Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Informed Generative Inverse Design of Stable Vanadium Oxide Crystals with First-Principles Validation

Danial Ebrahimzadeh
University of Oklahoma
Sarah Sharif
University of Oklahoma
Yaser Banad
bana@ou.edu
University of Oklahoma

Research Article

Keywords: Generative materials design, Inverse design, Vanadium oxides, Wasserstein GAN, Variational autoencoder, Density functional theory, Functional oxides, Crystal generation

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10456555/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Physics-Informed Generative Inverse Design of Stable Vanadium Oxide Crystals with First-Principles Validation

Danial Ebrahimzadeh[1,2], Sarah S. Sharif[1,2], Yaser M. Banad[1,2*]

[1]School of Electrical and Computer Engineering, University of Oklahoma, 110 W. Boyd St., Norman, 73019, OK, USA.
[2]INQUIRE Lab, University of Oklahoma, Norman, 73019, OK, USA.

*Corresponding author(s). E-mail(s): bana@ou.edu;
Contributing authors: danial.ebrahimzadeh@ou.edu; s.sharif@ou.edu;

Abstract

Vanadium oxides are a compositionally rich family of functional transition-metal oxides with diverse electronic, optical, thermochromic, electrochemical, and structural behaviors. Their multiple accessible oxidation states and complex phase space make them attractive targets for computational materials discovery, but make systematic first-principles screening computationally demanding. We propose an inverse-design workflow that couples a voxel-based variational autoencoder (VAE) with a formation-energy-constrained Wasserstein generative adversarial network (WGAN) to generate physically plausible V–O crystal candidates while mitigating adversarial training instability. Crystal structures are encoded as element-resolved 3D voxel grids, compressed into a continuous latent space, and sampled by a WGAN whose generator is biased toward low formation energy via an exponential penalty predicted by a frozen convolutional neural network (CNN) regressor. Using 10,981 density functional theory (DFT)-relaxed structures for training and validation, the framework generates 6,592 candidate V–O crystals with high generative quality (validity 81.5%, uniqueness 96.6%, novelty 99.5%). First-principles screening identifies 669 stable candidates (10.2%), 365 metastable candidates (5.5%), and a further 94 fall below the Materials Project–referenced convex hull and are reported as candidates for further validation. A composition-matched DFT control confirms that the formation-energy constraint raises the first-principles stable fraction from 15.9% to 25.4%. Phonon calculations on representative candidates, including one below the hull, support lattice-dynamical stability, with only minor soft modes attributable to finite-size effects or known phase transitions. The proposed pipeline reduces the DFT

screening burden and provides a scalable route to functional oxide compositions for thermochromic, electronic, electrochemical, and adaptive thermal-material applications.

Keywords: Generative materials design, Inverse design, Vanadium oxides, Wasserstein GAN, Variational autoencoder, Density functional theory, Functional oxides, Crystal generation

1 Introduction

Functional transition-metal oxides occupy a central position in modern materials science because small changes in stoichiometry, oxidation state, coordination environment, and crystal symmetry can produce large changes in electronic, optical, magnetic, catalytic, and electrochemical behavior. This sensitivity creates opportunities for discovering new functional materials, but it also makes oxide phase spaces difficult to explore systematically. Vanadium oxides (V-O) provide a representative and scientifically rich example. The V-O binary system contains multiple stable and metastable phases, including metallic VO, the Magnéli series V_nO_{2n-1}, VO_2, V_2O_3, and V_2O_5, whose functional behavior is governed by vanadium's accessible oxidation states and diverse coordination environments [1, 2]. VO_2 is widely studied for its first-order metal-insulator transition near 68 °C, which produces large changes in electrical resistivity and infrared optical response [3]. V_2O_3 is a prototypical correlated oxide with pressure- and temperature-driven electronic transitions, whereas V_2O_5 is widely investigated for electrochemical, electrochromic, and catalytic applications [2, 4, 5]. The polymorphism of V_xO_y compounds has also motivated studies of memristive switching, reconfigurable electronics, and neuromorphic device concepts [6, 7]. These diverse functionalities make the V–O system a useful model space for testing generative inverse-design strategies for functional oxides.

The breadth of functionalities summarized in Fig. 1 is mirrored by a rapidly expanding research literature. Web of Science records from 2000 to 2025 indicate that more than 30,000 publications span the V-O binary system, dominated by V_2O_5, VO_2, and V_2O_3, with sustained year-over-year growth across the major phases (Figs. S1 and S2). A clear and more recent emergence of lower-frequency Magnéli-series compositions such as V_3O_7, V_4O_9, V_6O_{13}, and V_7O_{13} is also evident (Fig. S3), reflecting the broadening interest in less-studied stoichiometries that lie outside the conventional VO_2-V_2O_5 axis. The most highly cited works in the field (Fig. S4) span thermochromic devices, infrared optics, electrochemical storage, electrochromism, and correlated-electron physics, underscoring both the diversity of applications and the persistent scientific interest in identifying new functional phases [8]. This expanding and increasingly diversified V-O literature motivates inverse-design strategies capable of operating across the broad V–O compositional range representable within the fixed voxel encoding rather than within a few well-characterized stoichiometries.

Systematically discovering stable and metastable V-O structures remains challenging because vanadium's multiple accessible oxidation states produce a rich phase space

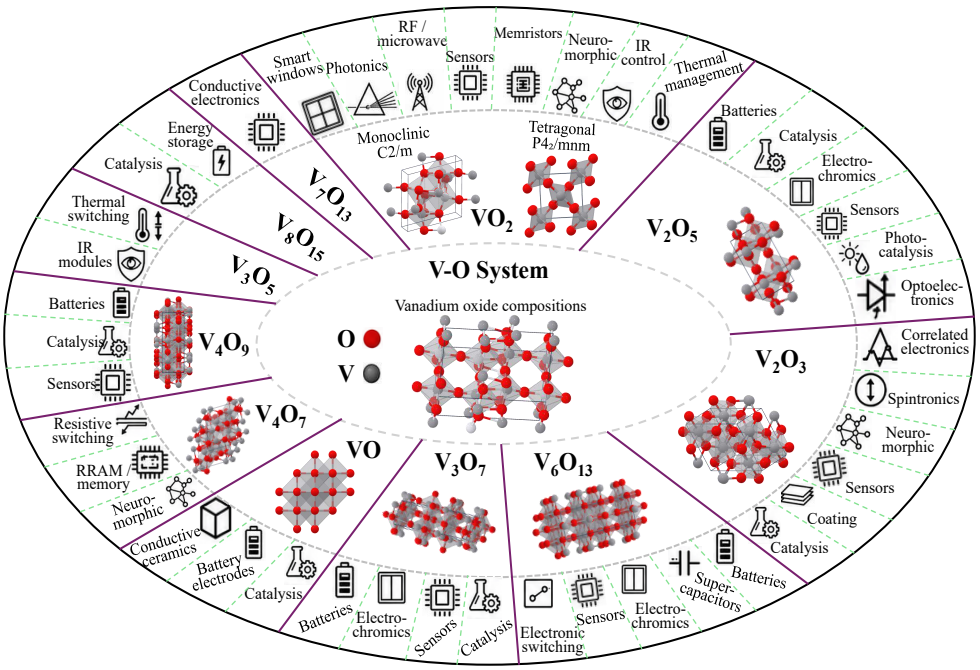


Fig. 1 Functional applications of the vanadium-oxide system organized by composition. Each V_xO_y phase exhibits distinct electronic, optical, electrochemical, and structural functionalities, spanning smart windows, infrared and thermal modulation, electrochromic coatings, photonics, energy storage, catalysis, sensing, and neuromorphic and memristive devices. The diversity of accessible compositions and functionalities motivates composition-agnostic inverse-design strategies capable of operating across the full V-O phase space

in which small stoichiometric or structural changes can substantially alter electronic structure and functional behavior [9]. Traditional materials exploration has relied on chemical intuition, iterative synthesis, and targeted first-principles calculations [10]. High-throughput DFT workflows and repositories such as Materials Project (MP) [11], the Open Quantum Materials Database (OQMD) [12], and AFLOW [13] have substantially accelerated computational screening by cataloging large numbers of relaxed structures and formation energies. However, direct high-throughput DFT remains a forward-design strategy. Candidate structures must be specified before they can be evaluated. For chemically flexible oxide systems such as V-O, exhaustive enumeration across stoichiometry, lattice geometry, and coordination motifs remains computationally expensive [14], motivating inverse-design strategies that can propose promising candidates before first-principles validation.

Materials informatics has emerged as a practical route for reducing this search burden by combining curated datasets, machine-learning models, and targeted validation workflows. Recent studies have shown that machine learning can accelerate materials discovery through property prediction, phase-space exploration, phase-diagram construction, simulation acceleration, multi-objective optimization, and inverse design. Related inverse-design studies have also demonstrated that trained surrogate models

can be coupled with optimization algorithms to identify candidate structures satisfying target properties. These developments motivate generative workflows that propose physically plausible crystal structures before expensive first-principles validation.

Inverse design reverses the conventional by specifying desired performance targets first and computationally identifying candidate structures that satisfy them (Fig. 2) [15, 16]. Surrogate models coupled with optimization algorithms have already demonstrated the broader utility of this paradigm beyond forward screening [17]. Physics-based inverse methods, including particle swarm optimization in CALYPSO [18] and evolutionary algorithms in USPEX [19], can locate stable polymorphs without experimental input, and Bayesian optimization further reduces costly energy evaluations [20]. However, these global optimizers scale poorly to high-dimensional configuration spaces and frequently converge on local minima. Deep learning (DL) [21] has provided a transformative complement by enabling generative models that learn continuous latent representations of known materials and sample new candidates from the learned distribution. Recent materials-informatics studies have shown that machine learning can accelerate materials discovery by learning structure-property relationships, guiding phase-space exploration, accelerating simulation workflows, and supporting multi-objective optimization and inverse design [22]. VAEs [23] encode crystals into smooth latent spaces supporting property-targeted optimization, while GANs [24] train competing generator-discriminator networks to produce realistic samples without explicit likelihood estimation. Court *et al.* demonstrated three-dimensional density-map representations for inorganic crystal generation and property prediction [25], and Ren *et al.* introduced an invertible crystallographic representation enabling inverse mapping within latent space [26]. Moreover, the Crystal Diffusion Variational Autoencoder (CDVAE) incorporated stability as a physical inductive bias through score-matching decoders and periodic graph neural networks [27], while MatterGen generates diverse inorganic structures across the periodic table with fine-tuning toward targeted property constraints [28]. A comprehensive review by Alverson *et al.* situates these architectures within a unified taxonomy for generative crystal design [29].

The crystal generative modeling landscape has evolved rapidly from early voxel-based VAEs to modern symmetry-aware and diffusion-based architectures, and it is important to position the present work within this progression. Voxel representations, as employed in iMatGen [30] and the Cond-DFC-VAE framework [25], encode three-dimensional electron density or atomic occupancy on fixed grids, offering direct compatibility with convolutional neural networks and naturally preserving spatial relationships. Their principal limitation is resolution: the cubic grid imposes a fixed spatial discretization that can break crystallographic symmetry and becomes memory-intensive for large unit cells. Graph-based and equivariant approaches circumvent these issues. CDVAE [27] operates on periodic graphs with SE(3)-equivariant message passing and uses a score-matching decoder to iteratively refine atomic coordinates toward low-energy configurations. DiffCSP [31] extends this concept by jointly diffusing atom types, coordinates, and lattice parameters under fractional-coordinate periodicity, while MatterGen [28] further removes the VAE bottleneck entirely and employs a fully joint diffusion process with fine-tunable property adapters. Physics-guided WGAN

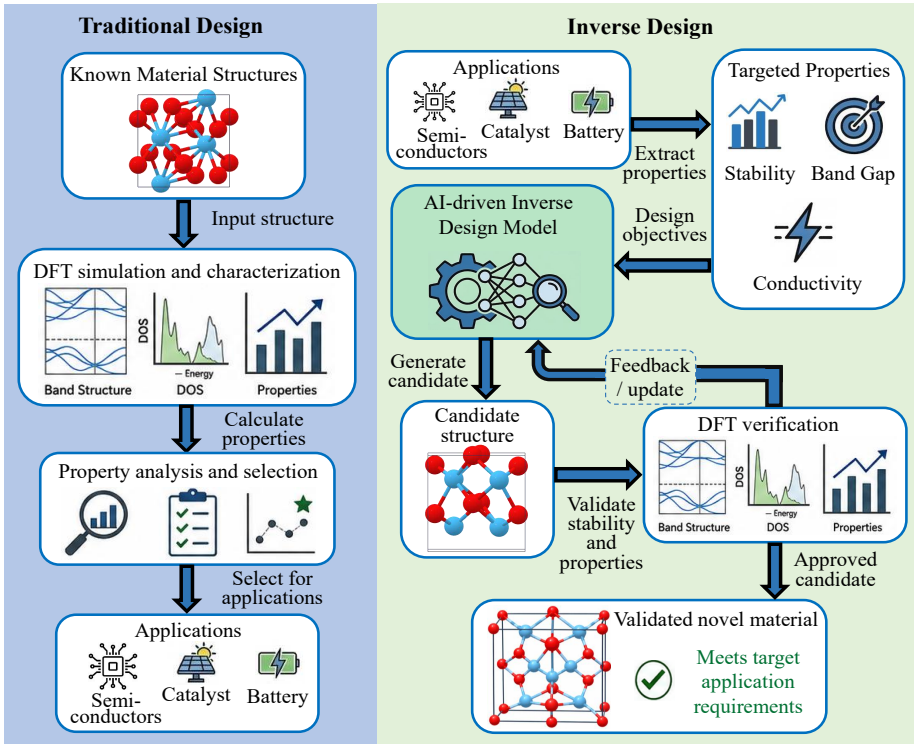


Fig. 2 Comparison of forward materials discovery and inverse design for functional inorganic materials. The forward route (left) proceeds from a known structure through DFT simulation and property analysis to application. The inverse route (right) begins from targeted properties (stability, band gap, conductivity), generates candidate structures with an AI-driven model, and validates them by DFT in a feedback loop until a candidate meets the target application requirements

architectures [32] and constrained deep convolutional GANs [33] take a complementary route by embedding energy or symmetry penalties directly into adversarial training, improving validity without requiring equivariant network design. Moreover, Kim *et al.* applied a WGAN to the Mg-Mn-O ternary system, discovering 23 crystal structures with plausible stability [34]. A related V-O generative study previously combined WGAN and VAE components for three fixed stoichiometries, VO_2, V_2O_3, and V_2O_5, generating 451 candidate compositions for electronic-materials design [35]. That study demonstrated the feasibility of adversarial generation in selected V-O subspaces, but its fixed-stoichiometry setting did not test whether stability-guided generation can generalize across the broader binary V-O phase space. The present work addresses this broader setting by expanding the training corpus, removing fixed-stoichiometry sampling constraints, and evaluating generated candidates across a broad compositional range of V_xO_y stoichiometries.

Despite this rapid progress, stability-constrained adversarial generation has not been demonstrated in a composition-agnostic setting across the multi-valent V–O binary system, nor has the contribution of the thermodynamic constraint itself been

isolated against a first-principles control. The present framework bridges element-resolved voxel encoding with a stability-constrained WGAN whose exponential energy penalty biases generation toward low-energy regions of the V-O phase space. Physics-informed ML strategies for inverse materials design, such as those employing neural networks to navigate microstructure-property mappings [36], share a similar philosophy of embedding domain knowledge into the learning objective; the present framework applies this principle at the crystal-structure level through a frozen formation-energy regressor coupled to the generator loss.

Two methodological challenges remain central to generative crystal design. First, ensuring realistic thermodynamic stability in generated transition-metal oxides is complicated by oxygen vacancies, mixed valence, and competing metastable polymorphs [37, 38], which together produce a rugged energy landscape that conventional adversarial objectives are ill-equipped to navigate. Second, periodic crystals require encoding schemes that preserve lattice periodicity, atomic connectivity, and symmetry while remaining compatible with deep learning architectures, a challenge not faced by molecular systems with compact representations such as SMILES [39]. WGANs partially address training instability by replacing binary classification with Earth Mover's Distance optimization [40], but integrating formation energy constraints directly into the generative training loop remains largely unexplored.

In this work, we introduce a physics-guided inverse-design framework for composition-agnostic generation of V-O crystals. The framework combines two components: (i) a voxel-based VAE that encodes element-resolved atomic densities and lattice geometry into a continuous latent space, and (ii) a WGAN whose generator loss incorporates a formation-energy penalty predicted by a frozen CNN regressor. This design couples distributional realism with a thermodynamic bias toward low-energy regions of the V-O phase space. Applying this framework to the V-O binary system, we generated 6,592 unique compositions spanning a broad range of V_xO_y stoichiometries. Of the 6,592 generated structures, 1,846 (28.0%) have negative formation energies and 4,746 (72.0%) have positive formation energies. A total of 669 structures (10.2%) satisfy stringent stability criteria (negative formation energy and ≤ 300 meV/atom above the MP convex hull), with an additional 365 (5.5%) meeting a metastability threshold of ≤ 500 meV/atom. Notably, 94 generated configurations exhibit formation energies falling below the MP-based convex-hull reference, representing potentially novel thermodynamically favorable phases not currently documented in the training database requiring further validation. These candidates were not present in the training set and therefore represent previously unseen outputs of the generative model. Phonon calculations on representative candidates, including one lying below the MP-referenced hull, confirm dynamic stability, with minor imaginary modes attributable to finite-size effects or known phase transitions rather than fundamental instabilities. These results suggest that the proposed framework provides a generalizable computational pathway for discovering stable and metastable functional oxides, with natural extensibility to ternary and higher-order oxide systems.

Section 2 details the inverse design methodology, including the VAE encoding scheme and stability-constrained WGAN architecture. Section 3 presents generation performance, latent-space organization, and thermodynamic and dynamic stability assessments.

2 Methodology

The inverse-design framework developed in this work couples generative deep learning with first-principles validation to discover thermodynamically stable and metastable V–O crystal structures. As illustrated in Fig. 3, the pipeline progresses through five interconnected stages: (a) data acquisition and curation from the MP database, (b,c) voxel-based crystal structure encoding into fixed-size three-dimensional grids compatible with convolutional architectures, (d) a coupled VAE–WGAN generative engine that operates in a learned latent space while enforcing formation-energy constraints, and (e) ab initio thermodynamic and dynamic stability verification of every generated candidate. Each stage is detailed in the subsections that follow. A high-level conceptual overview of this data-representation-coding-training-validation workflow is also provided in Fig. S5.

2.1 Dataset Construction

The scarcity of experimentally characterized V–O phases across the full binary composition range motivates a computational data-generation strategy. Starting from the curated binary compounds available in the MP, we applied systematic element substitution to populate V–O structural prototypes across a broad range of V_xO_y stoichiometries (Fig. 3(a)). In this procedure, existing binary oxide and binary metallic structures serve as templates: one element is replaced by vanadium and the other by oxygen (or vice versa), producing candidate configurations that inherit known crystallographic frameworks while exploring new chemical identities. This substitution-based augmentation strategy ensures structural diversity—spanning layered, rock-salt, spinel-related, and corundum-derived motifs—while remaining grounded in experimentally observed structure types.

Every candidate structure was relaxed with the Vienna Ab initio Simulation Package (VASP) [41] using the Perdew–Burke–Ernzerhof (PBE) generalized-gradient exchange-correlation functional and projector augmented wave (PAW) pseudopotentials for both vanadium and oxygen. Conjugate-gradient relaxations were converged to 10^{-5} eV in total energy and 0.05 eV/Å in residual forces, with a plane-wave cutoff of 500 eV and reciprocal-space sampling at 0.5 $Å^{-1}$ grid spacing. To capture the strong on-site Coulomb interaction of V-3d electrons—which governs the electronic and magnetic properties of V–O and critically affects computed formation energies—all calculations employed the rotationally invariant DFT+U formulation of Dudarev (`LDAUTYPE = 2`) with $U_{\text{eff}} = U - J = 3.25$ eV [42]. The resulting dataset of 10,981 relaxed structures and their DFT formation energies provides the training corpus for the generative models and spans formation energies from -2.68 to $+4.35$ eV/atom with oxygen fractions from 0.07 to 0.93, ensuring coverage of the V–O compositional and configurational space.

2.2 Voxel-Based Crystal Representation

Traditional crystallographic file formats such as CIF present significant obstacles for DL applications: they are discontinuous, vary in length from structure to structure, and do not naturally encode the spatial relationships exploited by convolutional operations.

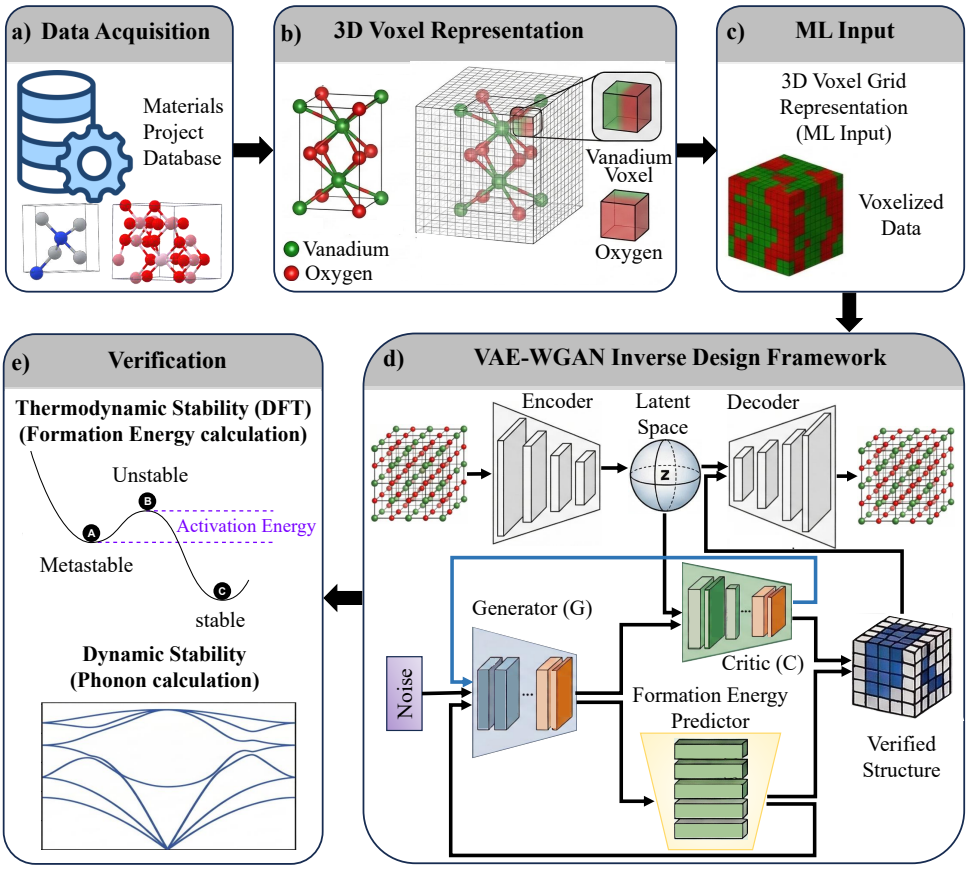


Fig. 3 Overview of the inverse-design pipeline for stable V–O crystals. (a) Data acquisition and augmentation: binary compounds from the Materials Project are expanded by element substitution to populate V_xO_y prototypes. (b,c) Element-resolved voxel encoding, in which each structure is mapped onto fixed-size three-dimensional grids (64^3 for the V and O atomic-site channels, 32^3 for lattice geometry) compatible with 3D convolutional networks. (d) The coupled VAE–WGAN generative engine: the VAE compresses voxel grids into a continuous latent space (upper path), while the WGAN generator and critic operate in that space (lower path), with a frozen formation-energy predictor biasing generation toward low-energy regions. (e) Ab initio verification of every generated candidate, combining DFT thermodynamic stability with phonon-based dynamic stability

We therefore transform each crystal into a set of fixed-size three-dimensional voxel grids that serve as direct inputs to the neural network (Fig. 3(b,c)).

The voxelization process begins by centering each structure within a standardized 15 Å cubic box to establish translational invariance. Atomic positions are represented by radial basis Gaussian functions placed at nuclear coordinates, with each element assigned a dedicated volumetric channel. The occupancy of voxel (i, j, k) due to an atom at position $\mathbf{r}$ is computed as

$$p_{ijk} = \exp\left(-d_{ijk}^2 \,/\, 2\sigma^2\right), \tag{1}$$

where d_{ijk} is the Euclidean distance from the voxel center to the atomic position and $\sigma = 0.26$ Å is the Gaussian width, chosen to produce well-localized, atom-centered density peaks on the 15 Å cubic grid at a resolution of 64^3 voxels (voxel spacing ≈ 0.23 Å). Voxels with occupancy below a threshold of 0.02 are set to zero to suppress background noise and define the foreground mask used consistently during both encoding and decoding. For V–O binary compounds, this yields two $64 \times 64 \times 64$ grids (one for vanadium, one for oxygen) where each atom is represented as a localized density peak; proximity relationships between neighboring atoms are captured through the spatial arrangement of these peaks within the shared volumetric channel. Because each element occupies its own channel, atom ordering within the input file is irrelevant, inherently satisfying permutational invariance. A third $32 \times 32 \times 32$ grid encodes lattice geometry: a single Gaussian is placed at the unit-cell center, and smoothed distances are recorded along the three principal axes and the three face diagonals, compactly capturing cell dimensions, lengths, and angles while preserving geometric relationships. Periodic boundary conditions are enforced during all distance evaluations so that crystal periodicity is faithfully maintained in the representation.

This voxel-based scheme offers several advantages over alternative representations. Compared with graph-based encodings, which require explicit construction of periodic neighbor lists and variable-size connectivity representations, voxel grids provide spatially uniform inputs of fixed dimensionality that are directly compatible with standard 3D convolutional layers [39, 43]. The fixed-resolution spatial encoding simultaneously preserves local atomic arrangements and global structural features, enabling efficient learning of the complex structure–property relationships that govern V–O phase stability.

The voxel-based representation adopted here discretizes space on a fixed 64^3 grid, which imposes an upper bound on the unit-cell size that can be faithfully encoded and may introduce minor symmetry-breaking artifacts at boundaries. The current framework does not explicitly enforce space-group symmetry during generation, relying instead on post-generation DFT relaxation to recover the correct symmetry. Oxidation-state assignments are inferred from stoichiometry and Bader charge analysis rather than being directly encoded in the voxel channels, which limits the model's ability to distinguish between iso-stoichiometric phases with different charge orderings. These limitations motivate future work on symmetry-aware voxel encodings, explicit oxidation-state channels, and extension to ternary and higher-order oxide systems.

2.3 Variational Autoencoder

The VAE compresses the high-dimensional voxel grids into a smooth, continuous latent space amenable to generative sampling and property-directed optimization (upper path in Fig. 3(d)). The encoder consists of successive 3D convolutional layers with Leaky ReLU activations, batch normalization, and dropout regularization, which progressively extract hierarchical structural features from the voxelized input. The final convolutional block outputs two vectors (a mean $\boldsymbol{\mu}$ and a log-variance $\log \boldsymbol{\sigma}^2$) that together parameterize a Gaussian posterior distribution in latent space. The prior is

set to the standard Gaussian $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Latent codes are sampled via the reparameterization trick, $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, which preserves end-to-end differentiability while introducing the stochasticity needed for generative sampling.

The decoder mirrors the encoder through 3D transposed convolutional layers augmented with ResNet-style residual connections [44]. In this design, feature maps from earlier (lower-resolution) decoder layers are upsampled and concatenated with those in later (higher-resolution) layers before the next convolution. This skip-connection architecture improves the decoder's capacity to reconstruct fine structural details—such as precise oxygen coordination environments and subtle lattice distortions—that would otherwise be lost during compression to the low-dimensional bottleneck.

Training jointly optimizes the encoder and decoder to minimize a composite loss balancing reconstruction fidelity and latent-space regularity:

$$\mathcal{L}_{\mathrm{VAE}} = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2}_{\mathcal{L}_{\mathrm{recon}}} + \beta \underbrace{\left[-\frac{1}{2}\sum_{j=1}^{d}\left(1 + \log\sigma_j^2 - \mu_j^2 - \sigma_j^2\right)\right]}_{\mathcal{L}_{\mathrm{KL}}} \tag{2}$$

where $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ are input and reconstructed voxel grids, d is the latent dimensionality, and β weights the Kullback–Leibler (KL) divergence term that regularizes the posterior toward the standard Gaussian prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. We employ KL annealing, linearly increasing β from 10^{-4} to 1.0 over the first 100 training epochs while holding it constant thereafter, to prevent posterior collapse, a well-known failure mode in which the decoder learns to ignore the latent code. Separate VAE instances encode the atomic-site channels and the lattice-parameter channel with latent dimensions of 200 and 25, respectively; their concatenated codes form the input to the subsequent WGAN stage.

After the VAE decoder reconstructs the element-resolved voxel grids from a generated latent code, the continuous Gaussian density fields must be converted back into discrete atomic positions and a lattice matrix. For each element channel, voxels with occupancy above the foreground threshold (0.02) are clustered using a peak-finding algorithm that identifies local maxima of the density field; cluster centroids are taken as atomic coordinates. Atoms within 0.5 Å of each other are merged to suppress duplicate peaks introduced by discretization noise. The lattice channel is decoded by recovering the six characteristic distances (three principal axes and three face diagonals) from the smoothed-distance Gaussian field; these six distances are inverted to reconstruct the lattice matrix using the analytical relations between distances, edge lengths, and angles. The resulting structure is written as a VASP POSCAR file and passed through pymatgen's structure validator to remove obvious geometric artifacts before DFT relaxation.

2.4 Wasserstein GAN with Formation-Energy Constraints

The WGAN [40] operates in the continuous latent space learned by the VAE, generating novel V–O candidates while explicitly favoring thermodynamic stability (lower path in Fig. 3(d)). A generator network G transforms random noise vectors $\mathbf{z}_{\mathrm{noise}} \sim$

$\mathcal{N}(\mathbf{0}, \mathbf{I})$ into synthetic latent codes through a sequential architecture of dense layers, two-dimensional upsampling operations, and convolutional layers with batch normalization, progressively building high-dimensional latent representations compatible with the VAE decoder input. A critic network C, replacing the binary discriminator of conventional GANs, assigns an unbounded, continuous realism score to each sample rather than a probability. Because the critic minimizes the Wasserstein (Earth Mover's) distance between real and generated latent distributions rather than a binary cross-entropy objective, it provides informative and non-vanishing gradient signals throughout training, even when the critic is near-optimal [45]. This property is fundamental to avoiding the mode collapse and training oscillations that plague standard adversarial architectures.

To enforce the 1-Lipschitz constraint required by the Wasserstein objective, the critic applies weight clipping with bounds $[-c, c]$, $c = 10^{-2}$, to all trainable weights, following Arjovsky *et al.* [40]. The critic is updated $n_{\text{critic}} = 5$ times for every generator update to ensure that the critic remains close to optimal throughout training, yielding a faithful estimate of the Wasserstein distance at each generator step.

The generator loss couples adversarial realism to a thermodynamic stability constraint by integrating a frozen formation-energy predictor directly into the objective. The critic minimizes:

$$\mathcal{L}_C \;=\; \mathbb{E}_{\mathbf{z}}\big[C(G(\mathbf{z}))\big] \;-\; \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}\big[C(\mathbf{x})\big], \tag{3}$$

which yields informative gradients even when the critic is near-optimal [40]. The generator minimizes the composite loss:

$$\mathcal{L}_G \;=\; -\mathbb{E}_{\mathbf{z}}\big[C(G(\mathbf{z}))\big] \;+\; \frac{\lambda}{B}\sum_{i=1}^{B}\exp\big(\hat{E}_f(G(\mathbf{z}_i))\big), \qquad \lambda = 0.1, \tag{4}$$

where $\hat{E}_f$ is the frozen CNN formation-energy regressor, $B = 128$ is the mini-batch size, and λ controls the balance between distributional fidelity and thermodynamic feasibility. The exponential penalty provides an asymmetric thermodynamic bias: when the regressor predicts a negative formation energy ($\hat{E}_f < 0$), $\exp(\hat{E}_f) < 1$ and the penalty contribution is small; when $\hat{E}_f > 0$, the penalty grows exponentially, strongly discouraging the generator from producing high-energy configurations. The exponential acts on the numerical formation energy in eV/atom, equivalently $\exp(\hat{E}_f/E_0)$ with $E_0 = 1$ eV/atom, so its argument is dimensionless. Over the training-energy range this yields the intended asymmetry: $\exp(-2.68) \approx 0.069$ versus $\exp(4.35) \approx 77.5$, so high-energy predictions dominate the batch-averaged penalty and are suppressed far more strongly than favorable negative-energy predictions. The value $\lambda = 0.1$ was selected to keep the formation-energy term and the Wasserstein term in approximately the same order of magnitude during training.

This choice was confirmed by a penalty-weight sweep, $\lambda \in \{0, 0.05, 0.10, 0.20, 0.50\}$ (Table S1). Because full DFT validation across all settings is computationally prohibitive, the frozen formation-energy regressor (test MAE 0.064 eV/atom) was used as a relative surrogate, applying the same stability criterion

used for DFT screening to the predicted formation energies. Without the penalty ($\lambda = 0$), only 5.1% of generated structures are predicted stable; the penalty raises this to 24.2% at $\lambda = 0.1$, isolating the formation-energy term as the active driver of thermodynamic bias. Predicted stability saturates for $\lambda \geq 0.1$ (24.2–24.8%) while uniqueness, novelty, and structural diversity decrease monotonically and validity peaks at $\lambda = 0.1$, placing $\lambda = 0.1$ at the elbow of the stability–diversity trade-off. The effect of the constraint on first-principles stability is verified independently of the regressor by a composition-matched DFT control.

2.5 Formation-Energy CNN Regressor

The formation-energy regressor $\hat{E}_f$ provides the thermodynamic signal that biases generation toward stable regions of latent space. Its input is the same encoded crystal representation used for WGAN training. For each structure, the representation contains three fixed-length 200-component rows: one row for the lattice encoding and two rows for the encoded atomic-site channels corresponding to V and O. The lattice VAE produces a 25-dimensional latent code, which is stored in the first 25 entries of the lattice row, with the remaining entries set to zero. Together, these three rows give 425 values, which are then padded with 359 additional zeros to obtain $784 = 28 \times 28$ values and reshaped into a 28×28 feature map for two-dimensional convolutional processing.

The network comprises four two-dimensional convolutional layers with 32, 64, 128, and 256 filters, respectively, using 3×3 kernels, LeakyReLU activations with negative slope 0.2, dropout, and batch normalization after the second through fourth convolutional blocks. The convolutional features are flattened and passed through six fully connected layers with 1024, 256, 256, 256, 64, and 1 units to regress the formation energy. The regressor was trained independently using mean-squared-error loss and the Adam optimizer. On a held-out test set the regressor achieves a mean absolute error of 0.064 eV/atom, sufficient for the relative ranking of candidate stability used in the penalty-weight ablation (Table S1) while remaining orders of magnitude faster than DFT. After training, its weights were loaded into the WGAN model and frozen, so that during adversarial training the formation-energy gradients affect the generator while the regressor parameters remain fixed.

2.6 Ab Initio Validation and Stability Classification

All decoded structures, including those failing the initial geometric screen, underwent DFT relaxation in VASP using the identical PBE+U settings described in Section 2.1, ensuring full consistency between the training data and the validation calculations (Fig. 3(e)). Following relaxation, formation energies are computed as

$$E_f = \frac{E_{\mathrm{V}_x\mathrm{O}_y} - \left(x\, E_{\mathrm{V}}^{\mathrm{ref}} + y\, E_{\mathrm{O}}^{\mathrm{ref}}\right)}{x + y} \tag{5}$$

where $E_{\mathrm{V}_x\mathrm{O}_y}$ is the total energy of the relaxed compound and $E_{\mathrm{V}}^{\mathrm{ref}}$, $E_{\mathrm{O}}^{\mathrm{ref}}$ are reference energies for elemental vanadium and molecular oxygen in their respective standard states. For MP convex hull analysis, generated V–O entries and database reference

entries were processed using the same pymatgen compatibility workflow to maintain consistency with Materials Project thermochemical corrections. Convex-hull distances were computed in the binary V–O chemical system using pymatgen phase–diagram tools. Thermodynamic classification is applied to the negative-formation-energy structures using the distance to the MP convex hull, ΔE_{hull}. A structure with $E_f < 0$ and $0 \leq \Delta E_{\text{hull}} \leq 300$ meV/atom is labeled stable; one with $300 < \Delta E_{\text{hull}} \leq 500$ meV/atom is metastable; and one with $\Delta E_{\text{hull}} > 500$ meV/atom is unstable. Structures lying below the current MP hull ($\Delta E_{\text{hull}} < 0$) are reported separately as potentially novel phases. These thresholds are more restrictive than the ≤ 500 meV/atom cutoff adopted in earlier generative studies [30] to ensure that flagged candidates possess genuine practical viability.

Dynamic stability is assessed independently through phonon dispersion calculations on selected low-formation-energy compositions. We employ the finite-displacement method as implemented in Phonopy [46, 47], coupled with VASP force evaluations on appropriately sized supercells. Prior to phonon analysis, each structure is re-optimized with tightened convergence criteria: residual forces below 0.01 eV/Å and energy convergence better than 10^{-6} eV. Force constants are extracted from supercells typically ranging from $3 \times 3 \times 2$ to $5 \times 4 \times 4$, with dimensions selected to ensure adequate sampling of long-range interatomic interactions. Supercell dimensions were chosen so that each lattice vector of the supercell extended at least 10 Å, ensuring sufficient real-space sampling of interatomic force constants. The absence of large imaginary phonon branches across the full Brillouin zone serves as the criterion for confirming lattice-dynamical stability.

3 Results and Discussion

3.1 Computational Implementation and Reproducibility

The complete framework was implemented in Python 3 using TensorFlow/Keras for generative model construction, with ASE (Atomic Simulation Environment) for crystal structure manipulation, and pymatgen for convex-hull analysis. A random 90/10 train–test split was applied independently for both the formation-energy CNN regressor and the GAN evaluation set. To prevent gradient leakage, the formation-energy regressor was trained to convergence and its weights frozen before any adversarial training commenced. The training set was resampled 20× with replacement (bootstrap resampling) to stabilize mini-batch statistics at the chosen batch size of 128.

Table 1 summarizes the key parameters of the model. The VAE employed latent dimensions of 200 for atomic sites and 25 for lattice parameters, trained with the Adam optimizer (learning rate 3×10^{-4}) for 201 and 101 epochs, respectively. The formation-energy CNN regressor was trained for 50 epochs with the Adam optimizer at learning rate 1×10^{-3}, batch size 128. For the WGAN, both the generator and critic were trained for 5,000 epochs using Adam (learning rate 2×10^{-4}, $\beta_1 = 0.5$, decay $= 10^{-4}$) with a batch size of 128 and formation-energy constraint weight $\lambda = 0.1$. DFT calculations employed VASP with the PBE+U parameters detailed in Section 2.1, and phonon calculations were performed on NVIDIA H100 GPUs.

Table 1 Model parameters and dataset statistics

Parameter	Value
Voxel grid – atomic sites	$64 \times 64 \times 64$
Voxel grid – lattice	$32 \times 32 \times 32$
Latent dimension – site VAE	200
Latent dimension – lattice VAE	25
Latent dimension – WGAN	200
Training structures	10,981
Generated structures	6,592

3.2 VAE Training Convergence

The fidelity of the inverse design pipeline depends on the VAE's ability to learn continuous, reconstructable representations of V–O crystal structures. Figure 4 presents the learning curves for both the lattice and site encoder–decoder pairs. The lattice VAE (Fig. 4(a)) converges rapidly during the first 50 epochs, with both training and test losses stabilizing below 1×10^{-3} by epoch 100 and reaching final values near 1×10^{-3} on a logarithmic scale. The close alignment between training (black) and test (orange) curves throughout the entire training process indicates minimal overfitting and strong generalization. The site VAE (Fig. 4(b)) exhibits a sharp initial transient followed by stable convergence near 4×10^{-3} by epoch 20, with training and test losses remaining tightly coupled thereafter. This consistent performance across both components confirms that the VAE successfully encodes crystal structures into a smooth latent manifold from which high-fidelity reconstructions of atomic positions and lattice geometry can be recovered.

3.3 WGAN Training Stability and Formation-Energy Constraint

A central motivation for adopting the Wasserstein framework over a conventional adversarial objective is the well-documented instability of standard GAN training—mode collapse, vanishing gradients, and generator–discriminator imbalance [45]. Figure 5 compares the training dynamics of both architectures over 5,000 epochs.

For the discriminator (Fig. 5(a)), the WGAN critic (green) achieves rapid convergence within approximately 500 epochs and maintains a stable loss near 0.1 for the remainder of training, whereas the conventional GAN discriminator (purple) exhibits persistent oscillations between 0.1 and 1.4 that never fully subside. The generator comparison (Fig. 5(b)) is even more revealing: the standard GAN generator loss oscillates between 2 and 16 across all 5,000 epochs—a hallmark of vanishing gradients and mode collapse—while the WGAN generator loss decreases smoothly to near zero within the first 500 epochs and remains stable thereafter. Table 2 quantifies this advantage after convergence: although discriminator losses are comparable (0.423 versus 0.418), the WGAN generator loss is 0.289 and standard GAN loss is 0.511.

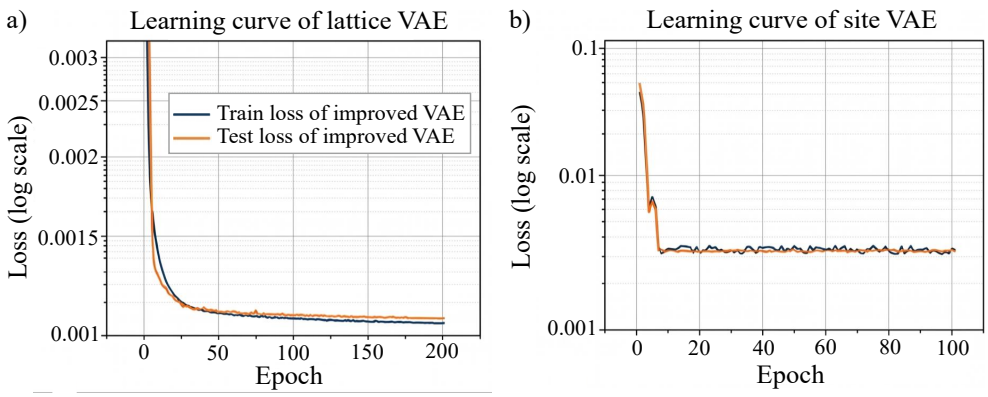


Fig. 4 Training and test loss curves for the two VAE components on a logarithmic scale. (a) The lattice encoder–decoder converges to $\sim 10^{-3}$ within ~ 50 epochs; (b) the site encoder–decoder converges to $\sim 4 \times 10^{-3}$ within ~ 20 epochs. The close overlap of training and test curves in both panels indicates minimal overfitting and stable reconstruction of atomic and lattice channels

Table 2 Converged loss values for GAN and WGAN after training

Model	Critic / D loss	Generator loss
GAN	0.423	0.511
WGAN	0.418	0.289

This stability arises from the Wasserstein critic objective and the composite generator loss defined in Section 2.4 (Eqs. 3 and 4). The unbounded critic score and the asymmetric exponential formation-energy penalty together provide informative, non-vanishing gradients to the generator throughout training, in contrast to the saturating sigmoid-based cross-entropy of the conventional GAN. The resulting smooth descent of the WGAN generator loss in Fig. 5(b) is the direct empirical signature of this design.

3.4 Generation Quality Metrics

To characterize the generative performance of the framework, we evaluated the standard validity–uniqueness–novelty (V/U/N) metric triplet [27] across the 6,592 generated V–O structures (Fig. 6(a)).

Structural validity reaches 81.5%, indicating that more than four out of five generated structures satisfy geometric feasibility criteria including absence of overlapping atoms, realistic interatomic distances, and physically plausible densities. Among valid structures, 96.6% are unique, confirming that the generator explores the latent space broadly with virtually no mode collapse. Novelty is 99.5%, meaning that nearly all unique valid structures differ from every entry in the training set; the model creates new compositions rather than memorizing known prototypes. XRD-fingerprint-based structural diversity reaches 0.973 on a $[0, 1]$ scale, indicating that the generated crystals

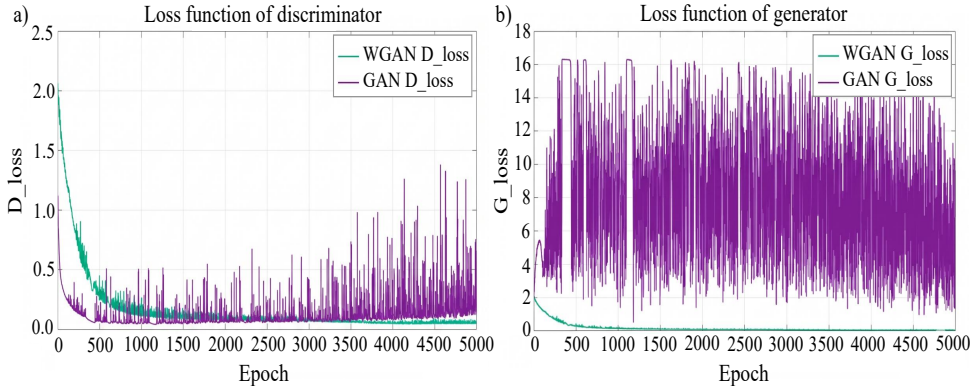


Fig. 5 GAN (purple) vs. WGAN (green) training over 5,000 epochs. (a) Discriminator/critic loss. (b) Generator loss. The WGAN converges stably in both, while the GAN oscillates persistently

span a broad range of structural topologies. Resolving these metrics by composition (Fig. 6(b)) confirms that generation quality is maintained across stoichiometries rather than being driven by a single dominant phase. Validity is higher for the most heavily represented compositions, VO_2 (94.9%) and V_2O_5 (93.3%), and somewhat lower for VO (82.9%) and V_2O_3 (85.5%), consistent with the greater difficulty of resolving well-separated atomic density peaks in the more metal-rich, densely packed VO and V_2O_3 lattices on a fixed voxel grid. Uniqueness and novelty remain at or near 100% for V_2O_5 and V_2O_3 and above 86% and 98% respectively for VO, indicating that the per-composition quality is not achieved by memorizing a narrow set of prototypes.

3.5 Latent-Space Organization

A key question for any generative model is whether the learned latent space encodes physically meaningful information or merely captures superficial statistical correlations. To investigate this, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) to the concatenated VAE latent representations of the training data. Each crystal was represented by concatenating its lattice (25-dimensional) latent code with the two 200-dimensional atomic-site latent codes (vanadium and oxygen) into a 425-dimensional vector, zero-padded to 784 dimensions, standardized to zero mean and unit variance, compressed to 50 principal components via PCA, and embedded in two dimensions (perplexity = 30, 1,500 iterations).A stratified sample of $n = 5{,}000$ structures was drawn from the full 10,981-structure training set to ensure representative coverage. Figure 7 presents the resulting embedding colored by three complementary physical descriptors.

The formation-energy panel (Fig. 7(a)) reveals that thermodynamically low-energy candidates within the predefined stability window ($E_f < 0$ eV/atom, dark purple) concentrate in a dense central cluster, while progressively less stable and metastable phases ($E_f > 0$, teal to yellow) radiate into distinct peripheral arms. Crucially, this spatial organization was not imposed during training, neither the VAE nor the WGAN received explicit stability labels as inputs. The oxygen-fraction panel (Fig. 7(b)) demonstrates that each radial arm carries a distinct color, meaning that the model

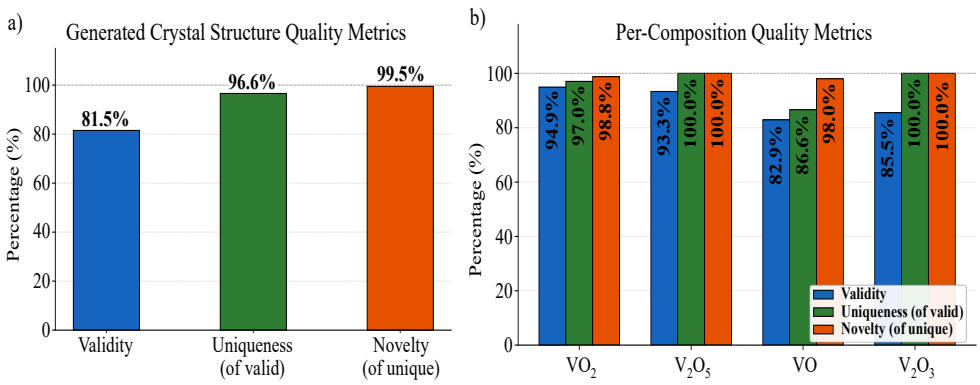


Fig. 6 Generation-quality metrics for the generated V–O structures. (a) Aggregate validity (81.5%), uniqueness among valid structures (96.6%), and novelty among unique structures (99.5%). (b) The same three metrics resolved by composition for VO_2, V_2O_5, VO, and V_2O_3, showing that generation quality is maintained across stoichiometries. The high uniqueness and novelty indicate broad latent-space exploration with negligible mode collapse or memorization of training prototypes

learned to separate different V:O stoichiometries into geometrically distinct regions without any compositional label being provided during training.

The residual-energy panel (Fig. 7(c)) provides a discriminating test. After removing a weak linear composition–energy baseline ($E_f = -2.65\,x_O + 0.11$, $R^2 = 0.29$), the remaining signal reflects differences in crystal topology, coordination environment, and bonding character that cannot be explained by stoichiometry alone. The moderate correlation between E_f and x_O (Pearson $r = -0.53$, Spearman $\rho = -0.58$) and the large residual standard deviation after detrending ($\sigma_{res} = 0.765$ eV/atom) confirm that the beyond-composition structural information encoded in the latent space is important. Because t-SNE preserves local neighborhoods rather than global geometry, Fig. 7 is interpreted only as qualitative evidence of local organization. To quantify the physical information accessible from the encoded representation, we trained five-fold cross-validated ridge-regression probes on the combined lattice and site representation. This representation predicts oxygen fraction with $R^2 = 0.939$ (MAE $= 0.027$) and formation energy with $R^2 = 0.766$ (MAE $= 0.269$ eV/atom), whereas oxygen fraction alone predicts formation energy with only $R^2 = 0.29$. The representation therefore carries linearly accessible structural and compositional information well beyond the weak composition-energy correlation. Because the WGAN samples from this organized latent space and the exponential energy penalty (Eq. (4)) further penalizes high-$\hat{E}_f$ regions, structures drawn from the high-density stable core are inherently biased toward low formation energies.

3.6 Thermodynamic Stability Analysis

DFT validation of all 6,592 generated compositions partitions the set by formation energy and convex-hull distance (Fig. 8). Of the full set, 1,846 structures (28.0%) have negative E_f and 4,746 (72.0%) have positive E_f. Within the negative-E_f population, 669 structures (10.2% of all generated) are stable, 365 (5.5%) are metastable,

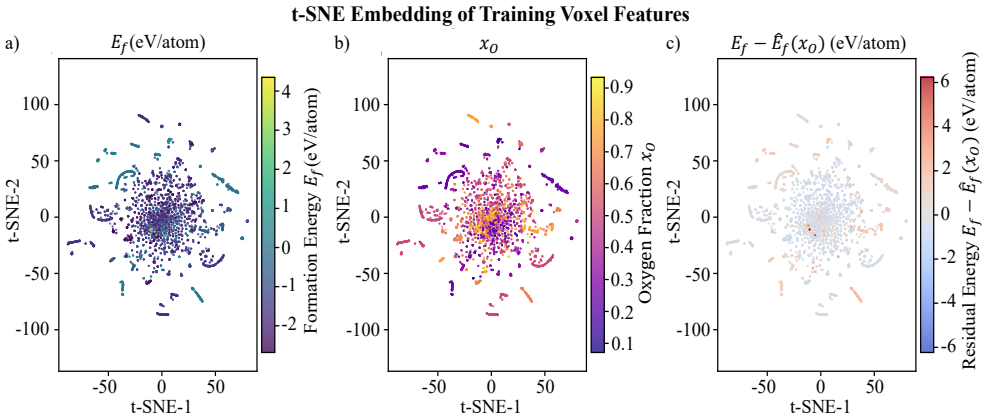


Fig. 7 t-SNE embedding of 5,000 V–O training structures, colored by (a) formation energy E_f, (b) oxygen fraction x_O, and (c) residual energy $E_f - \hat{E}_f(x_O)$ after removing the linear composition-energy trend. Low-energy structures concentrate in a central cluster and distinct stoichiometries separate into radial arms, indicating that the latent space organizes by stability and composition without explicit labels. The local color organization is consistent with composition- and energy-related information in the encoded representation

and 718 (10.9%) lie more than 500 meV/atom above the hull; a further 94 structures fall below the MP-referenced hull. The DFT-calculated formation energies of the full set are shown against oxygen fraction in Fig. 9, where the majority cluster near or below the MP convex hull with no systematic bias toward particular stoichiometries, confirming broad coverage from vanadium-rich to oxygen-rich regimes. Although coverage spans the full compositional range, the stable and below-hull candidates concentrate at intermediate-to-low oxygen fractions. This reflects the underlying V–O thermochemistry. The most thermodynamically favorable binary phases occur at intermediate vanadium oxidation states, where the convex hull is deepest, whereas fully oxidized V_2O_5 and highly reduced VO lie at shallower or higher points on the hull. The generator inherits this bias from the training distribution, which is itself densest at intermediate stoichiometries, and the formation-energy penalty further concentrates sampling toward these low-energy compositions.

Table 3 lists the lowest-formation-energy structure identified in the generated data for selected V–O stoichiometries. The most negative formation energies occur at intermediate oxygen fractions (V_2O_3, -2.546; V_3O_4,-2.534 eV/atom). This behavior is consistent with the weak population-level composition-energy trend ($R^2 = 0.29$), indicating that maximal thermodynamic favorability arises at intermediate oxidation rather than at the oxygen-rich extreme.

The relevant measure of generative performance is not the raw stable fraction but the yield of stable structures that are new. The training corpus consists of known V–O phases and their direct substitutional derivatives and therefore contains, by construction, no structures absent from existing databases. Of the 669 stable candidates generated here, 99.5% are novel with respect to the training set, so the framework does not recapitulate its input distribution but produces previously unseen structures that satisfy the same first-principles stability criterion.

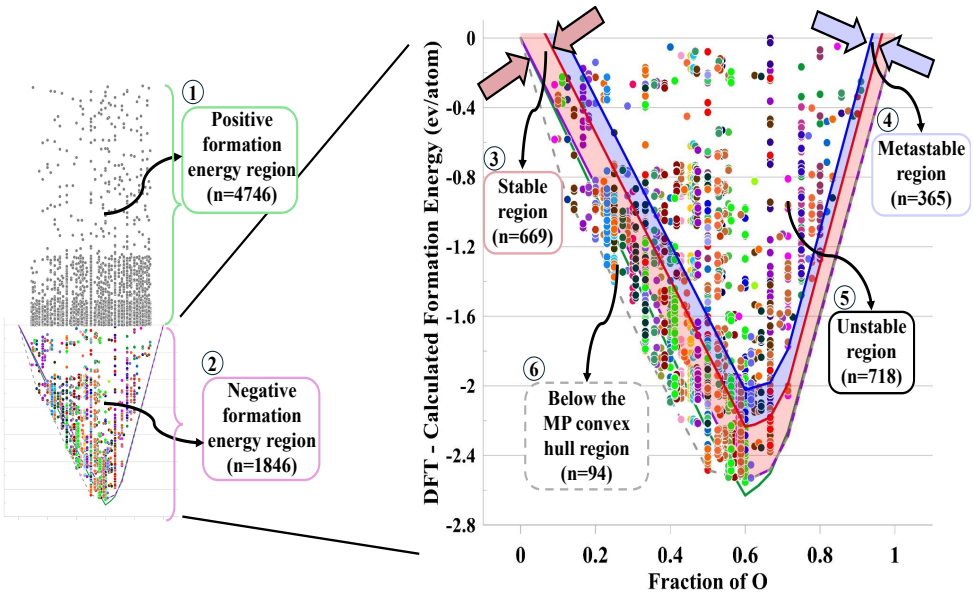


Fig. 8 Thermodynamic classification of the 6,592 generated V–O structures. (Left) Partition by sign of the DFT formation energy: 4,746 structures have positive E_f (gray) and 1,846 have negative E_f (colored); positive-E_f structures are thermodynamically unfavorable and are not classified further. (Right) The negative-E_f population resolved relative to the MP convex hull into stable $0 \leq \Delta E_{\text{hull}} \leq 300$ meV/atom; $n = 669$), metastable ($300 < \Delta E_{\text{hull}} \leq 500$ meV/atom; $n = 365$), unstable ($\Delta E_{\text{hull}} > 500$ meV/atom; $n = 718$), and below-hull ($\Delta E_{\text{hull}} < 0$; $n = 94$) regions, which together account for all 1,846 negative-E_f structures. Shaded bands mark the 300 and 500 meV/atom windows

Table 3 Reference V–O structures from the generated data with the lowest formation energy for selected stoichiometries

Composition	E_f (eV/atom)	N_{atom}	Density ρ (g/cm^3)
VO_2	−2.384	6	4.66
V_2O_5	−2.098	7	1.89
V_3O_5	−2.370	16	4.97
V_2O_3	−2.546	10	5.23
V_3O_4	−2.534	14	5.43
VO	−2.480	6	5.54
V_3O_2	−2.092	20	5.46

3.7 Effect of the Formation-Energy Constraint on DFT-Validated Stability

The penalty-weight sweep of Section 2.4 evaluates candidate stability using the frozen regressor, which is also the source of the penalty term in Eq. (4). To quantify the effect of the constraint on first-principles stability directly, we performed a composition-matched DFT control in which stability is judged solely from DFT formation energies

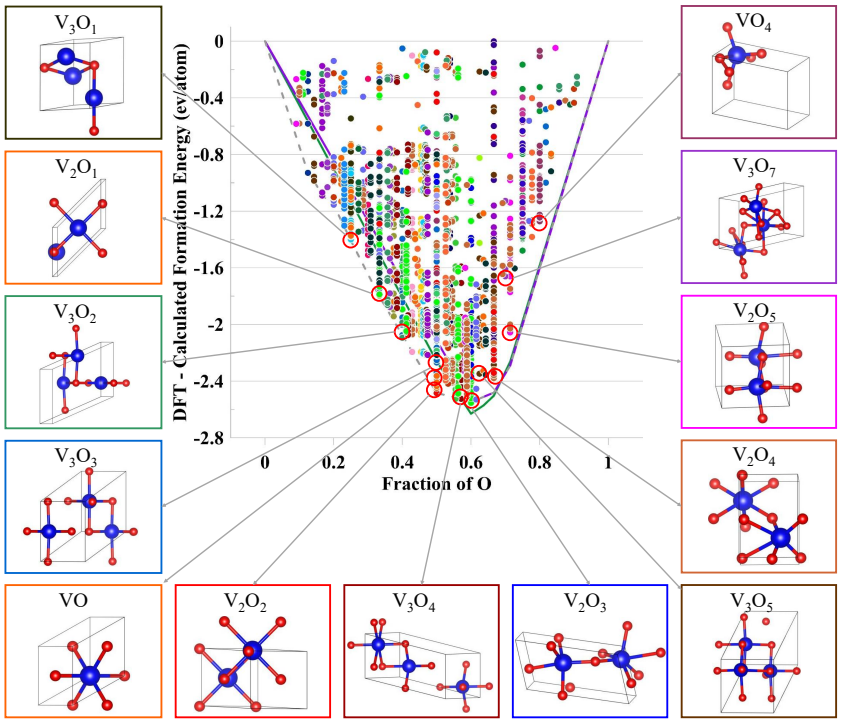


Fig. 9 DFT-calculated formation energies of all 6,592 generated V–O compositions versus oxygen fraction. Solid purple and green curves denote the MP and OQMD convex hulls, respectively. Insets show representative lowest-energy generated structures spanning vanadium-rich to oxygen-rich stoichiometries (V in blue, O in red); red circles mark the corresponding points on the scatter. Of the full set, 1,846 structures have negative E_f and 4,746 have positive E_f. The solid lines denote the MP and OQMD convex hulls

and MP convex-hull distances, independent of the regressor. Two representative stoichiometries, VO_2 and V_2O_3, were selected for DFT tractability, and a matched subsample of 232 generated structures was drawn for each setting and relaxed with the identical PBE+U protocol of Section 2.1.

The results are summarized in Table 4. With the penalty ($\lambda = 0.1$), 129 of 232 structures (55.6%) exhibit negative formation energy, of which 59 (25.4% of the subsample) are stable and 29 (12.5%) metastable. Without the penalty ($\lambda = 0$), 117 structures (50.4%) have negative formation energy, with 37 (15.9%) stable and 22 (9.5%) metastable. The penalty therefore raises the DFT-confirmed stable fraction from 15.9% to 25.4%, a relative increase of approximately 60% within a composition-matched set.

The mechanism is clarified by conditioning on the sign of the formation energy. The penalty increases the fraction of structures with $E_f < 0$ only modestly (50.4% to 55.6%), but among those structures it raises the stable fraction from 31.6% to 45.7%. The exponential penalty thus does not merely drive formation energies below zero; it shifts the generated distribution toward lower formation energy in the vicinity of the

Table 4 Composition-matched DFT control of the formation-energy penalty, restricted to the VO_2 and V_2O_3 compositions ($n = 232$ generated structures per setting)

Setting	$E_f < 0$	Stable	Metastable	Stable | $E_f < 0$
$\lambda = 0$ (no penalty)	117 (50.4%)	37 (15.9%)	22 (9.5%)	31.6%
$\lambda = 0.1$ (production)	129 (55.6%)	59 (25.4%)	29 (12.5%)	45.7%

MP convex hull, which is precisely the asymmetric bias that Eq. (4) is designed to impose. Because this control is restricted to two intermediate-oxygen-fraction compositions lying near the hull minimum, its stable fractions are higher than the full-set average reported in Section 3.6; the physically meaningful quantity is the λ-dependent difference within the matched set rather than the absolute values.

3.8 Dynamic Stability Assessment

Thermodynamic favorability alone does not guarantee structural viability, lattice-dynamical stability must also be assessed. Phonon calculations are the most expensive step of the workflow and were therefore performed on selected candidates rather than on the full validated set. Figure 10 presents three representative case studies, one drawn from each thermodynamic class defined in Fig. 8 (an unstable structure, a stable structure, and a structure lying below the MP-referenced hull) chosen so that the three classes can be compared on a common footing. The three cases and their key phonon characteristics are summarized in Table S2.

The unstable V_2O_5 structure (Fig. 10(a)), computed with a $3 \times 3 \times 2$ supercell, exhibits extensive imaginary frequencies extending to approximately -20 THz across multiple high-symmetry paths. The pervasive nature and large magnitude of these imaginary modes indicate genuine dynamic instability. Such structures are expected to relax toward lower-symmetry configurations or dynamically more stable polymorphs and are therefore excluded from the stable candidate set.

In contrast, the stable VO_2 structure (Fig. 10(b)), computed with a $5 \times 5 \times 4$ supercell, displays a predominantly positive phonon spectrum throughout the Brillouin zone, with frequencies extending to approximately 23 THz consistent with V–O bond stretching modes. Minor imaginary branches with magnitudes below 2 THz appear in isolated high-symmetry directions. These shallow soft modes are attributable to finite-size supercell effects inherent to the displacement method or to the well-documented rutile-to-monoclinic structural transition of VO_2 near 340 K [3, 48]. Minor imaginary frequencies of this character presage temperature-driven phase transitions rather than fundamental structural failure and do not disqualify the structure from the stable candidate set.

The third case is a candidate at the VO stoichiometry (a four-atom V_2O_2 cell) drawn from the sub-hull population of Section 3.6. Its dispersion (Fig. 10(c)) is entirely positive throughout the sampled Brillouin zone. the three acoustic branches disperse linearly from Γ with no imaginary excursion along any high-symmetry direction, and the optical manifold extends to ≈ 13.5 THz. The phonon density of states (Fig. S6) carries no spectral weight at negative frequency, confirming that no soft mode is missed

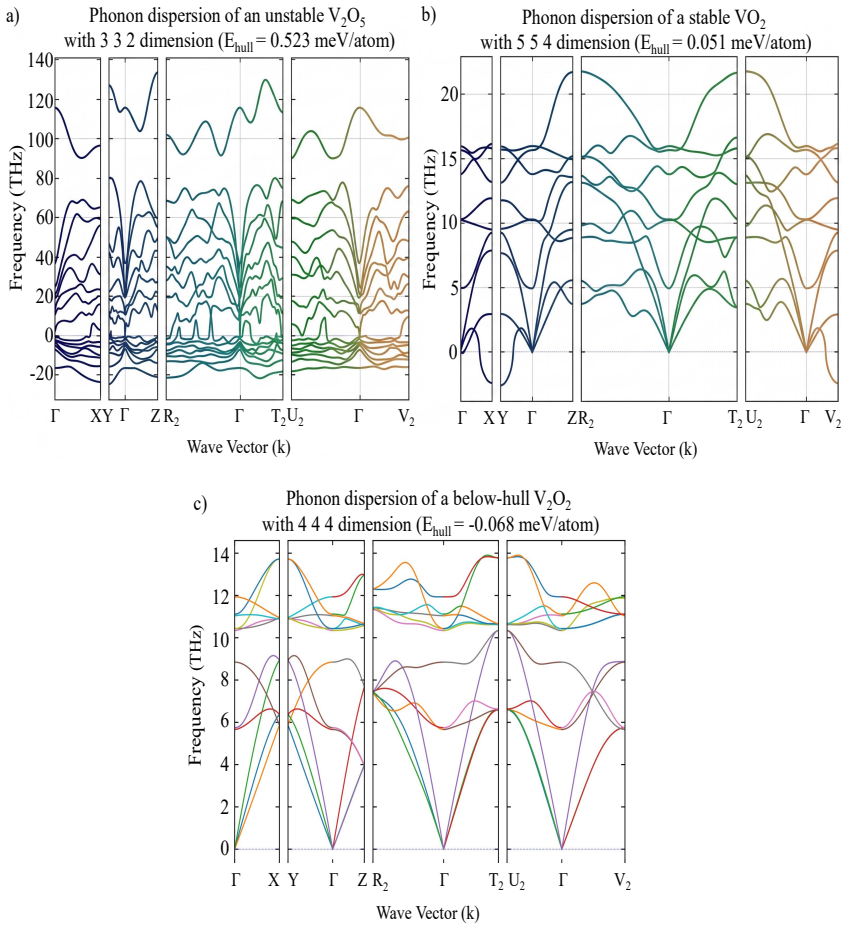


Fig. 10 Phonon dispersions for three generated candidates spanning the three thermodynamic classes of Fig. 8, computed by the finite-displacement method. (a) An unstable V_2O_5 candidate ($3 \times 3 \times 2$ supercell): imaginary branches extending to ~ -20 THz across multiple high-symmetry paths, indicating genuine lattice-dynamical instability. (b) A stable VO_2 candidate ($5 \times 5 \times 4$ supercell): a predominantly positive spectrum extending to ~ 23 THz, with shallow imaginary branches (< 2 THz) confined to isolated high-symmetry directions. (c) A below-hull V_2O_2 candidate ($4 \times 4 \times 4$ supercell): a fully positive spectrum throughout the sampled Brillouin zone, with no imaginary frequencies

by the finite sampling of the band path. This structure is therefore dynamically stable in the harmonic approximation, and unlike the VO_2 case above, requires no appeal to finite-size effects or to a known phase transition. The harmonic thermodynamic functions (Fig. S7) are correspondingly well-behaved. The Helmholtz free energy decreases monotonically from its zero-point value, the entropy increases monotonically, and the constant-volume heat capacity saturates near 100 J K^{-1} mol^{-1}.

The combined thermodynamic and dynamic screening yields a curated set of V–O compositions possessing both energetic favorability and lattice robustness. The broad stoichiometric range of stable and metastable structures, spanning oxygen-poor to

oxygen-rich compositions, provides a candidate pool for subsequent property-targeted screening. Because vanadium oxides are known to exhibit metal–insulator transitions, tunable optical response, electrochemical activity, and memristive behavior, the generated structures may be relevant to future studies of thermochromic materials, electrochromic coatings, energy-storage electrodes, catalytic oxides, and reconfigurable electronic or neuromorphic devices. However, the present work focuses on structural generation and stability validation; application-specific properties such as optical constants, ionic transport, catalytic activity, or device-level switching behavior remain targets for future investigation.

4 Conclusion

This work presented a physics-guided inverse-design framework for discovering stable V–O crystals by integrating a voxel-based VAE with a formation-energy-constrained WGAN. The voxel representation enables fixed-size, element-resolved encoding of atomic environments and lattice geometry, while the Wasserstein objective provides stable gradients during adversarial training. By embedding an exponential formation-energy penalty through a frozen CNN regressor, the generator is explicitly steered toward thermodynamically favorable regions of latent space without requiring explicit stability labels.

Applied to the V–O binary system, the framework produced 6,592 candidate structures with high generative quality (81.5% validity, 96.6% uniqueness, and 99.5% novelty). Of these, 1,846 have negative formation energies and 4,746 have positive formation energies. Consistent DFT validation identified 669 candidates (10.2%) that meet strict stability requirements ($E_f < 0$ and ≤ 300 meV/atom above the MP convex hull), plus 365 additional candidates (5.5%) within a metastability window of ≤ 500 meV/atom. An additional 94 candidates fall below the MP-referenced hull and are flagged as candidates for further validation. Phonon dispersion calculations on representative low-energy candidates further support dynamical stability, with shallow imaginary modes attributable to finite supercell effects or known temperature-driven transitions. These phonon calculations were performed on three representative candidates, one from each thermodynamic class.

These results demonstrate that enforcing thermodynamic bias directly in adversarial training can increase the yield of stable candidates and accelerate discovery in chemically complex oxide spaces. Current limitations include discretization-induced symmetry breaking in fixed voxel grids, the lack of explicit space-group constraints during generation, and implicit rather than explicit encoding of oxidation states. The penalty-weight ablation additionally relies on a regressor-based stability surrogate that inherits the regressor's prediction error and was confirmed against full DFT only at the production setting. Future work will incorporate symmetry-aware encodings, oxidation-state or charge descriptors, and extension to ternary oxides to enable property-targeted generation of multifunctional materials for thermal, electronic, electrochemical, catalytic, and neuromorphic applications. More broadly, the

results demonstrate how thermodynamic priors can be embedded directly into generative models to improve the efficiency of first-principles materials discovery in chemically complex oxide systems.

Supplementary information. Supplementary information accompanies this manuscript, including the Web of Science literature analysis (Figs. S1–S4) and the formation-energy penalty-weight sweep (Table S1).

Acknowledgements. The authors acknowledge the computational resources and software tools used for first-principles calculations, structure manipulation, and post-processing.

Declarations

Funding. This work was supported in part by the U.S. National Science Foundation under Award No. 2432082 and the Air Force Office of Scientific Research under Award No. FA9550-25-1-0322. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the funding agencies.

Competing interests. The authors declare no competing interests.

Ethics approval and consent to participate. Not applicable.

Consent for publication. Not applicable.

Data availability. The generated V–O crystal structures in VASP POSCAR format are publicly available at https://github.com/INQUIRELAB/VAE-WGAN-VanadiumOxides.

Materials availability. Not applicable.

Code availability. All implementation code for this work is publicly available at https://github.com/INQUIRELAB/VAE-WGAN-VanadiumOxides.

Author contribution. D.E. designed and implemented the generative framework, performed the simulations and analysis, and drafted the manuscript. S.S.S. and Y.M.B. supervised the study, contributed to the methodology and interpretation, and revised the manuscript. All authors read and approved the final manuscript.

References

[1] Hu P, Hu P, Vu TD, Li M, Wang S, Ke Y, et al. Vanadium oxide: phase diagrams, structures, synthesis, and applications. Chemical Reviews. 2023;123(8):4353–4415. https://doi.org/10.1021/acs.chemrev.2c00546.

[2] Liu K, Lee S, Yang S, Delaire O, Wu J. Recent progresses on physics and applications of vanadium dioxide. Materials Today. 2018;21(8):875–896. https://doi.org/10.1016/j.mattod.2018.03.029.

[3] Qazilbash MM, Brehm M, Chae BG, Ho PC, Andreev GO, Kim BJ, et al. Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging. Science. 2007;318(5857):1750–1753. https://doi.org/10.1126/science.1150124.

[4] Le TK, Pham PV, Dong CL, Bahlawane N, Vernardou D, Mjejri I, et al. Recent advances in vanadium pentoxide (V_2O_5) towards related applications in chromogenics and beyond. Journal of Materials Chemistry C. 2022;10(11):4019–4071. https://doi.org/10.1039/D1TC04872D.

[5] Granqvist CG. Transparent conductors as solar energy materials: a panoramic review. Solar Energy Materials and Solar Cells. 2007;91(17):1529–1598. https://doi.org/10.1016/j.solmat.2007.04.031.

[6] Seo D, Kim D, Ryu J, Pyo C, Lee S, Yoon TS, et al. VOx-based nonvolatile radio-frequency switches for reconfigurable filter. Advanced Science. 2025;12(31):e2501989. https://doi.org/10.1002/advs.202501989.

[7] Yi W, Tsang KK, Lam SK, Bai X, Crowell JA, Flores EA. Biological plausibility and stochasticity in scalable VO_2 active memristor neurons. Nature Communications. 2018;9(1):4661. https://doi.org/10.1038/s41467-018-07052-w.

[8] Snook GA, Kao P, Best AS. Conducting-polymer-based supercapacitor devices and electrodes. Journal of Power Sources. 2011;196(1):1–12. https://doi.org/10.1016/j.jpowsour.2010.06.084.

[9] Silversmit G, Depla D, Poelman H, Marin GB, De Gryse R. Determination of the V 2p XPS binding energies for different vanadium oxidation states (V^{5+} to V^{0+}). Journal of Electron Spectroscopy and Related Phenomena. 2004;135(2–3):167–175. https://doi.org/10.1016/j.elspec.2004.03.004.

[10] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547–555. https://doi.org/10.1038/s41586-018-0337-2.

[11] Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. APL Materials. 2013;1(1):011002. https://doi.org/10.1063/1.4812323.

[12] Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. npj Computational Materials. 2015;1(1):15010. https://doi.org/10.1038/npjcompumats.2015.10.

[13] Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, Taylor RH, et al. AFLOW: an automatic framework for high-throughput materials discovery. Computational Materials Science. 2012;58:218–226. https://doi.org/10.1016/j.

commatsci.2012.02.005.

[14] Jain A, Hautier G, Moore CJ, Ong SP, Fischer CC, Mueller T, et al. A high-throughput infrastructure for density functional theory calculations. Computational Materials Science. 2011;50(8):2295–2310. https://doi.org/10.1016/j.commatsci.2011.02.023.

[15] Zunger A. Inverse design in search of materials with target functionalities. Nature Reviews Chemistry. 2018;2(4):0121. https://doi.org/10.1038/s41570-018-0121.

[16] Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: generative models for matter engineering. Science. 2018;361(6400):360–365. https://doi.org/10.1126/science.aat2663.

[17] Chang Y, Wang H, Dong Q. Machine learning-based inverse design of auxetic metamaterial with zero Poisson's ratio. Materials Today Communications. 2022;30:103186. https://doi.org/10.1016/j.mtcomm.2022.103186.

[18] Wang Y, Lv J, Zhu L, Ma Y. CALYPSO: a method for crystal structure prediction. Computer Physics Communications. 2012;183(10):2063–2070. https://doi.org/10.1016/j.cpc.2012.05.008.

[19] Lyakhov AO, Oganov AR, Stokes HT, Zhu Q. New developments in evolutionary structure prediction algorithm USPEX. Computer Physics Communications. 2013;184(4):1172–1182. https://doi.org/10.1016/j.cpc.2012.12.009.

[20] Zuo Y, Qin M, Chen C, Ye W, Li X, Luo J, et al. Accelerating materials discovery with Bayesian optimization and graph deep learning. Materials Today. 2021;51:126–135. https://doi.org/10.1016/j.mattod.2021.08.012.

[21] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444. https://doi.org/10.1038/nature14539.

[22] Fang J, Xie M, He X, Zhang J, Hu J, Chen Y, et al. Machine learning accelerates the materials discovery. Materials Today Communications. 2022;33:104900. https://doi.org/10.1016/j.mtcomm.2022.104900.

[23] Kingma DP, Welling M.: Auto-encoding variational Bayes. Preprint at https://doi.org/10.48550/arXiv.1312.6114.

[24] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Advances in Neural Information Processing Systems. 2014;27. https://doi.org/10.48550/arXiv.1406.2661.

[25] Court CJ, Yildirim B, Jain A, Cole JM. 3-D inorganic crystal structure generation and property prediction via representation learning. Journal of Chemical Information and Modeling. 2020;60(10):4518–4535. https://doi.org/10.1021/acs.

jcim.0c00464.

[26] Ren Z, Tian SIP, Noh J, Oviedo F, Xing G, Li J, et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. Matter. 2022;5(1):314–335. https://doi.org/10.1016/j.matt.2021.11.032.

[27] Xie T, Fu X, Ganea OE, Barzilay R, Jaakkola TS. Crystal diffusion variational autoencoder for periodic material generation. In: International Conference on Learning Representations (ICLR); 2022. .

[28] Zeni C, Pinsler R, Zügner D, Fowler A, Horton M, Fu X, et al. A generative model for inorganic materials design. Nature. 2025;639(8055):624–632. https://doi.org/10.1038/s41586-025-08628-5.

[29] Alverson M, Baird SG, Murdock R, Johnson J, Sparks TD. Generative adversarial networks and diffusion models in material discovery. Digital Discovery. 2024;3(1):62–80. https://doi.org/10.1039/D3DD00137G.

[30] Noh J, Kim J, Stein HS, Sanchez-Lengeling B, Gregoire JM, Aspuru-Guzik A, et al. Inverse design of solid-state materials via a continuous representation. Matter. 2019;1(5):1370–1384. https://doi.org/10.1016/j.matt.2019.08.017.

[31] Jiao R, Huang W, Lin P, Han J, Chen P, Lu Y, et al. Crystal structure prediction by joint equivariant diffusion. Advances in Neural Information Processing Systems. 2023;36:17464–17497. https://doi.org/10.48550/arXiv.2309.04475.

[32] Zhao Y, Siriwardane EMD, Wu Z, Fu N, Al-Fahdi M, Hu M, et al. Physics guided deep learning for generative design of crystal materials with symmetry constraints. npj Computational Materials. 2023;9(1):38. https://doi.org/10.1038/s41524-023-00987-9.

[33] Long T, Fortunato NM, Opahle I, Zhang Y, Samathrakis I, Shen C, et al. Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. npj Computational Materials. 2021;7(1):66. https://doi.org/10.1038/s41524-021-00526-4.

[34] Kim S, Noh J, Gu GH, Aspuru-Guzik A, Jung Y. Generative adversarial networks for crystal structure prediction. ACS Central Science. 2020;6(8):1412–1420. https://doi.org/10.1021/acscentsci.0c00426.

[35] Ebrahimzadeh D, Sharif SS, Banad YM. Accelerated discovery of vanadium oxide compositions: a WGAN-VAE framework for materials design. Materials Today Electronics. 2025;13:100155. https://doi.org/10.1016/j.mtelec.2025.100155.

[36] Hasan M, Eger ZE, Senthilnathan A, Acar P. Microstructure-sensitive deformation modeling and materials design with physics-informed neural networks. AIAA

Journal. 2024;62(5):1864–1874. https://doi.org/10.2514/1.J062708.

[37] Xia Z, Sun X, Wang Z, Meng J, Jin B, Wang T. Low-power memristor for neuromorphic computing: from materials to applications. Nano-Micro Letters. 2025;17(1):217. https://doi.org/10.1007/s40820-025-01705-4.

[38] Ganduglia-Pirovano MV, Hofmann A, Sauer J. Oxygen vacancies in transition metal and rare earth oxides: current state of understanding and remaining challenges. Surface Science Reports. 2007;62(6):219–270. https://doi.org/10.1016/j.surfrep.2007.03.002.

[39] Hoffmann J, Maestrati L, Sawada Y, Tang J, Sellier JM, Bengio Y.: Data-driven approach to encoding and decoding 3-D crystal structures. Preprint at https://doi.org/10.48550/arXiv.1909.00949.

[40] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning (PMLR); 2017. p. 214–223.

[41] Kresse G, Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B. 1996;54(16):11169–11186. https://doi.org/10.1103/PhysRevB.54.11169.

[42] Munro JM, Latimer K, Horton MK, Dwaraknath S, Persson KA. An improved symmetry-based approach to reciprocal space path selection in band structure calculations. npj Computational Materials. 2020;6(1):112. https://doi.org/10.1038/s41524-020-00383-7.

[43] Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Physical Review Letters. 2018;120(14):145301. https://doi.org/10.1103/PhysRevLett.120.145301.

[44] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–778.

[45] Wiatrak M, Albrecht SV, Nystrom A.: Stabilizing generative adversarial networks: a survey. Preprint at https://doi.org/10.48550/arXiv.1910.00927.

[46] Togo A, Chaput L, Tadano T, Tanaka I. Implementation strategies in Phonopy and Phono3py. Journal of Physics: Condensed Matter. 2023;35(35):353001. https://doi.org/10.1088/1361-648X/acd831.

[47] Togo A. First-principles phonon calculations with Phonopy and Phono3py. Journal of the Physical Society of Japan. 2023;92(1):012001. https://doi.org/10.7566/JPSJ.92.012001.

[48] Morin FJ. Oxides which show a metal-to-insulator transition at the Néel temperature. Physical Review Letters. 1959;3(1):34–36. https://doi.org/10.1103/PhysRevLett.3.34.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf