[6] Clement Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation. In *ECML PKDD*, pages 560–576. Springer, 2023.

[7] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *ICML*, pages 8867–8887. PMLR, 2022.

[8] Yi-Lun Liao and Tess E. Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *ICLR*, 2023.

[9] Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. In *NeurIPS*, volume 34, pages 13757–13769, 2021.

[10] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *ICLR*, 2022.

[11] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2020.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, volume 33, pages 6840–6851, 2020.

[13] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, pages 8162–8171. PMLR, 2021.

[14] Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *NeurIPS*, volume 34, pages 17981–17993, 2021.

[15] Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.

[16] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

[17] Lorenz C Blum and Jean-Louis Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database gdb-13. *Journal of the American Chemical Society*, 131(25):8732–8733, 2009.

[18] Davide Rigoni, Navarin Nicolo, and Sperduti Alessandro. A systematic assessment of deep learning models for molecule generation. In *ESANN*, pages 547–552, 2020.