# Inverse-design of organometallic catalysts with guided equivariant diffusion

**François Cornet**[1] **Bardi Benediktsson**[2] **Bjarke Hastrup**[2]
**Arghya Bhowmik**[2] **Mikkel N. Schmidt**[1]
[1]DTU Compute [2]DTU Energy
{frjc,bardbe,bjaha,arbh,mnsc}@dtu.dk

## Abstract

Organometallic complexes are ubiquitous in homogenous catalysis, and their optimisation is of particular interest for many technologically relevant reactions. However, due to the large variety of possible metal-ligand and ligand-ligand interactions, finding the best combination of metal and ligands is an immensely challenging task. Here we present an inverse design framework based on a diffusion generative model for *in-silico* design of such complexes. Given the importance of the spatial structure of a catalyst, the model directly operates on all-atom (including explicit H) representations in 3D space. To handle the symmetries inherent to that data representation, it combines an equivariant diffusion model and an equivariant property predictor to drive sampling at inference time. We illustrate the potential of the proposed framework by optimising catalysts for the Suzuki-Miyaura cross-coupling reaction, and validating a selection of novel proposed complexes with DFT.

## 1 Introduction

In-silico catalyst design is a grand chemical challenge [1, 2], and the combination of machine learning (ML) and quantum chemistry (QC) methods is an appealing strategy. On one hand, surrogate ML models can be trained on reference databases and, in turn, used to speed up property evaluation while preserving the accuracy of the reference method. On the other hand, generative ML models can learn the distribution of the chemical space of interest, and in turn generate novel chemical structures that share aggregate properties with the training data (*distribution learning*). The combination of surrogate and generative modelling opens the door to the inverse-design of compounds with optimised properties, i.e. *goal-directed generation* [3].

ML inverse catalyst design involves defining the relevant chemical space for a reaction under study, as well as collecting a suitable amount of QC data for training. While directly learning a conditional generative model can sometimes be an option, it requires a sufficient number of labeled samples with the desired properties, which is often not attainable. Instead, *guidance* decouples the generative process from the conditional information, by only using the latter at inference time to steer the generative model towards the target properties. Larger databases with relevant candidates [4–6] can then be leveraged for training the generative model, and the amount of task-specific data, e.g. an energy barrier for a particular reaction, can be limited to what the surrogate requires to be satisfactorily accurate.

The geometry of a catalyst plays a great role in catalysis, and methods that operate on molecular graphs or string representations lack information about the 3D structure of studied complexes. A given molecular graph, or string, can potentially correspond to multiple spatial arrangements with greatly varying properties. Additionally, bonding information is not properly defined for complexes involving transition metals—requiring non-standardized descriptors [7]. Recently, generative models

37th Conference on Neural Information Processing Systems (NeurIPS 2023) – AI4Mat Workshop.

for 3D atomistic structures [8–10] have reached a point where they can compete with geometry-free models, and the diffusion paradigm [9, 11, 12] seems particularly promising. Since the original equivariant diffusion for molecules [9], numerous variants have been proposed to tackle various problems such as conformer generation [13], linker design [14], structure-based design [15], or target-aware design [16].

In this work, we introduce a guided equivariant diffusion model specifically designed to generate organometallic complexes with optimised properties. In what follows, we assume that (1) we know the kind of compounds that catalyse the reaction, and (2) the mechanism of the catalysed reaction, along with the rate-determining step.

An overview of the conditional generation process is presented in Fig. 1. We summarise our main contributions as follows:

- We implement a 3D equivariant diffusion generative model, specifically designed for organometallic complexes;
- We train an equivariant property predictor, and use it in combination with the diffusion model to perform regressor-guidance and sample organometallic catalysts with optimised properties;
- For the specific problem of optimising a critical step in the Suzuki-Miyaura cross-coupling reaction, we close the loop by validating a selection of generated complexes using DFT.

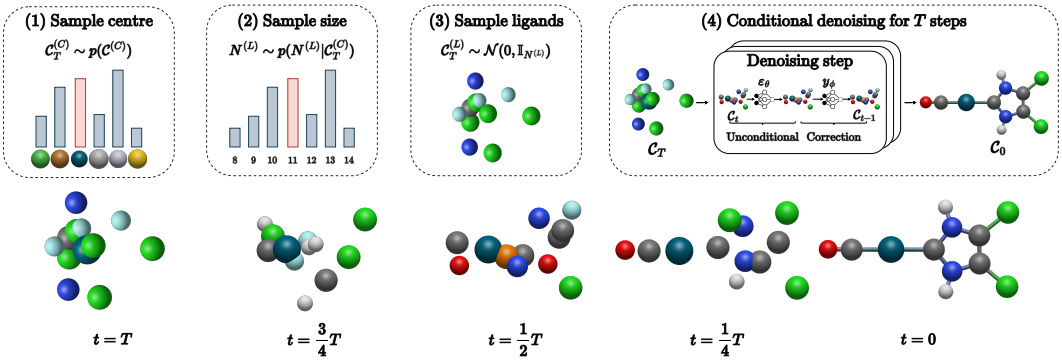


Figure 1: (**Top**) Overview of the conditional generation process of an organometallic catalyst $\mathcal{C}$: (1) A metal centre atomic type, $\mathcal{C}_T^{(C)}$, is sampled, (2) based on the centre, the number of atoms contained in the coordinated ligands is sampled, (3) random atomic types and positions are assigned to all ligand atoms, and (4) the conditional denoising runs for $T$ steps. Each denoising step involves an unconditional denoising update (steering towards valid molecules, via $\varepsilon_\theta$) followed by a property target correction (steering towards molecules with the desired properties, via $y_\phi$). (**Bottom**) Example of denoising trajectory for a complex with a Pd centre. The position and atomic type of the centre are kept fixed during the whole trajectory, and only the surrounding atoms are denoised.

## 2 Methods

**Data representation** An organometallic complex is (often) composed of a (or multiple) transition metal centre surrounded by organic ligands coordinated in very specific ways. It can be represented as a set of atoms $\mathcal{C} = [\mathcal{C}^{(C)}, \mathcal{C}^{(L)}] = \{[x^{(C)}, h^{(C)}], [x^{(L)}, h^{(L)}]\}$, where $x^{\{(C),(L)\}} \in \mathbb{R}^{\{N_C, N_L\}\times 3}$ represents the atomic coordinates and $h^{\{(C),(L)\}} \in \mathbb{R}^{\{N_C, N_L\}\times M}$ the atomic types. Superscript $(C)$ refers to the centre (and possibly some proximal atoms), while $(L)$ denotes all other atoms belonging to the ligands coordinated around the centre. Depending on the problem under study, $x^{(C)}$ and/or $h^{(C)}$, or possibly parts thereof, can be fixed and viewed as a type of context, such as when the centre or the coordination pattern is known. Finally, we denote by $y \in \mathbb{R}$ the property of interest associated with a complex.

**Equivariant diffusion model for organometallic catalysts** Diffusion models [11, 12] are generative models that include a *diffusion process* that iteratively corrupts data points towards a known

prior through additive noise, and a generative *denoising process* that approximates the reverse of the diffusion process.

In this work, we directly tailor the equivariant diffusion framework [9] to work with organometallic complexes. For ease of notation, we always consider the centre to be fixed in what follows, both in terms of composition and position. The tailored diffusion process, as it does not corrupt the centre, writes

$$q(\mathcal{C}_t|\mathcal{C}) = \delta\big([\mathcal{C}^{(C)}, \mathcal{C}_t^{(L)}]\big) \cdot \mathcal{N}\big(\mathcal{C}_t^{(L)}|\alpha_t \mathcal{C}^{(L)}, \sigma_t^2 \mathbb{I}\big), \tag{1}$$

for $t = 1, ..., T$ with $T = 1000$, where $\alpha_t \in \mathbb{R}^+$ controls how much signal is retained, $\sigma_t \in \mathbb{R}^+$ the amount of noise added, and $\delta(\cdot)$ is the Dirac delta distribution. We use a variance preserving process—a special case of noising process for which $\alpha_t = \sqrt{1 - \sigma_t^2}$, and model $\alpha_t$ with a polynomial scheduler [9].

To account for the fixed centre, the approximate denoising process is also modified, and writes

$$p_\theta(\mathcal{C}_{t-1}|\mathcal{C}_t) = \delta\big([\mathcal{C}_t^{(C)}, \mathcal{C}_{t-1}^{(L)}]\big) \cdot \mathcal{N}\big(\mathcal{C}_{t-1}^{(L)}|\mu_{\mathcal{C}_{t-1}^{(L)}}, \sigma^2_{\mathcal{C}_{t-1}^{(L)}} \mathbb{I}\big), \tag{2}$$

$$= \delta\big([\mathcal{C}_t^{(C)}, \mathcal{C}_{t-1}^{(L)}]\big) \cdot \mathcal{N}\Big(\mathcal{C}_{t-1}^{(L)}|\frac{\alpha_{t-1}}{\alpha_t}\big(\mathcal{C}_t^{(L)} - \frac{\sigma^2_{t|t-1}}{\sigma_t}\varepsilon_\theta(\mathcal{C}_t, t)\big), \frac{\sigma^2_{t|t-1}\sigma^2_{t-1}}{\sigma_t^2}\mathbb{I}\Big), \tag{3}$$

where $\varepsilon_\theta$ is a denoising neural network trained to predict the noise $\epsilon$ sampled to obtain $\mathcal{C}_t$ from $\mathcal{C}$, and $\sigma^2_{t|t-1} = \sigma_t^2 - \sigma^2_{t-1}(\alpha_t/\alpha_{t-1})^2$.

To ensure that the induced likelihood $p_\theta(\mathcal{C})$ is invariant to rotation and translation, we parametrise our denoising neural network, $\varepsilon_\theta$, with an equivariant MPNN inspired from the PaiNN architecture [17]. The latter is provably more expressive than the original EGNN [18], as it can resolve local angular information [19]. Due to the fixed centre, we are not required to subtract the centre of gravity for ensuring translation invariance.

The training objective is the simplified loss objective [20],

$$L(\theta) = \mathbb{E}_{\substack{\mathcal{C}\sim q(\mathcal{C}) \\ \mathcal{C}_t \sim q(\mathcal{C}_t|\mathcal{C})}} \Big[||\epsilon_{x^{(L)}} - \hat{\epsilon}_{x^{(L)}}||^2 + ||\epsilon_{h^{(L)}} - \hat{\epsilon}_{h^{(L)}}||^2\Big], \tag{4}$$

where $[\hat{\epsilon}_{x^{(L)}}, \hat{\epsilon}_{h^{(L)}}] = \varepsilon_\theta(\mathcal{C}_t, t)$ denotes the output of the denoising network, and $[\epsilon_{x^{(L)}}, \epsilon_{h^{(L)}}] \sim \mathcal{N}(0, \mathbb{I})$ is the noise sampled to form $\mathcal{C}_t$ according to Eq. (1).

To accommodate the changes introduced in the noising and denoising processes, we adapt the original sampling procedure [9] in the following way: (1) we start by fixing the centre, by e.g. drawing it from an empirical distribution over central atoms, (2) sample the number of atoms that will compose the ligands given the centre, $N_L \sim p(N_L|C)$, and (3) execute the standard ancestral sampling procedure by iteratively applying Eq. (3).

**Regressor guidance** Inspired from classifier-guidance [21], regressor-guidance builds on the observation that conditioning on a label $y$ can be done by sampling from $p_{\theta,\phi}(\mathcal{C}_{t-1}|\mathcal{C}_t, y) \propto p_\theta(\mathcal{C}_{t-1}|\mathcal{C}_t)p_\phi(y|\mathcal{C}_{t-1})$, where $p_\theta(\mathcal{C}_{t-1}|\mathcal{C}_t)$ is the unconditional denoising process from Eq. (3), and $p_\phi(y|\mathcal{C}_{t-1})$ is a conditional distribution over properties induced by a property predictor $y_\phi$. Here, we define $p_\phi(y|\mathcal{C}_{t-1})$ in terms of an energy function $f_\phi(y, \mathcal{C}_t) = ||y - y_\phi(\mathcal{C}_t, t)||^2$, such that $p_\phi(y|\mathcal{C}_{t-1}) \propto \exp\big(-f_\phi(y, \mathcal{C}_t)\big)$.

In practice, the conditional denoising process is a corrected version of the unconditional one from Eq. (3),

$$p_{\theta,\phi}(\mathcal{C}_{t-1}|\mathcal{C}_t, y) = \delta\big([\mathcal{C}_t^{(C)}, \mathcal{C}_{t-1}^{(L)}]\big) \cdot \mathcal{N}\big(\mathcal{C}_{t-1}^{(L)}|\mu_{\mathcal{C}_{t-1}^{(L)}} - \sigma_{\mathcal{C}_{t-1}^{(L)}} \nabla_{\mu_{\mathcal{C}_{t-1}^{(L)}}} f_\phi(y, \mu_{\mathcal{C}_{t-1}}), \sigma^2_{\mathcal{C}_{t-1}^{(L)}} \mathbb{I}\big), \tag{5}$$

with $\mu_{\mathcal{C}_{t-1}} = [\mathcal{C}_t^{(C)}, \mu_{\mathcal{C}_{t-1}^{(L)}}]$, where the correction is obtained by evaluating the gradient of $f_\phi$ with respect to $\mu_{\mathcal{C}_{t-1}^{(L)}}$.

The property predictor, $y_\phi$, is parametrised by another equivariant neural network with a similar backbone to that of the diffusion model, followed by a graph-level readout. The model is trained

using the same diffusion process as the generative diffusion model, and its parameters are optimised to minimise

$$L(\phi) = \mathbb{E}_{\substack{(\mathcal{C},y)\sim q(\mathcal{C}) \\ \mathcal{C}_t \sim q(\mathcal{C}_t|\mathcal{C})}} \Big[ ||y - y_\phi(\mathcal{C}_t)||^2 \Big]. \tag{6}$$

**Dataset and Task** We perform our experiments on the Suzuki–Miyaura C–C cross-coupling dataset [22]. The database features 7054 catalysts. Each of them is composed of a metal centre, to which 2 ligands are bound: $\mathrm{L}_1 - \mathrm{M} - \mathrm{L}_2$ ($\mathrm{M} \in \{\mathrm{Ni}, \mathrm{Pd}, \mathrm{Pt}, \mathrm{Cu}, \mathrm{Ag}, \mathrm{Au}\}$). For each complex, a DFT-level optimised geometry is available, along with the reaction energy associated with the oxidative addition of the substrate and the transition metal ($\Delta E_{\text{Rxn A}}$) – depicted by Rxn A in Fig. 2.

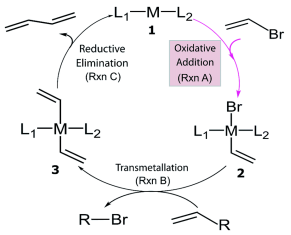


Figure 2: Reaction under study in this work—the model generates $\mathrm{L}_1 - \mathrm{M} - \mathrm{L}_2$, and optimises the reaction energy of Rxn A ($\Delta E_{\text{Rxn A}}$). Figure reproduced from [22].

While full kinetic profiles are usually required for accurate description of catalytic performance, previous work [22, 23] has shown that in the case Suzuki reaction $\Delta E_{\text{Rxn A}}$ could be used as a descriptor for analysing the catalytic cycle thermodynamics with volcano plots. To be relevant, catalysts should display a reaction energy $\Delta E_{\text{Rxn A}}$ between $-32.1$ and $-23.0$ kcal/mol.

## 3 Experiments and Results

**Unconditional Generation** In this experiment, we evaluate and compare the ability of different variants of the generative model to learn the unconditional data distribution. After training each model, we generate 10000 samples, and evaluate their properties in terms of: chemistry (via validity, formula uniqueness and formula novelty) geometry around M (via $W^{\mathrm{L}_{1,2}-\mathrm{M}}$ and $W^{\mathrm{L}_1-\mathrm{M}-\mathrm{L}_2}$), and binding energy (via $W^{\Delta E}$). The details of the evaluation procedure and different baselines are provided in Section 5.1 (see Appendix).

The results are presented in Table 1. In line with previous work [24], we first find that more expressive geometric neural networks, i.e. EDM-PaiNN and OM-EDM-PaiNN, yield noticeably improved validity. Second, our proposed OM-EDM-PaiNN reproduces the training distribution of the geometry around M better, and produces samples that have an aggregated binding energy distribution that is closest to the training distribution. In Fig. 3, we show that distribution for Pd, Pt and Cu.

Table 1: Results of the unconditional sampling experiment. V stands for Validity. UF and NF stand for Unique Formulas and Novel Formulas.

| | V (↑) | UF (↑) | NF (↑) | $W^{\mathrm{L}_{1,2}-\mathrm{M}}$ (↓) | $W^{\mathrm{L}_1-\mathrm{M}-\mathrm{L}_2}$ (↓) | $W^{\Delta E}$ (↓) |
|---|---|---|---|---|---|---|
| EDM | 0.352 | 0.327 | 0.268 | 0.256 | 0.0113 | 0.0044 |
| EDM-PaiNN | 0.594 | 0.493 | 0.389 | 0.203 | 0.0089 | 0.0040 |
| OM-EDM | 0.491 | 0.435 | 0.328 | 0.165 | 0.0084 | 0.0044 |
| **OM-EDM-PaiNN** | **0.603** | 0.470 | 0.337 | **0.147** | **0.0079** | **0.0036** |
| FF | – | – | – | 0.814 | 0.0257 | – |
| Dataset | 0.878 | 0.644 | – | – | – | – |

**Conditional Generation** We now leverage our framework for inverse-designing relevant organometallic catalysts, and study how the distribution of the property of interest is impacted when performing conditional generation.

We first find the modified diffusion introduced in Section 2 to be critical. Without information about the metal-centre, the property predictor $y_\phi$ has a hard time producing meaningful predictions for increasing levels of noise, and can not effectively be used for guidance. All produced samples are unphysical. On the contrary, when using OM-EDM-PaiNN, conditional sampling can effectively be performed and the property distribution can be steered towards the target value, as displayed in Fig. 3.

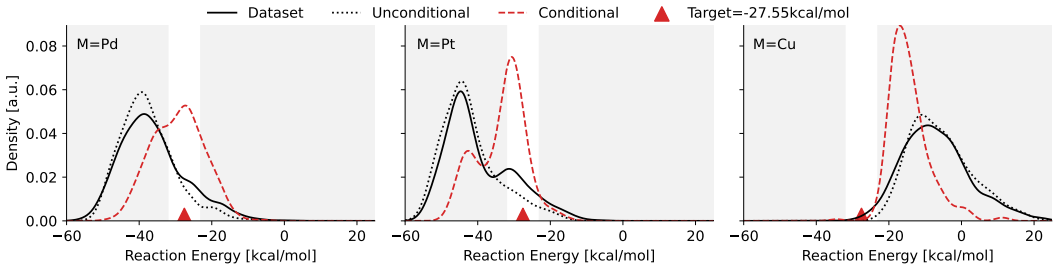


Figure 3: Results of sampling OM-EDM-PAINN for $\mathrm{M} \in \{\mathrm{Pd}, \mathrm{Pt}, \mathrm{Cu}\}$.

We perform further validation by randomly selecting 6 samples (3 Pd and 3 Pt) featuring novel formulas and an estimated binding energy (by an auxiliary surrogate) in the range of interest, and recompute energy values with DFT, using the same protocol as the one used to generate the training database [22]. We display the results in Fig. 4.

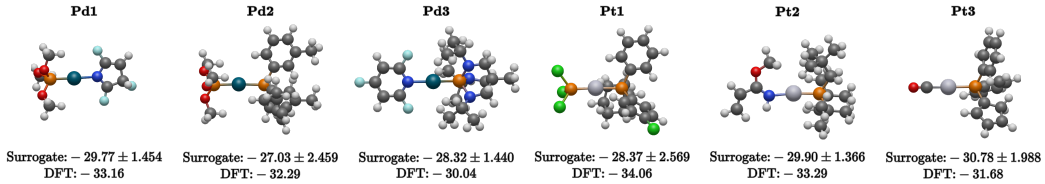


Figure 4: Results of the DFT validation (values in kcal/mol). For the predictions, we also provide the standard deviation across an ensemble of 10 surrogates. The target value for conditional generation was $-27.55$ kcal/mol, mid value of the range of interest ($[-32.1, -23.0]$ kcal/mol). For each metal centre, we randomly selected 3 conditional samples that were deemed valid and featured a novel formula.

## 4 Conclusion

In this work, we have introduced a guided equivariant diffusion model specifically designed for the generation of organometallic complexes with targeted properties. Our preliminary results showed that the increased expressivity of the denoising neural network combined with a proper modelling of the centre enables controllable property optimisation. We checked 6 novel conditional samples with DFT, and found them to have a binding energy in relatively good agreement with their predicted value (within 5 kcal/mol at most). Our future efforts will focus on further validating the proposed method. This includes a thorough analysis of the generated compounds in terms of coverage and diversity, along with additional DFT calculations.

Avenues for future work are numerous. Pre-training the generative diffusion model on large databases, or alternative conditional sampling methods are natural options. Explicit modelling of the binding point by joint generation of the catalyst and substrate might also be worth exploring.

## References


[1] Jessica G Freeze, H Ray Kelly, and Victor S Batista. Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. *Chemical reviews*, 119(11):6595–6612, 2019.

[2] Marco Foscato and Vidar R Jensen. Automated in silico design of homogeneous catalysts. *ACS catalysis*, 10(3):2354–2377, 2020.

[3] Wenhao Gao and Connor W Coley. The synthesizability of molecules proposed by generative models. *Journal of chemical information and modeling*, 60(12):5714–5723, 2020.

[4] John J Irwin and Brian K Shoichet. Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[5] Maho Nakata and Tomomi Shimazaki. Pubchemqc project: a large-scale first-principles electronic structure database for data-driven chemistry. *Journal of chemical information and modeling*, 57(6):1300–1308, 2017.

[6] David Balcells and Bastian Bjerkem Skjelstad. tmqm dataset—quantum geometries and properties of 86k transition metal complexes. *Journal of chemical information and modeling*, 60(12):6135–6146, 2020.

[7] Oliver Schilter, Alain Vaucher, Philippe Schwaller, and Teodoro Laino. Designing catalysts with deep generative models and computational data. a case study for suzuki cross coupling reactions. *Digital Discovery*, 2(3):728–735, 2023.

[8] Niklas Gebauer, Michael Gastegger, and Kristof Schütt. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. *Advances in neural information processing systems*, 32, 2019.

[9] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

[10] Niklas WA Gebauer, Michael Gastegger, Stefaan SP Hessmann, Klaus-Robert Müller, and Kristof T Schütt. Inverse design of 3d molecular structures with conditional generative neural networks. *Nature communications*, 13(1):973, 2022.

[11] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[13] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2021.

[14] Ilia Igashov, Hannes Stärk, Clément Vignac, Victor Garcia Satorras, Pascal Frossard, Max Welling, Michael Bronstein, and Bruno Correia. Equivariant 3d-conditional diffusion models for molecular linker design. *arXiv preprint arXiv:2210.05274*, 2022.

[15] Arne Schneuing, Yuanqi Du, Charles Harris, Arian Jamasb, Ilia Igashov, Weitao Du, Tom Blundell, Pietro Lió, Carla Gomes, Max Welling, et al. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2022.

[16] Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. In *The Eleventh International Conference on Learning Representations*, 2022.

[17] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

[18] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[19] Chaitanya K. Joshi, Cristian Bodnar, Simon V Mathis, Taco Cohen, and Pietro Lio. On the expressive power of geometric graph neural networks, 2023.

[20] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

[21] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[22] Benjamin Meyer, Boodsarin Sawatlon, Stefan Heinen, O Anatole Von Lilienfeld, and Clémence Corminboeuf. Machine learning meets volcano plots: computational discovery of cross-coupling catalysts. *Chemical science*, 9(35):7069–7077, 2018.

[23] Michael Busch, Matthew D Wodrich, and Clémence Corminboeuf. Linear scaling relationships and volcano plots in homogeneous catalysis–revisiting the suzuki reaction. *Chemical science*, 6(12):6754–6761, 2015.

[24] Alex Morehead and Jianlin Cheng. Geometry-complete diffusion for 3d molecule generation. In *ICLR 2023 - Machine Learning for Drug Discovery workshop*, 2023.

[25] Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8:31, 2013.

# 5 Supplementary Material

## 5.1 Evaluation details

### 5.1.1 Validity, Uniqueness, Novelty

All the presented metrics are proportions of the generated samples.

**Validity** A generated complex has to pass a series of checks to be deemed valid:

1. **(one TM check)** It has to have **exactly** one transition metal atom;
2. **(distance check)** All pairwise distances should be at least $0.9$Å, and no atom can be disconnected from the rest of the complex (i.e. its closest neighbour is located at distance larger than the cutoff of $3.0$Å);
3. **(RDKit check)** The ligands, i.e. complex where the TM has been removed, have to be valid according to RDKit [25].

As the algorithm implemented in RDKit to determine bonds can not handle transition metals, we proceed as follows: we remove the metal centre, and we then try to find a feasible bond allocation using `rdDetermineBonds.DetermineBonds`. If the allocation succeeds, the sample is deemed valid. As the removal of the metal centre can introduce local charges, we apply `rdDetermineBonds.DetermineBonds` for different charges until one matches. If none matches, the configuration is deemed invalid.

The validation method is not perfect, as only around $88\%$ of the training database is deemed valid by our algorithm (Table 1).

**Uniqueness and Novelty** As bonding is not properly defined for transition metal complexes, we study uniqueness and novelty in terms of chemical formulas. This does not provide the full picture as two identical formulas can correspond to different complexes. However, when encountering new formulas, we are ensured that the corresponding complexes are novel. Uniqueness and Novelty are defined as follows,

$$\text{UF} = \frac{\text{\# (valid and unique formulas)}}{\text{\# samples}}, \tag{7}$$

$$\text{NF} = \frac{\text{\# (valid, unique and novel formulas)}}{\text{\# samples}}. \tag{8}$$

### 5.1.2 Geometry and Binding Energy

Given the importance of the direct neighbourhood of the centre, we assess the geometry of centre and the two proximal atoms by comparing the empirical distribution of the $\text{L}_{1,2} - \text{M}$ distances and the $\text{L}_1 - \text{M} - \text{L}_2$ angle. Similarly, we compare the training distribution of binding energy with the distribution induced by the generated samples.

We measure the discrepancy between training distributions and distributions induced by the generated samples using the 1-Wassertein distance. If $P_z$ denotes the empirical measure for centre $z \in \mathcal{Z}$ across the dataset, and $Q_z$ denotes the empirical measure the same centre across the samples generated by the diffusion model, the distance between the two empirical distributions is given by

$$W(P_z, Q_z) = \big(\frac{1}{n}\sum_{i=1}^{n} ||X_{(i)} - Y_{(i)}||\big), \tag{9}$$

where $X_{(i)}$ and $Y_{(i)}$ denote samples from $P_z$ and $Q_z$ respectively.

To obtain an aggregated distance value, we compute a weighted sum over the different metal-centres,

$$W(P, Q) = \sum_{z \in \mathcal{Z}} p(z) W(P_z, Q_z), \tag{10}$$

where $p(z)$ denotes the empirical categorical distribution over the metal centre obtained from the training data.

### 5.1.3 Baselines

Our method, coined **OM-EDM-PaiNN** implements Eqs. (1) and (3), and the more expressive denoising neural network inspired from the PaiNN architecture [17]. We compare it with 3 different baselines:

- **EDM**: that reimplements the vanilla equivariant denoising diffusion [9];
- **EDM-PaiNN**: that reimplements the vanilla equivariant denoising diffusion with a more expressive denoising neural network identical to that of **OM-EDM-PaiNN**;
- **OM-EDM**: that implements Eqs. (1) and (3), but uses EGNN [18] as denosing neural network.

We additionally include a baseline based on geometry-free representations, where RDKit, or any cheap force-field, is used to generate geometries. We denote that method by FF. We compute the geometry statistics from the additional $18064$ force-field-level data points released along with the DFT-level data [22].