CompGen: A Conditional Generation Framework for Inverse Composition Design of Catalytic Surfaces

Shuizhou Chen[1]* **Chenghan Sun**[2]* **Zhiyuan Liu**[1] **Andi Han**[3,4]†

Ichigaku Takigawa[2,4,5]† **Quan Qian**[1]†

[1]School of Computer Engineering & Science, Shanghai University
[2]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University
[3]School of Mathematics and Statistics, University of Sydney
[4]RIKEN, Center for Advanced Intelligence Project
[5]Department of Computational Biology and Medical Sciences,Graduate School of Frontier Sciences, The University of Tokyo
{shuizhou,zhiyuan88,qqian}@shu.edu.cn, chsun@icredd.hokudai.ac.jp, andi.han@sydney.edu.au, takigawa@k.u-tokyo.ac.jp

Abstract

Generating adsorption configurations, that is, how small atoms or molecules bind to complex catalyst surfaces, remains underexplored in inverse materials design. We present CompGen, a conditional generative framework that reformulates 3D structure prediction as a 2D shell-wise composition task centered on the adsorption site. CompGen uses a Chemically Informed Autoencoder (CIAE) to embed sparse compositions into a continuous, periodic table aware latent space learned with a multi-stage pretraining process. A conditional diffusion model then samples in this latent space under multi-physical conditions, including adsorbate identity, adsorption energy, and relevant elements, enabling inverse composition design of catalytic surfaces. Pretrained on a subset of Open Catalyst 2020, CompGen is fine-tuned to more complex high-entropy alloy (HEA) surfaces and achieves strong fine-tuned performance. Extensive experiments show robust zero-shot and few-shot behavior, highlighting CompGen's effectiveness for data-efficient, domain-transferable inverse design of catalytic surfaces.

1 Introduction

Inverse design of catalytic materials is crucial for advancing energy storage technologies and promoting environmental sustainability, representing an important challenge within the AI-for-Science domain (Seh et al., 2017; Freeze et al., 2019; Zitnick et al., 2020; Noh et al., 2020; Wang et al., 2023). At the core of catalyst design is the accurate identification and optimization of active sites, which refer to the localized regions on catalyst surfaces where chemical reactions occur (Vogt and Weckhuysen, 2022). These active sites determine how efficiently a catalyst accelerates specific reactions.

Despite recent developments in generative models for accelerating inverse materials design conditioned on desired physical or chemical properties (Sanchez-Lengeling and Aspuru-Guzik, 2018; Gebauer et al., 2022; Anstine and Isayev, 2023; Xiao et al., 2023; Zheng et al., 2024; Park et al., 2024), applying such techniques to catalytic surfaces remains challenging due to the uncertainty and

*Equal contribution.
†Corresponding authors

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI4Science.

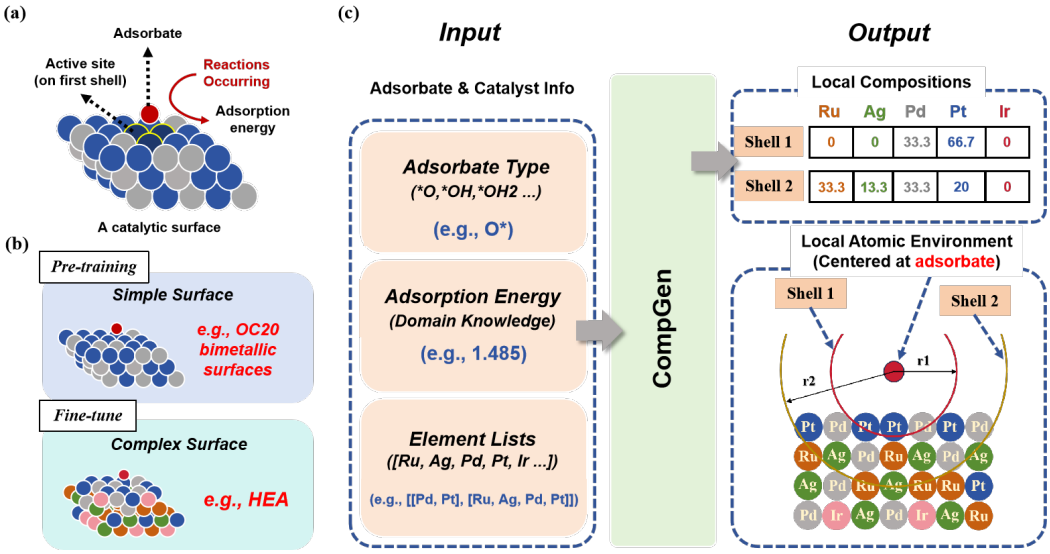

Figure 1: The illustration of (a) An example of adsorption configuration on a catalytic surface. (b) CompGen domain adaptation: pretraining on simple surfaces (OC20) and fine-tuning on compositionally complex HEA surfaces. (c) An example of input and output using CompGen to generate target local compositions.

complexity of local atomic environments around active sites. In particular, it remains an open problem to identify optimal local atomic arrangements of active sites, guided by key reactivity indicators such as adsorption energy (Qin et al., 2020; Wang et al., 2025b). Most existing approaches require to generate explicit 3D coordinates of the local atoms, which however, face challenges of efficiently representing molecule symmetries in translation, rotation, and periodic boundary conditions (Kolluru et al., 2022; Duval et al., 2023; Wang and You, 2025). Recent diffusion-based models (Cornet et al., 2024; Kolluru and Kitchin, 2024) and equivariant PaiNN-based surface generation model (Schütt et al., 2021; Rønne et al., 2024) have made progress by modeling adsorbate placements on surfaces or generating full surface structures. However, they are often limited to relatively simple systems (e.g., silver-oxide surfaces or homogeneous organometallics), narrowly focus on adsorption site prediction, or lack explicit conditional generation strategies. Alternatives based on reinforcement learning or molecular representation learning face challenges from limitations of single-objective learning setup or reliance on domain-specific features (Lacombe et al., 2023; Wang et al., 2025a).

In this work, we reformulate inverse design of catalytic surfaces as a simpler, yet chemically meaningful **composition generation task**. We focus on generating the proportions of each type of elements in two concentric spherical shells around a surface active site: *an inner shell 1* that interacts directly with the adsorbed atoms or molecules (i.e., the adsorbates) and *an outer shell 2* that surrounds the inner shell. This two-shell approach bypasses explicit 3D geometry, but preserves essential ensemble and ligand effects used to tune catalytic properties (Li et al., 2018; Pedersen et al., 2022). *shell 1* controls ensemble feasibility by fixing element counts near the adsorbate, while *shell 2* enables ligand tuning through second shell composition. This setup provides useful priors that shrink the design space and reduce the number of candidate microstates, enabling more efficient high-throughput screening.

Despite the simplified setup, the composition generation task remains challenging due to the sparse, high-dimensional element composition data and the need for chemically guided, property-driven design. To address this, we introduce **CompGen**, a conditional generative framework for inverse composition design of catalytic surfaces. CompGen couples a Chemically Informed Autoencoder (CIAE) with a conditional diffusion model backbone. The CIAE creates a chemically-aware latent space for compositions projection by leveraging a periodic table representation (PTR) as a relational prior (Feng et al., 2021) and a multi-stage pretraining process. The conditional diffusion model then samples from latent space under physical conditions, including categorical labels and numerical targets such as adsorption energy.

CompGen is pretrained on a subset of the Open Catalyst 2020 (OC20) dataset (Chanussot et al., 2021), which spans diverse catalytic surfaces up to three elements. To show adaptability of our pretrained model to more complex surface compositions, we fine-tune the framework on a high-entropy alloy (HEA) dataset (Clausen et al., 2024) with surfaces composed of five different elements and observe significant performance gains. We further evaluate CIAE design, diffusion backbone choices, and conditional generation strategies, and demonstrate strong zero-shot and few-shot generalization. The main contributions of this work are summarized as follows:

- We present CompGen, a conditional generative framework that reformulates 3D inverse catalysts design as a tractable 2D, shell-based composition generation task. To the best of our knowledge, this is the first framework explicitly focused on inverse *composition* design of catalytic surfaces.
- We design CIAE, which transforms element-wise composition profiles into a chemically consistent latent space using a periodic table-aware encoder and a two-stage pretraining strategy.
- We demonstrate CompGen can robustly transfer from simple surfaces to compositionally complex HEAs via parameter-efficient fine-tuning, and validate the model architecture with comprehensive experiments.

2 Related work

Inverse materials design. Generative models have advanced inverse materials design by mapping target properties to novel structures. Early GAN-based methods (Nouira et al., 2018; Kim et al., 2020; Zhao et al., 2021) laid the groundwork, followed by diffusion-based (Hoogeboom et al., 2022; Jiao et al., 2023; Pakornchote et al., 2024) and flow-based approaches (AI4Science et al., 2023) that further improved the task with high-fidelity generalization. Recent representative diffusion models include MatterGen (Zeni et al., 2025), which uses property-conditioned denoising to improve the generation of stable, unique, and novel inorganic materials, and All-atom Diffusion Transformer (ADiT) (Joshi et al., 2025), which maps molecules and crystals into a shared VAE latent space and jointly generates periodic and non-periodic atomic structures via a latent-space diffusion Transformer.

Inverse catalysts design. Inverse design of catalytic surfaces raises distinct challenges beyond bulk generation. A range of paradigms has been explored: reinforcement learning, language-model–based generation, and diffusion models. AdsorbRL (Lacombe et al., 2023) employs a Deep Q-Network (Mnih et al., 2015) to navigate vast compositional spaces and identify catalysts optimized with given adsorption energy targets. CatGPT (Mok and Back, 2024) leverages a GPT-2-based language model (Radford et al., 2019) to generate string-based representations of catalyst surfaces, enabling fine-tuning for downstream property prediction tasks validated by Density Functional Theory (DFT) (Kohn and Sham, 1965) calculations. Among diffusion models, Rønne et al. (Rønne et al., 2024) propose a rotationally equivariant diffusion framework with force-field guidance to sample low-energy silver-oxide surface structures, though without explicit conditioning on target properties. OM-Diff (Cornet et al., 2024) introduces a guided equivariant diffusion model that generates 3D structures of homogeneous organometallic catalysts, conditioned on specific metal centers via regressor-informed denoising. AdsorbDiff (Kolluru and Kitchin, 2024) predicts the optimal adsorbate binding orientations and placements on catalytic surfaces. However, it focuses on improving placement success rates rather than generating the local surface environment of active sites. Complementary to these generative efforts, structure-search workflows have been proposed for inverse catalysts such as metal-oxide interfaces (Kempen and Andersen, 2025). An thermodynamics-guided search over Zn_yO_x and In_yO_x clusters on pure metal surfaces uncovers stable active site motifs and highlights the importance of site diversity in design. Most recently, PGH-VAE (Wang et al., 2025a) applies a topology-aware VAE with features from topological algebraic analysis to inverse design of active sites on IrPdPtRhRu HEA surfaces.

3 Problem Setup

We cast inverse catalytic surfaces design task as generating two shell-wise compositions under given conditions. We define the local neighborhood of an adsorption site by two concentric shells of surface

atoms around the adsorbate's *central* (binding) atom[3]. (1) **First shell** $\mathcal{C}_1$: surface atoms within 2.5 Å cutoff radius of the central atom, i.e., its nearest neighbors by distance. (2) **Second shell** $\mathcal{C}_2$: surface atoms within 5.0 Å of the central atom that are the nearest neighbors of the first shell atoms.

For each shell $k \in \{1, 2\}$, we use a composition vector $\mathbf{c}^k \in \mathbb{R}^D$ over $D = 118$ chemical elements from the periodic table ordered by atomic number. The i-th entry $c_i^k \in [0, 1]$ denotes the normalized atomic percentage of the i-th element within shell k, with $\sum_{i=1}^{D} c_i^k = 1$. We then stack the two shell compositions as $\mathbf{C} = [\mathbf{c}^1, \mathbf{c}^2] \in \mathbb{R}^{2\times D}$ to be the generation target of CompGen framework.

We conditionally generate on three practical descriptors of an arbitrary pair of adsorbate-surface configuration: the adsorbate S, the adsorption energy E, and an element list L that specifies the allowed element types in each shell. In standard computational workflows for catalyst screening (Nørskov et al., 2009; Schlexer Lamoureux et al., 2019), S is fixed by the target reaction and denotes the atomic or molecular species that binds to the surface; E is a quantitative measure of binding strength between the adsorbate and the surface and is a widely used reactivity descriptor and prediction target (Ghanekar et al., 2022; Ock et al., 2024); and L defines the per-shell set of allowed elements, i.e., the composition design space for generation. In our simplified setting, these three descriptors compactly define the adsorption model as presented in Figure 1.

Taken together, formally, we aim to learn the conditional distribution of the surface composition $p(\mathcal{C}_1, \mathcal{C}_2|S, E, L)$, so that we can sample novel, constraint-consistent shell compositions. Our proposed CompGen framework reduces the search complexity by at least a polynomial factor in the number of sites. See more details in Appendix C.

4 Method

As illustrated in Figure 2, our CompGen framework follows the design of latent diffusion model (Rombach et al., 2022), which consists of two core modules: CIAE and a conditional latent diffusion model (CLDM). CIAE places the two-shell composition vectors on a periodic table grid and encodes them with a convolutional autoencoder to obtain a compact, reconstructable latent. CLDM then learns the conditional distribution over these latents and samples new compositions given the specified, mainly using U-Net (Ronneberger et al., 2015) and Diffusion Transformer (DiT) (Peebles and Xie, 2023).

4.1 Chemically Informed Autoencoder (CIAE)

Vanilla autoencoders learn generic latents and are largely blind to chemistry-specific semantics. We design the CIAE by imposing chemical prior information in two ways: (i) it represents each shell composition on a periodic table grid as inspired by prior works (Zheng et al., 2018; Feng et al., 2021), and (ii) it uses a two-stage pretraining scheme that first transfers property-aware features into the encoder and then trains the full autoencoder for reconstruction. This design exploits two simple facts: the periodic table's layout encodes element periods and groups, providing a physically meaningful spatial inductive bias; and staged pretraining yields a latent space that is easy to map into and reconstruct from while preserving chemically relevant relationships.

Before applying the two-stage training strategy to obtain a compact latent space that preserves core chemical regularities, we pre-process each shell composition $\mathbf{c}^k \in \mathbb{R}^D$ by mapping it to a 2D grid representation $\mathbf{P}$ that encodes the periodic table's layout. Formally, the entry at position (h, w) for shell k is defined as:

$$(\mathbf{P}_k)_{h,w} = \begin{cases} c_i^k & \text{if cell } (h, w) \text{ corresponds to element } i \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{P}_k$ is called the PTR for shell k. Cells for absent elements and empty positions are zero-padded. We then stack the two shell PTRs to form a two-channel input tensor $\mathbf{x}_0 = [\mathbf{P}_1; \mathbf{P}_2] \in R^{2\times H\times W}$. Construction details for the PTR are provided in Appendix A.

[3]The *central atom of the adsorbate* is the atom that forms the primary bond to the surface, e.g., `O` in *`OH`, `C` in *`CH3`, and `N` in *`NH3`.

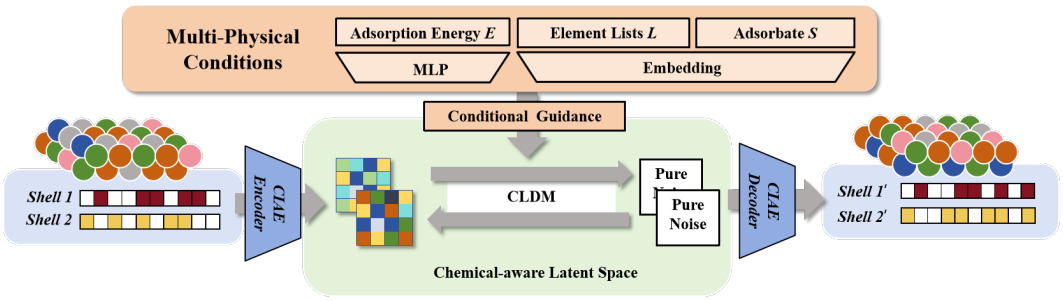

Figure 2: CompGen framework for multi-physical conditional composition generation over chemical-aware latent space with CIAE and CLDM.

Stage I: Supervised Encoder Pretraining. To inject property awareness into the latent space, we adopt the General and Transferable Deep Learning (GTDL) framework (Feng et al., 2021) by using a VGG-like CNN backbone pretrained on the Materials Project dataset (Jain et al., 2013) for a binary classification task to serve as our encoder f_{enc}. It maps $\mathbf{x_0}$ to a latent vector $\mathbf{z} = f_{\text{enc}}(\mathbf{x_0}; \theta_{\text{enc}}) \in \mathbb{R}^{2\times d_z \times d_z}$. However, unlike the original framework, we do not augment PTRs with handcrafted features, as our objective is only for composition reconstruction.

Stage II: Pretraining the Autoencoder With the stage I encoder as initialization, we train the full autoencoder end-to-end for high-fidelity reconstruction of shell-wise compositions. As mentioned in stage I, the stacked PTR tensor $\mathbf{x_0} \in \mathbb{R}^{2\times H\times W}$ is encoded to a low-dimensional latent vector $\mathbf{z}$. In stage II, a decoder f_{dec} maps $\mathbf{z}$ back to the composition space, producing $\hat{\mathbf{C}} = f_{\text{dec}}(\mathbf{z}; \theta_{\text{dec}}) \in \mathbb{R}^{2\times D}$, i.e., per-shell elemental fractions rather than PTR pixels. The decoder uses a convolutional block with a residual connection followed by fully connected layers. The training objective is to minimize the reconstruction error between the original composition tensor $\mathbf{C}$ and the reconstructed tensor $\hat{\mathbf{C}}$. We optimize $(\theta_{\text{enc}}, \theta_{\text{dec}})$ with a *Mean Squared Error* (MSE) loss between the ground-truth compositions $\mathbf{C}$ and reconstructions $\hat{\mathbf{C}}$, averaged over shells and elements.

4.2 Conditional Latent Diffusion Model

The chemically informed latent space learned by CIAE in pretraining stages provides the foundation for the generation task. Given a latent embedding $\mathbf{z}$ from the CIAE encoder, we train a conditional diffusion model to learn the distribution of $\mathbf{z}$ conditioned on:

- **Adsorbate S.** We encode S as a one-hot vector of dimension K_s ($K_s = 13$ in our experiments, covering 13 types of the most basic adsorbates such as *O, *CH, *NH2) from the OC20 dataset. A learned embedding layer maps this to $\mathbf{s} \in \mathbb{R}^{d_h}$.
- **Adsorption energy E.** The scalar target energy (in unit of eV) is repeated to length K_s for the stability of training and then passed through a MLP to produce $\mathbf{e} \in \mathbb{R}^{d_h}$.
- **Element list L.** The allowed element types (e.g., [Ru, Pt, Pd, Ag, Ir]) of each shell are embedded to yield $\mathbf{l} \in \mathbb{R}^{d_h}$.

Latent Diffusion. Let $\mathbf{z}_0 = \mathbf{z}$ denote the latent representation produced by the pretrained CIAE, we implement the standard CLDM process. The forward process corrupts latent $\mathbf{z}_0$ into $\mathbf{z}_t$ over discrete timesteps $t \in \{0, ..., T\}$ by gradually adding Gaussian noise. T is chosen large enough that z_T is approximately standard normal. A neural network ϵ_θ is trained to run the reverse process, denoising from $t = T$ to $t = 0$ by predicting the additive noise.

The training objective minimizes the expected MSE between the true noise and the network's prediction over noisy latents $\mathbf{z}_t$:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{z}_0, \epsilon\sim\mathcal{N}(0,I), t}\left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l})\|^2\right].$$

where ϵ_θ is conditioned on $(\mathbf{s}, \mathbf{e}, \mathbf{l})$.

5 Experiment

Datasets Two datasets are considered during the pretraining and fine-tuning stages of CompGen. For pretraining, we utilize a subset of the per-adsorbate trajectories from the OC20 dataset[4]. Specifically, we extract trajectories involving 13 common adsorbates (*O, *H, *OH, *OH2, *C, *CH, *CH2, *CH3, *CH4, *N, *NH, *NH2, and *NH3)[5], with well-defined central atoms as reference to calculate the distance to the surface atoms for each shell. These adsorbates are adsorbed on distinct surfaces, each composed of at most three different element types. In total, 52 unique surface elements are presented, and the final dataset consists of 131,740 adsorption configurations. We first randomly split the dataset with 5% as test set, and further split the remaining 95% of the dataset into training and validation set with ratio of 90:10.

HEAs have recently gained attention as promising catalysts due to their highly diverse local structural composition, which enable fine-tuning of catalytic properties across a broad design space. To evaluate CompGen's fine-tuning capabilities, we adopt the HEA dataset from Clausen et al. (Clausen et al., 2024), which includes 4,892 distinct adsorption configurations of *OH and *O on HEA surfaces composed of five elements: Ag, Ir, Pd, Pt, and Ru. The same pre-processing protocol is applied as in the pretraining dataset by extracting normalized first and second shell element compositions. We follow the original data split as used in the original paper, keeping the same 80:10:10 ratio for the fine-tuning experiments of CompGen.

Metrics We evaluate the performance of CompGen with three complementary metrics: (i) *Fréchet Distance* (FD) for distributional similarity of generated CIAE latent space, (ii) *Leakage* for compliance with compositional constraints, and (iii) *Cosine Similarity* for point-wise measure of quality of the final generated two-shell composition.

- **Fréchet Distance (FD).** Motivated by image generation evaluation (Heusel et al., 2017), we adopt the Fréchet distance for comparing the generated distribution and input distribution of surface compositions. We use CIAE encoder to project both real (X_r) and generated (X_g) compositions into the same chemical-aware latent space. The FD is then calculated based on the means ($\boldsymbol{\mu}_r, \boldsymbol{\mu}_g$) and covariances ($\boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_g$) of the latent embeddings:

$$\text{FD} = ||\boldsymbol{\mu}_r - \boldsymbol{\mu}_g||_2^2 + \text{Tr}\left(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_g - 2(\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_g)^{\frac{1}{2}}\right)$$

  Lower FD scores indicate a closer match between the real and generated distributions.

- **Leakage.** For a ground truth composition $\mathbf{c}$, define its support (allowed set of elements) as $S = \{i \mid c_i > 0\}$. Given a generated composition $\hat{\mathbf{c}}$, leakage quantifies the sum of all proportions fall outside the support in the disallowed set S^c. For a batch of N samples, we define:

$$\text{Leakage} = \frac{1}{N}\sum_{j=1}^{N}\sum_{i\in S_j^c}\hat{c}_i^{(j)}$$

  where S_j^c is the disallowed set for the j-th ground truth sample $\mathbf{c}^{(j)}$. Lower values indicate better consistency of generated compositions.

- **Cosine Similarity (Sim).** It provides a direct point-wise measure of generation quality by evaluating the alignment between a pair of ground truth vector $\mathbf{c}$ and generated vector $\hat{\mathbf{c}}$. Sim computes the average cosine similarity over N samples:

$$\text{Sim} = \frac{1}{N}\sum_{j=1}^{N}\frac{\mathbf{c}^{(j)}\cdot\hat{\mathbf{c}}^{(j)}}{||\mathbf{c}^{(j)}||_2||\hat{\mathbf{c}}^{(j)}||_2}$$

  where the $\cdot$ sign denotes the dot product. Higher values indicate higher similarity.

Baseline To quantify the contribution of CIAE and the overall CompGen framework, we construct a chemically agnostic baseline that bypasses the CIAE and feeds an unstructured, image-based representatio directly to the DiT model. Starting from a composition vector $\mathbf{c} \in \mathbb{R}^{118}$, we zero-pad it

[4]The OC20 dataset is available at https://fair-chem.github.io/catalysts/datasets/oc20.html

[5]The * sign indicates that the atom or molecule is bound to the surface of the catalyst.

Table 1: CIAE pretraining reconstruction performance. The evaluation includes test sets from both the pretraining domain dataset (OC20) and the out-of-domain dataset (HEA). Metrics reported are **Mean Square Error (MSE)** and **Cosine Similarity (Sim)**.

Dataset	**Avg. MSE** ↓	**Avg. Sim (shell 1)** ↑	**Avg. Sim (shell 2)** ↑	**Avg. Sim (all shells)** ↑
OC20 test set	3.4974	0.8767	0.9935	0.9351
HEA test set	5.1630	0.9785	0.9471	0.9628

to an augmented vector $\mathbf{c}' \in \mathbb{R}^{120}$ for reshaping compatibility, where $\mathbf{c}' = \mathrm{pad}(\mathbf{c}; (0, 2))$; $\mathbf{c}'$ is then reshaped into a matrix $\mathbf{M} = reshape(\mathbf{c}', (15, 8)) \in \mathbb{R}^{15\times 8}$, mapping the elemental features into a rudimentary spatial grid devoid of any embedded chemical priors. Finally, to create a standardized input, this matrix $\mathbf{M}$ is padded into a larger zero-tensor $\mathbf{X} \in \mathbb{R}^{48\times 48}$, which is defined as:

$$\mathbf{X}_{i,j} = \begin{cases} M_{i,j} & \text{if } 1 \le i \le 15 \text{ and } 1 \le j \le 8 \\ 0 & \text{otherwise} \end{cases}$$

This final tensor X serves as the direct input to the diffusion model. From this baseline experiment, any performance gap relative to CompGen thus isolated from the CIAE's structured latent space.

CIAE Pretraining We pretrain the CIAE on 95% of the selected OC20 subset to learn a dense, chemically structured latent space from shell-wise compositions. We report (i) MSE of latent reconstruction and (ii) point-wise Cosine Similarity between reconstructed and ground truth 118-dimensional composition vectors for shell 1 and shell 2. Evaluation is performed on held-out OC20 and out-of-domain HEA test sets. Results are presented in Table 1.

These results show that CIAE can efficiently map the sparse inputs to the chemically consistent latent space and achieve low reconstruction error and high alignment with target elemental compositions across both shells, with reasonable transferability to unseen HEA dataset.

OC20 Pretraining and Fine-tune on HEA We pretrain CompGen on the OC20 subset to evaluate both U-Net and DiT backbones with the given three conditions (S, E, and L) in the chemically informed latent space. Model performance is assessed using three metrics: FD, Leakage, and point-wise Cosine Similarity (Sim), across the baseline model, U-Net, and DiT diffusion backbones, as summarized in Table 2.

To assess the transferability and adaptability of CompGen, we fine-tune the model on the HEA dataset whose compositional and structural characteristics differ markedly from OC20. We consider two strategies: (i) *Low-Rank Adaptation* (LoRA) (Hu et al., 2022), which adds trainable low-rank adapters to a frozen backbone (here, DiT or U-Net) for parameter-efficient adaptation; and (ii) *full-parameter fine-tuning*, which updates all model weights and serves as an upper bound on adaptation performance. As in pretraining experiments, we report metrics of FD, Leakage, and Cosine Similarity (Sim) of both shells for the fine-tuning regime appear in Table 2.

From Table 2 we draw four conclusions. **(i)** The value of a chemistry-aware latent space is clear: the chemically agnostic baseline fails to learn a meaningful distribution (e.g., FD > 137), whereas our pretrained models achieve orders of magnitude better FD (e.g., U-Net with FD of 0.4753). **(ii)** Both backbones exhibit notable zero-shot transferability to HEA, and the U-Net even improves on shell 1 metrics without any fine-tuning with FD of 0.4276 and Sim score of 0.9271. **(iii)** Fine-tuning consistently improves target domain performance: full-parameter fine-tuning of the U-Net yields the best FD on both shells (shell 1: 0.2297, shell 2: 2.4086) and the highest shell 1 Sim score of 0.9549. **(iv)** LoRA offers a strong parameter-efficient alternative approach, often matching or exceeding the performance of full-parameter fine-tuning. For example, with DiT on the harder shell 2 composition generation task, LoRA yields the highest Sim score (0.6497) compared to full-parameter fine-tuning (0.5303). Finally, to mitigate ambiguity in the one-to-many mapping between conditions and compositions, a promising extension is to pre-align their latents using CLIP-style contrastive learning (Radford et al., 2021) prior to diffusion.

Table 2: Performance comparison of model architecture (**DiT Baseline, CIAE+U-Net vs. CIAE+DiT**) and fine-tuning methods (**Zero-shot vs. Full-parameter Fine-tune vs. LoRA Fine-tune**) for CompGen over metrics of **Fréchet Distance (FD), Leakage, and Cosine Similarity (Sim)**. Results are reported separately on Shell 1 and Shell 2.

Experiments on Shell 1	**OC20**			**HEA**		
	FD ↓	Leakage↓	Sim ↑	FD↓	Leakage↓	Sim ↑
DiT Baseline Pretrain	137.2204	20.4389	0.1051	143.4563	20.4632	0.1043
DiT Baseline + LoRA Fine-tune	–	–	–	121.5242	20.3875	0.1082
U-Net Pretrain	0.4753	0.0292	0.9167	0.4276	0.046	0.9271
U-Net + Full-param Fine-tune	–	–	–	**0.2297**	**0.0343**	**0.9549**
U-Net + LoRA Fine-tune	–	–	–	0.2672	0.0435	0.9427
DiT Pretrain	0.7148	0.1304	0.9037	2.5538	0.1778	0.8988
DiT + Full-param Fine-tune	–	–	–	0.2484	0.0365	0.9535
DiT + LoRA Fine-tune	–	–	–	0.8596	0.1211	0.9336

Experiments on Shell 2	**OC20**			**HEA**		
	FD↓	Leakage↓	Sim ↑	FD↓	Leakage↓	Sim ↑
DiT Baseline Pretrain	26.2408	11.2958	-0.112	43.4831	11.133	-0.1316
DiT Baseline + LoRA Fine-tune	–	–	–	46.1376	11.3084	-0.1518
U-Net Pretrain	1.7876	0.206	0.7136	11.7167	0.3441	0.4758
U-Net + Full-param Fine-tune	–	–	–	**2.4086**	0.3078	0.5284
U-Net + LoRA Fine-tune	–	–	–	2.7984	0.3117	0.5927
DiT Pretrain	2.3681	0.261	0.8107	4.5157	0.3093	0.5627
DiT + Full-param Fine-tune	–	–	–	6.7187	**0.2758**	0.5303
DiT + LoRA Fine-tune	–	–	–	4.5004	0.3805	**0.6497**

6 Conclusion and Future Work

In this work, we cast inverse catalysts design as a two-shell composition generation task and introduce CompGen, a modular framework that couples CIAE with CLDM to supply compositionally actionable priors that collapse the otherwise vast 3D search space. Across multiple metrics, CompGen achieves high generation quality and strong transfer from subset of OC20 surfaces to compositionally complex HEA surfaces via efficient fine-tuning. The framework is architecture-agnostic and can accommodate alternative encoders, generators, and conditioning schemes as the task evolves.

Looking ahead, we plan to redesign the current PTR-initialized VGG encoder with a purpose-built CIAE tailored especially for catalytic surface compositions, exploring novel architectures (e.g., axial attention) and conducting broader ablation studies. For stage I pretraining, we will move beyond generic classification pretraining task toward objectives that better align the latent space with representation of surface compositions. We will also enrich the multi-physical conditions with obtainable surface information related with ensemble and ligand effects, such as the active site motifs, facet labels, and simple geometric details such as distances from the adsorbate center to first shell atoms, in order to develop more expressive embeddings for these inputs. Together, these upgrades aim to yield more uniquely determined generations and a stronger, end-to-end path from target properties to deployable catalyst candidates.

7 Acknowledgment

SC acknowledges support from Advanced Materials - National Science and Technology Major Project (2025ZD0620100). CS and IT acknowledge support from JSPS KAKENHI Grant Number 24K23848 and 25K03174.

References

AI4Science, M., Hernandez-Garcia, A., Duval, A., Volokhova, A., Bengio, Y., Sharma, D., Carrier, P. L., Benabed, Y., Koziarski, M., and Schmidt, V. (2023). Crystal-gfn: sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*.

Anstine, D. M. and Isayev, O. (2023). Generative models as an emerging paradigm in the chemical sciences. *Journal of the American Chemical Society*, 145(16):8736–8750.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., et al. (2021). Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072.

Clausen, C. M., Rossmeisl, J., and Ulissi, Z. W. (2024). Adapting oc20-trained equiformerv2 models for high-entropy materials. *The Journal of Physical Chemistry C*, 128(27):11190–11195.

Cornet, F., Benediktsson, B., Hastrup, B., Schmidt, M. N., and Bhowmik, A. (2024). Om-diff: inverse-design of organometallic catalysts with guided equivariant denoising diffusion. *Digital Discovery*, 3(9):1793–1811.

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Lio, P., Bengio, Y., and Bronstein, M. (2023). A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511*.

Feng, S., Fu, H., Zhou, H., Wu, Y., Lu, Z., and Dong, H. (2021). A general and transferable deep learning framework for predicting phase formation in materials. *npj Computational Materials*, 7(1):10.

Freeze, J. G., Kelly, H. R., and Batista, V. S. (2019). Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. *Chemical reviews*, 119(11):6595–6612.

Gebauer, N. W., Gastegger, M., Hessmann, S. S., Müller, K.-R., and Schütt, K. T. (2022). Inverse design of 3d molecular structures with conditional generative neural networks. *Nature communications*, 13(1):973.

Ghanekar, P. G., Deshpande, S., and Greeley, J. (2022). Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nature Communications*, 13(1):5788.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. (2022). Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. (2022). Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., et al. (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1).

Jiao, R., Huang, W., Lin, P., Han, J., Chen, P., Lu, Y., and Liu, Y. (2023). Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36:17464–17497.

Joshi, C. K., Fu, X., Liao, Y.-L., Gharakhanyan, V., Miller, B. K., Sriram, A., and Ulissi, Z. W. (2025). All-atom diffusion transformers: Unified generative modelling of molecules and materials. *arXiv preprint arXiv:2503.03965*.

Kempen, L. H. and Andersen, M. (2025). Inverse catalysts: tuning the composition and structure of oxide clusters through the metal support. *npj Computational Materials*, 11(1):8.

Kim, S., Noh, J., Gu, G. H., Aspuru-Guzik, A., and Jung, Y. (2020). Generative adversarial networks for crystal structure prediction. *ACS central science*, 6(8):1412–1420.

Kohn, W. and Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133.

Kolluru, A. and Kitchin, J. R. (2024). Adsorbdiff: Adsorbate placement via conditional denoising diffusion. *arXiv preprint arXiv:2405.03962*.

Kolluru, A., Shuaibi, M., Palizhati, A., Shoghi, N., Das, A., Wood, B., Zitnick, C. L., Kitchin, J. R., and Ulissi, Z. W. (2022). Open challenges in developing generalizable large-scale machine-learning models for catalyst discovery. *ACS Catalysis*, 12(14):8572–8581.

Lacombe, R., Hendren, L., and El-Awady, K. (2023). Adsorbrl: Deep multi-objective reinforcement learning for inverse catalysts design. *arXiv preprint arXiv:2312.02308*.

Li, H., Shin, K., and Henkelman, G. (2018). Effects of ensembles, ligand, and strain on adsorbate binding to alloy surfaces. *The Journal of chemical physics*, 149(17).

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

Mok, D. H. and Back, S. (2024). Generative pretrained transformer for heterogeneous catalysts. *Journal of the American Chemical Society*, 146(49):33712–33722.

Noh, J., Gu, G. H., Kim, S., and Jung, Y. (2020). Machine-enabled inverse design of inorganic solid materials: promises and challenges. *Chemical Science*, 11(19):4871–4881.

Nørskov, J. K., Bligaard, T., Rossmeisl, J., and Christensen, C. H. (2009). Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37–46.

Nouira, A., Sokolovska, N., and Crivello, J.-C. (2018). Crystalgan: learning to discover crystallographic structures with generative adversarial networks. *arXiv preprint arXiv:1810.11203*.

Ock, J., Badrinarayanan, S., Magar, R., Antony, A., and Barati Farimani, A. (2024). Multimodal language and graph learning of adsorption configuration in catalysis. *Nature Machine Intelligence*, 6(12):1501–1511.

Pakornchote, T., Choomphon-Anomakhun, N., Arrerut, S., Atthapak, C., Khamkaeo, S., Chotibut, T., and Bovornratanaraks, T. (2024). Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. *Scientific Reports*, 14(1):1275.

Park, H., Li, Z., and Walsh, A. (2024). Has generative artificial intelligence solved inverse materials design? *Matter*, 7(7):2355–2367.

Pedersen, J. K., Clausen, C. M., Skjegstad, L. E. J., and Rossmeisl, J. (2022). A mean-field model for oxygen reduction electrocatalytic activity on high-entropy alloys. *ChemCatChem*, 14(18):e202200699.

Peebles, W. and Xie, S. (2023). Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205.

Qin, R., Liu, K., Wu, Q., and Zheng, N. (2020). Surface coordination chemistry of atomically dispersed metal catalysts. *Chemical Reviews*, 120(21):11810–11899.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 10684–10695.

Rønne, N., Aspuru-Guzik, A., and Hammer, B. (2024). Generative diffusion model for surface structure discovery. *Physical Review B*, 110(23):235427.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Sanchez-Lengeling, B. and Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365.

Schlexer Lamoureux, P., Winther, K. T., Garrido Torres, J. A., Streibel, V., Zhao, M., Bajdich, M., Abild-Pedersen, F., and Bligaard, T. (2019). Machine learning for computational heterogeneous catalysis. *ChemCatChem*, 11(16):3581–3601.

Schütt, K., Unke, O., and Gastegger, M. (2021). Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pages 9377–9388. PMLR.

Seh, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., and Jaramillo, T. F. (2017). Combining theory and experiment in electrocatalysis: Insights into materials design. *Science*, 355(6321):eaad4998.

Vogt, C. and Weckhuysen, B. M. (2022). The concept of active site in heterogeneous catalysis. *Nature Reviews Chemistry*, 6(2):89–111.

Wang, B., Zheng, S., Wu, J., Li, J., and Pan, F. (2025a). Inverse design of catalytic active sites via interpretable topology-based deep generative models. *npj Computational Materials*, 11(1):147.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60.

Wang, Z., Li, W., Wang, S., and Wang, X. (2025b). The future of catalysis: Applying graph neural networks for intelligent catalyst design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 15(2):e70010.

Wang, Z. and You, F. (2025). Leveraging generative models with periodicity-aware, invertible and invariant representations for crystalline materials design. *Nature Computational Science*, pages 1–12.

Xiao, H., Li, R., Shi, X., Chen, Y., Zhu, L., Chen, X., and Wang, L. (2023). An invertible, invariant crystal representation for inverse design of solid-state materials using generative deep learning. *Nature Communications*, 14(1):7027.

Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Wang, Z., Shysheya, A., Crabbé, J., Ueda, S., et al. (2025). A generative model for inorganic materials design. *Nature*, 639(8055):624–632.

Zhao, Y., Al-Fahdi, M., Hu, M., Siriwardane, E. M., Song, Y., Nasiri, A., and Hu, J. (2021). High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Advanced Science*, 8(20):2100566.

Zheng, S., He, J., Liu, C., Shi, Y., Lu, Z., Feng, W., Ju, F., Wang, J., Zhu, J., Min, Y., et al. (2024). Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence*, 6(5):558–567.

Zheng, X., Zheng, P., and Zhang, R.-Z. (2018). Machine learning material properties from the periodic table using convolutional neural networks. *npj Computational Materials*, 4(1):1–7.

Zitnick, C. L., Chanussot, L., Das, A., Goyal, S., Heras-Domingo, J., Ho, C., Hu, W., Lavril, T., Palizhati, A., Riviere, M., et al. (2020). An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*.

A Periodic table representation (PTR) details

Our approach begins with how we represent the atomic environment in the model. Instead of a simple list of atoms, we leverage the inherent structure of the periodic table to create a chemically-aware input representation. The compositional vector for each shell mentioned in Probelm Setup, $\mathbf{c}^k \in \mathbb{R}^D$, is transformed into a 2D grid-like representation. This is achieved by mapping the fractional concentration of each element to a specific location on a 2D grid that mirrors the layout of the periodic table. Specifically, we construct $\mathbf{P}_k \in \mathbb{R}^{H \times W}$, with dimensions corresponding to the periodic table (i.e., $H = 9, W = 18$). Each cell (h, w) in this grid is uniquely assigned to a chemical element based on its position. The value of each cell is then fitted with the fractional concentration of its corresponding element from the composition vector $\mathbf{c}^k$. Formally, the entry at position (h, w) for shell k is defined as:

$$(\mathbf{P}_k)_{h,w} = \begin{cases} c_i^k & \text{if cell } (h, w) \text{ corresponds to element } i \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{P}_k$ is called the PTR for shell k. This representation transforms the compositional data into a grid-like tensor that spatially encodes chemical relationships. For instance, elements in the same group appear in the same column, while elements in the same period appear in the same row. This structure provides a powerful inductive bias, rendering a 2D Convolutional Neural Network (CNN) well-suited architecture for our autoencoder. This allows the model to recognize patterns of chemically similar elements, effectively capturing complex relationships that are missing with non-spatial representation.

B Analysis of CompGen data efficiency for training

To evaluate the data efficiency of our model, we conduct a systematic study using 40%, 60%, and 80% of the training data. These results are compared to a backbone model trained on the full dataset (100%). The model performance is evaluated on both OC20 and HEA tasks, by using MSE between latent embeddings and cosine similarity between shell compositions. Results are shown in Table 3 and figure 3.

Table 3: Performance under varying proportion of training data. Lower MSE and higher similarity values indicate better performance.

Data	MSE (OC20) ↓	OC20 shell 1 ↑	OC20 shell 2 ↑	MSE (HEA) ↓	HEA shell 1 ↑	HEA shell 2 ↑
40%	31.7891	0.8640	0.8207	48.7329	0.8484	0.5634
60%	28.9034	0.8848	0.8288	46.1000	0.8662	0.5715
80%	29.5925	0.8813	0.8343	44.3454	0.9014	0.5776
100%	26.1264	0.9037	0.8107	27.8641	0.8988	0.5627

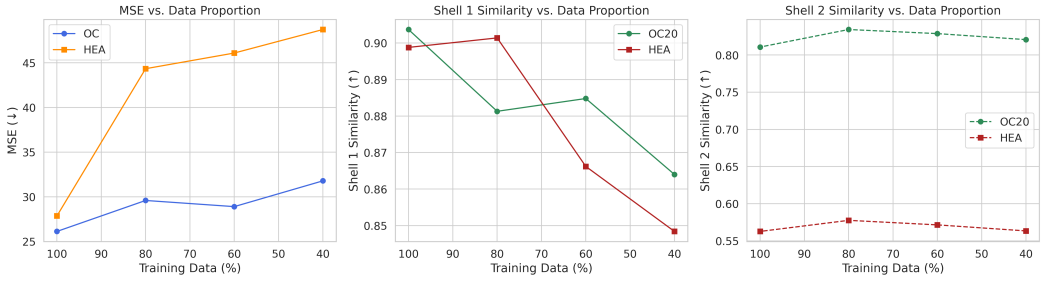

Figure 3: **Data scaling effect across multiple metrics.** Performance of the model under varying proportion of training data (40%, 60%, 80%, 100%). (**Left**) Mean squared error (MSE) for both OC20 and HEA tasks. (**Middle**) Similarity of shell 1 composition across generated structures. (**Right**) Similarity of shell 2 composition. Results demonstrate consistent model improvement with more training data, following typical deep learning scaling laws.

From Table 3, we observe a general trend between training data size and model performance. On the OC20 task, the MSE steadily improves from 31.79 at 40% to 28.90 at 60%, and reaches 29.59 at 80%, approaching the full-data baseline of 26.12. The OC20 shell similarity scores also improve, with OC20 shell 2 increasing from 0.8207 to 0.8343, demonstrating better local structural consistency.

Similarly, on the HEA task, the MSE drops from 48.73 (40%) to 44.34 (80%), and HEA shell 2 similarity grows from 0.5634 to 0.5776. Notably, this slightly exceeds the backbone score (0.5627), suggesting that fine-tuning on a subset can potentially outperform full-data pretraining under certain domain shifts.

Overall, these results are consistent with the expected data scaling behavior in deep models. Model performance improves sublinearly with more data, and even partial training (60 - 80%) is sufficient to recover most of the backbone performance. This demonstrates the data efficiency and generalizability of the proposed method.

C CompGen efficiency analysis

Setup and notation. Fix an adsorption site type (e.g., atop, bridge, hollow) on a specified facet. Let n_1 be the number of *labeled* first shell sites (directly coordinated surface atoms; e.g., n_1=1, 2, 3 for atop/bridge/hollow respectively), and let n_2 be the number of *labeled* second shell sites (near-surface/subsurface neighbors within a fixed cutoff). Let $\mathcal{E}$ be the full element set under consideration (e.g., HEA elements), with $|\mathcal{E}| = N_E$. For each shell $s \in \{1, 2\}$, let $\mathcal{A}_s \subseteq \mathcal{E}$ denote the *allowed element set* for that shell (provided by design rules or domain constraints), with size $|\mathcal{A}_s| = N_s$. A *microstate* is a distinct assignment of elements to the n_s labeled sites of shell s (and likewise for the other shell); we count microstates for the two shells jointly by multiplication.

We consider three levels of constraints:

1. **Unconstrained (U):** Any site may take any element in $\mathcal{E}$.
2. **Subset-constrained (S):** Shell s may take only elements in $\mathcal{A}_s$ (no composition counts enforced).
3. **CompGen-constrained (C):** In addition to $\mathcal{A}_s$, shell s has a *composition count vector* $\mathbf{K}_s = (K_{s,i})_{i\in\mathcal{A}_s}$ with $K_{s,i} \in \mathbb{Z}_{\geq 0}$ and $\sum_{i\in\mathcal{A}_s} K_{s,i} = n_s$.

Unless stated, we treat sites as labeled.[6]

Case U: Unconstrained. Each of the n_1 first shell sites and n_2 second shell sites may independently take any of the N_E elements:

$$\#\mathcal{M}_{\mathrm{U}} \;=\; N_E^{n_1+n_2}. \tag{1}$$

Case S: Subset-constrained by shell. Restricting to $\mathcal{A}_1$ and $\mathcal{A}_2$ yields

$$\#\mathcal{M}_{\mathrm{S}} \;=\; N_1^{n_1}\, N_2^{n_2}, \tag{2}$$

which is a factor of $\left(\frac{N_1}{N_E}\right)^{n_1}\left(\frac{N_2}{N_E}\right)^{n_2}$ reduction relative to (1).

Case C: CompGen composition-constrained. Given count vectors $\mathbf{K}_1$ and $\mathbf{K}_2$, the number of assignments for each shell is a multinomial coefficient:

$$\#\mathcal{M}_{\mathrm{C}} \;=\; \underbrace{\frac{n_1!}{\prod_{i\in\mathcal{A}_1} K_{1,i}!}}_{\text{first shell}} \times \underbrace{\frac{n_2!}{\prod_{i\in\mathcal{A}_2} K_{2,i}!}}_{\text{second shell}}. \tag{3}$$

This follows from counting the permutations of n_s labeled sites subject to exact element counts $K_{s,i}$ in shell s.

[6] Labeled sites reflect distinct geometric positions around the adsorption center (e.g., the three specific metal atoms forming an fcc/hcp hollow). If certain positions are symmetry-equivalent, one may divide by the symmetry group size to obtain a reduced count; our formulas give an upper bound that is sufficient to show the reduction factors.

Reduction factors. Equations (1), (2), (3) imply the hierarchy

$$\#\mathcal{M}_{\mathrm{C}} \;\le\; \#\mathcal{M}_{\mathrm{S}} \;\le\; \#\mathcal{M}_{\mathrm{U}}.$$

In particular,

$$\frac{\#\mathcal{M}_{\mathrm{C}}}{\#\mathcal{M}_{\mathrm{S}}} \;=\; \frac{n_1!}{\prod_{i\in\mathcal{A}_1} K_{1,i}!}\,\frac{n_2!}{\prod_{i\in\mathcal{A}_2} K_{2,i}!}\cdot\frac{1}{N_1^{n_1}N_2^{n_2}}, \qquad \frac{\#\mathcal{M}_{\mathrm{C}}}{\#\mathcal{M}_{\mathrm{U}}} \;=\; \frac{n_1!}{\prod_{i} K_{1,i}!}\,\frac{n_2!}{\prod_{i} K_{2,i}!}\cdot\frac{1}{N_E^{n_1+n_2}},$$

showing that CompGen yields exponential-in-n_s reductions when n_s is moderate (with exact factors governed by the multinomial denominators).

Concrete HEA example with motif-aware n_1. Consider a quinary HEA with $\mathcal{E} = \{\mathrm{Ag},\mathrm{Ir},\mathrm{Pd},\mathrm{Pt},\mathrm{Ru}\}$ ($N_E = 5$). Fix a hollow adsorption motif on an fcc(111) facet so $n_1 = 3$ (three directly coordinated surface atoms).[7] Let the near-surface cutoff yield $n_2 = 6$ second shell sites (adjustable to your geometry). Assume design constraints: $\mathcal{A}_1 = \{\mathrm{Pt},\mathrm{Pd}\}$ (N_1=2), $\mathcal{A}_2 = \{\mathrm{Pt},\mathrm{Pd},\mathrm{Ag},\mathrm{Ru}\}$ (N_2=4).

$$\text{Unconstrained:}\quad \#\mathcal{M}_{\mathrm{U}} \;=\; 5^{\,n_1+n_2} \;=\; 5^9 \;=\; 1{,}953{,}125.$$
$$\text{Subset-constrained:}\quad \#\mathcal{M}_{\mathrm{S}} \;=\; 2^{\,3}\cdot 4^{\,6} \;=\; 8\cdot 4{,}096 \;=\; 32{,}768.$$

Suppose CompGen proposes the first shell counts $\mathbf{K}_1 = (K_{1,\mathrm{Pt}}, K_{1,\mathrm{Pd}}) = (2,1)$ and the second shell counts $\mathbf{K}_2 = (K_{2,\mathrm{Pt}}, K_{2,\mathrm{Pd}}, K_{2,\mathrm{Ag}}, K_{2,\mathrm{Ru}}) = (3,1,1,1)$. Then

$$\#\mathcal{M}_{\mathrm{C}} \;=\; \underbrace{\frac{3!}{2!\,1!}}_{=3} \;\times\; \underbrace{\frac{6!}{3!\,1!\,1!\,1!}}_{=120} \;=\; 360.$$

Thus, relative to the unconstrained case, the search is reduced by a factor of $1{,}953{,}125/360 \approx 5{,}425$, and relative to the subset-only case by $32{,}768/360 \approx 91$. These counts are *before* any geometric relaxation or symmetry pruning.

Remarks on symmetry and unlabeled variants. If certain labeled positions are symmetry-equivalent (e.g., the three hollow sites under a C_3 rotation), one may divide the counts by the appropriate group size to obtain a tighter estimate. Alternatively, if one prefers to treat sites as *unlabeled*, the first shell count reduces to the number of distinct compositions only (one per feasible $\mathbf{K}_1$), which is $\binom{n_1+N_1-1}{N_1-1}$; CompGen then selects a single $\mathbf{K}_1$, and the count is 1 for that shell. Our labeled-site model provides a conservative (larger) count and hence a conservative reduction factor.

D Dataset details

Here we provide more details about the OC20 dataset used in CompGen pretraining stage. First we have the heatmap for composition distribution over element vs. adsorption energy on both shell over adsorbate *C. The composition heatmap for OC20 is presented in Figure 4 and 5.

To further investigate the relationship between composition and adsorption energy, we perform a detailed analysis on a narrow slice of the property landscape. Figure 6 and Figure 7 specifically examine the "middle energy bin", visualizing the element compositions of the extracted 15 surfaces that exhibit the lowest adsorption energies within this range. This examination provides a high resolution of the diversity for element combinations that condition on a constrained range of target adsorption energies.

[7] For atop and bridge motifs, set n_1=1 or 2.

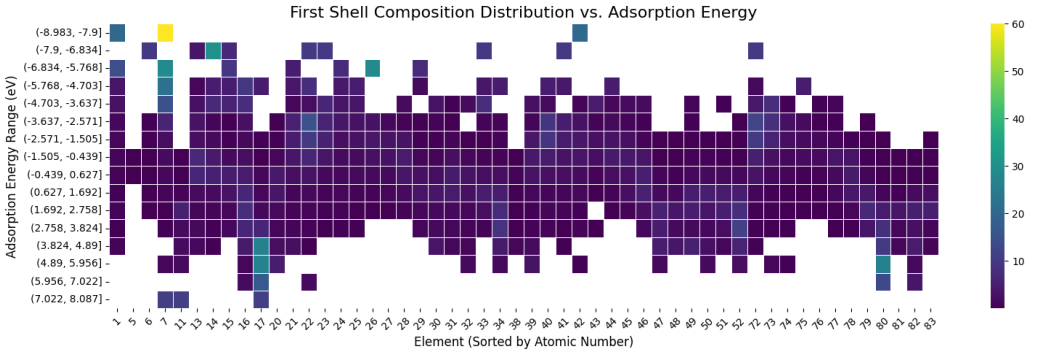

Figure 4: Composition heatmap for shell 1.

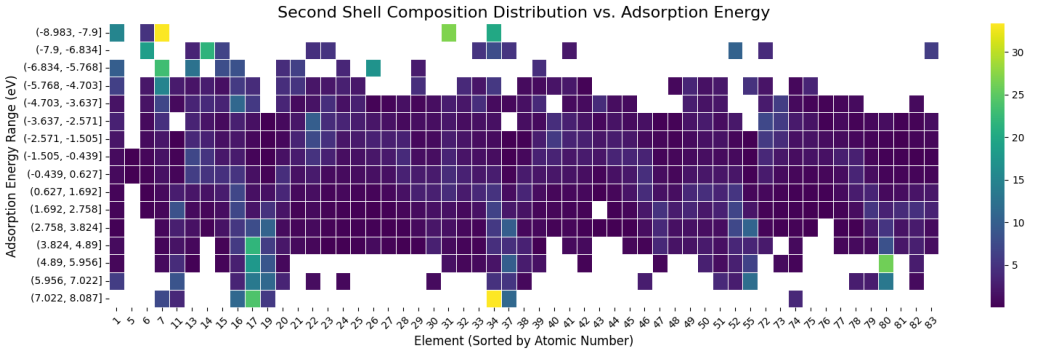

Figure 5: Composition heatmap for shell 2.

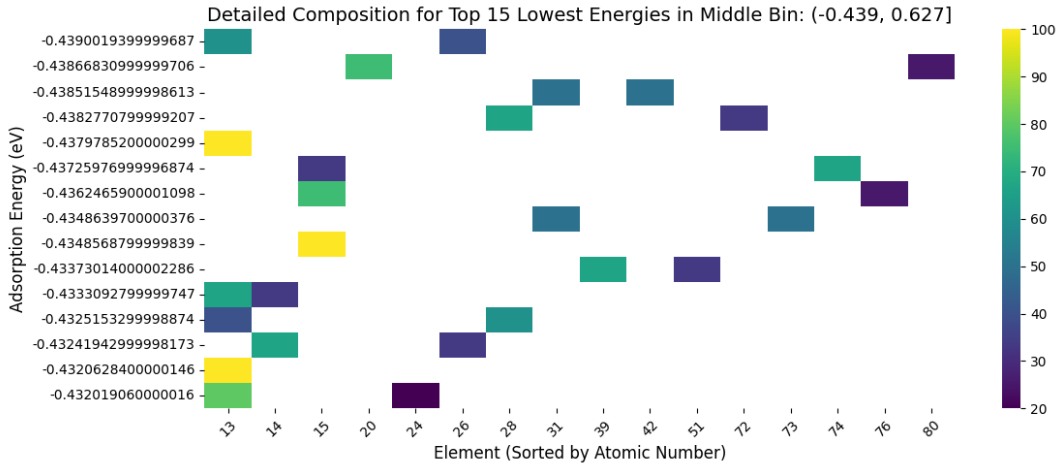

Figure 6: Composition heatmap for specific range of adsorption energies on shell 1.

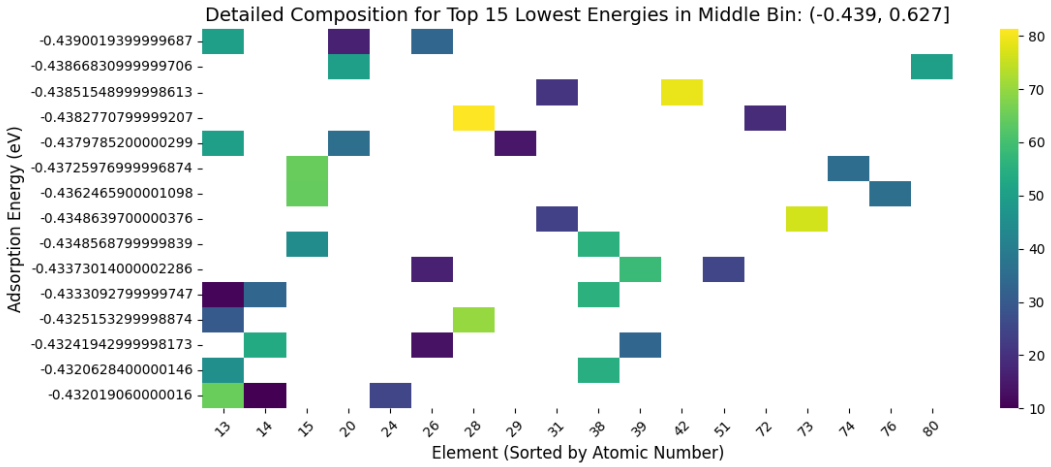

Figure 7: Composition heatmap for specific range of adsorption energies on shell 2.

E Diffusion model details

Forward Noising Process

Let $\mathbf{z}$ be the initial input x_0 for diffusion, we define a fixed forward (noising) kernel that does not depend on $\mathbf{y}$:

$$q\left(x_{1:T} \mid \mathbf{z}\right)=\prod_{t=1}^{T} q\left(x_t \mid x_{t-1}\right), \quad q\left(x_t \mid x_{t-1}\right)=\mathcal{N}\left(x_t ; \sqrt{1-\beta_t} x_{t-1}, \beta_t \mathbf{I}\right)$$

$$x_T \sim \mathcal{N}(0, \mathbf{I}).$$

Here, x_T is the noisy input at timestep T, drawn from a multivariate normal distribution with zero mean and identity covariance matrix ($\mathbf{I}$). This x_T will undergo a series of denoising steps to gradually transform into an idea latent embedding on the given information using the formula given above. Marginally,

$$q\left(x_t \mid \mathbf{z}\right)=\mathcal{N}\left(x_t ; \sqrt{\bar{\alpha}_t} \mathbf{z},\left(1-\bar{\alpha}_t\right) \mathbf{I}\right), \quad \bar{\alpha}_t=\prod_{s=1}^{t}\left(1-\beta_s\right)$$

Learned Reverse (Denoising) Process To reverse the forward diffusion process, we train a neural network ϵ_θ to predict the noise added to a noised sample x_t at time step t. This prediction is conditioned not only on x_t and t, but also on additional conditioning variables specific to the task:

$$\epsilon_\theta = \epsilon_\theta(x_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l})$$

where $\mathbf{s}$, $\mathbf{e}$ and $\mathbf{l}$ denote external conditioning inputs, composed by three counterparts as language description for elements and properties, elements in categories and properties in numerical value.

Based on this predicted noise, we define the reverse transition distribution as a Gaussian:

$$p_\theta(x_{t-1} \mid x_t, \mathbf{s}, \mathbf{e}, \mathbf{l}) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l}), \Sigma_t)$$

Here, the mean μ_θ and the variance Σ_t are typically parameterized as follows:

$$\Sigma_t = \beta_t \mathbf{I}, \quad \mu_\theta(x_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l}) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon_\theta(x_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l})\right)$$

This formulation follows the DDPM framework, where ϵ_θ learns to approximate the true noise, and the predicted mean guides the reverse sampling trajectory toward denoised data $\mathbf{z}$.

Training Objective

We still minimize the simple noise prediction loss, but conditioned on $(\mathbf{s}, \mathbf{e}, \mathbf{l})$:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{z},\mathbf{s},\mathbf{e},\mathbf{l},\epsilon\sim\mathcal{N}(0,I),t}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}}\mathbf{z} + \sqrt{1-\bar{\alpha}_t}\epsilon, t, \mathbf{s}, \mathbf{e}, \mathbf{l}\right)\right\|^2\right].$$

Here, $\mathbf{s}, \mathbf{e}, \mathbf{l}$ are drawn from their respective empirical distributions, ϵ is standard Gaussian noise, and t is uniform on $\{1, \ldots, T\}$.

Inference (Sampling) At inference time, we begin by drawing a pure noise sample x_T from the standard normal distribution $\mathcal{N}(0, \mathbf{I})$. Then, for each timestep t counting down from T to 1 , we denoise by sampling

$$x_{t-1} \sim \mathcal{N}\left(\mu_\theta\left(x_t, t, \mathbf{s}, \mathbf{e}, \mathbf{l}\right), \beta_t \mathbf{I}\right)$$

where μ_θ is the predicted mean of our network conditioned on the current noisy state x_t, the timestep t, and the three conditioning signals s (adsorbate), e (adsorption energy) and ℓ (language description) which will be transferred as vectors $\mathbf{s}$, $\mathbf{e}$ and $\mathbf{l}$ correspondingly. As we step backward through time, the noise magnitude β_t shrinks according to our predefined schedule, gradually transforming the initial noise into our target latent sample. After completing the final step at $t = 1$, the resulting $\mathbf{z}$ is returned as a sample approximately drawn from the desired conditional distribution $p_{\text{data}}\ (\mathbf{z} \mid \mathbf{s}, \mathbf{e}, \mathbf{l})$.