

MagGen: A Graph-Aided Deep Generative Model for Inverse Design of Permanent Magnets

Sourav Mal, Gaurav Seal, and Prasenjit Sen*



Cite This: *J. Phys. Chem. Lett.* 2024, 15, 3221–3228



Read Online

ACCESS |



Metrics & More

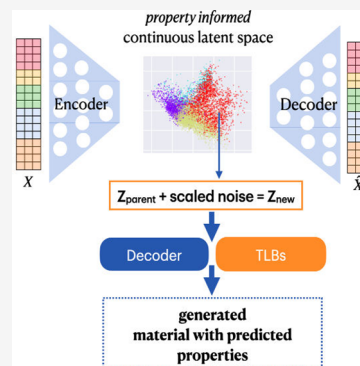


Article Recommendations



Supporting Information

ABSTRACT: A deep generative model based on a variational autoencoder (VAE), conditioned simultaneously by two target properties, is developed to inverse design stable magnetic materials. The structure of the physics-informed, property embedded latent space of the model is analyzed using graph theory. An impressive $\sim 96\%$ of the generated materials are found to satisfy the target properties as per predictions from the target-learning branches. This is a huge improvement over approaches that do not condition the VAE latent space by target properties or that do not consider the connectivity of the parent materials from which the new materials are generated. This impressive feat is achieved by using a simple real-space-only representation that can be directly read from material cif files. Model predictions are finally validated by density functional theory calculations on a randomly chosen subset of materials. The performance of the present model is comparable or superior to that of models reported earlier. This model (MagGen) is applied to the problem of designing rare earth-free permanent magnets with promising results.



Designing new materials with desired properties is one of the grand challenges in materials research. In the recent past, high-throughput virtual screening (HTVS)^{1–10} has achieved some success in discovering new materials. HTVS screens an initial pool of materials, usually generated via chemical intuition or other heuristic methods, by using density functional theory (DFT) calculations in a high-throughput mode. More recently, machine learning (ML)-based property prediction models^{11,12} have greatly accelerated this screening step. Despite its successes, the method is limited to exploring only a small subspace of the material space as represented by the initial pool and is also not efficient for property driven materials design.

Deep learning-based generative ML models (GMs) have made possible the efficient exploration of the vast material space, enabling the long-standing goal of inverse design. Inverse design in essence means inverting the usual structure and composition to property mapping and deriving structure and composition given a desired set of properties.

GMs create a continuous representation of materials in a space called the latent space. This can be embedded with properties during the training process, and new materials can be generated by sampling points in this space. Although GMs have been successful in inverse designing molecules,^{13–16} they face considerable challenge in the case of crystalline materials due to the absence of an invertible representation of the latter.

The two most common invertible representations used to inverse design crystalline materials are based on 3D voxel representation^{17–21} and 2D point cloud representation.^{22–24} In spite of their success in specific situations, all of these representations are restricted to encode either materials with a

fixed composition^{17–19,22,23} or those with a fixed crystal structure.^{20,21,24,25} Moreover, training generative models with voxel representation is highly memory intensive.

Recently, Ren et al.²⁶ proposed an invertible point cloud representation called Fourier-transformed crystal properties (FTCP), which can encode crystalline materials with different compositions and structures using a combination of real space and reciprocal space information. It was used to inverse design inorganic materials with user-defined formation energies, bandgap, thermoelectric power factor, and combinations thereof.

In this paper, we present an efficient generative model using only real space features to inverse design stable ferromagnetic materials with large saturation magnetization (M_s) for permanent magnet application. This representation is termed invertible real space crystallographic representation (IRCR).

The design of materials with large saturation magnetization and magnetic anisotropy energy (AE), particularly those without rare earth (RE) elements, is a challenging and important problem.²⁷ The target material properties are $M_s \geq 1 \text{ T}$ ²⁸ and a uniaxial AE with an anisotropy constant of $K_1 > 1 \text{ MJ/m}^3$.²⁸

Received: January 7, 2024

Revised: March 3, 2024

Accepted: March 5, 2024

Published: March 14, 2024



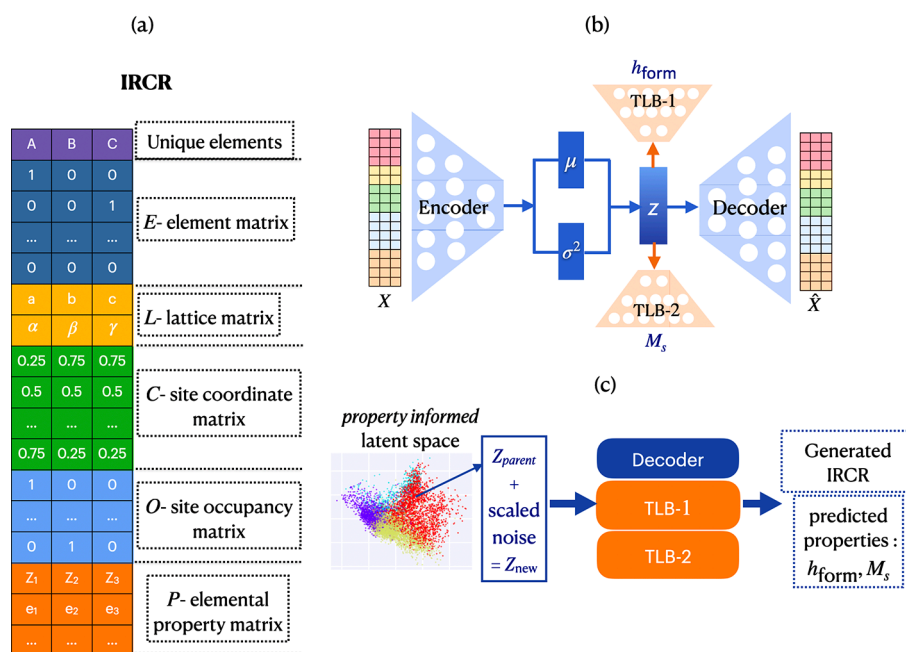


Figure 1. (a) Invertible real space crystallographic representation (IRCR) based on only real space features. (b) Schematic diagram of MagGen consisting of an encoder, a decoder, and two target-learning branches. (c) Scheme for new material generation with predicted h_{form} and M_s .

The design of such materials has been addressed by us²⁹ and other authors.^{28,30–37} These works used either high-throughput DFT calculations or ML-assisted screening to identify candidate materials, followed by validation with DFT calculations. So far, generative models have not been employed to address the problem, primarily due to the paucity of materials data for AE. Given this constraint, we aim to design stable ferromagnetic materials using a GM and subsequently screen these for AE via DFT. We deliberately set a lower threshold of $M_s \geq 0.5$ T during the generation process for convenience in this first magnetic materials generation model with magnetization as one of the targets. But as we present later, we do generate materials having $M_s > 1$ T that also satisfy the AE target.

We use a conditional variational autoencoder (cVAE)^{38,39} as the generative model, in which the latent space is conditioned by formation energy (h_{form}) and M_s simultaneously. h_{form} is a proxy for the stability of a material with respect to its isolated constituents. The distance from the convex hull (E_{hull}) is calculated for only the newly designed materials. This framework for stable magnetic materials generation is called MagGen. To the best of our knowledge, this is the first attempt to inverse design permanent magnet materials using any GM and only the second work after ref 26 attempting to organize the materials latent space with more than one target property simultaneously. Very recently, inverse design models have been used to discover new 2D magnetic materials.⁴⁰

A schematic diagram of IRCR for materials is given in Figure 1a. It consists of five matrices concatenated vertically:

$$\text{IRCR} = \{E, L, C, O, P\} \quad (1)$$

in which E , L , C , O , and P are matrices carrying information about the constituent elements, lattice structure, atomic basis, site occupancy, and elemental chemical properties of the constituent atoms, respectively. These are discussed in Section S1 of the Supporting Information (SI).

MagGen (schematic diagram shown in Figure 1b) consists of an encoder, a decoder, and two target-learning branches (TLBs). The encoder uses IRCR as input (X) and maps it to an isotropic Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ in a lower-dimensional latent space. A point z is sampled from the latent space according to eq 2:

$$z = \mu + \sigma \odot \epsilon, \epsilon \sim \mathcal{N}(0, I) \quad (2)$$

where ϵ is a random noise sampled from a unit Gaussian and \odot denotes element-wise multiplication. The decoder maps z back to the original input, called the reconstructed output (\hat{X}).

The latent space is regularized by enforcing the encoding distributions to follow a unit Gaussian prior. Any deviation from the prior condition is expressed as the Kullback–Leibler (KL) divergence between the encoding distribution and the prior. It enforces the encoded distributions to overlap as much as possible, centered around the origin, and ensures the creation of a continuous latent space.^{38,39}

The two TLBs map the latent points to the physical properties of interest. The branches predicting h_{form} and M_s are termed TLB-1 and TLB-2, respectively. This combination of regularization and property embedding builds in a structure in the latent space that ensures that latent points close to each other lead to similar properties, as is established in greater detail later. We call this space the “property informed continuous latent space”. The model architecture, loss function, and hyperparameters are discussed in Section S2 in the SI.

Training data include inorganic bulk materials taken from three publicly available databases: the Materials Project (MP),⁴¹ novomag,³³ and Novomag.²⁸ DFT-computed materials properties are listed in these databases. We include up to ternary materials with the number of sites in the unit cell (n_{sites}) up to 20 and lattice constants less than or equal to 25 Å. However, MagGen is not limited to these choices and can be extended beyond these limits if required.

The data set, discussed in Section S3 in the SI, is split in an 80:20 ratio to create training and test sets. The performance of MagGen is assessed by analyzing the reconstruction performance of the decoder, the property prediction accuracy of the two TLBs, and the organization of the latent space for the test set materials. The first two are summarized in Table 1. We compare our results with those of Ren et al.²⁶ wherever they overlap.

Table 1. Reconstruction Performance of the Decoder and Property Prediction Performances of the Two Target-Learning Branches of MagGen Using IRCR; Similar Quantities for FTCP Are Mentioned for Comparison

component	quantity	IRCR	FTCP
decoder	accuracy of constituent elements (%)	98.64	99.0
	MAPE of lattice constants (<i>abc</i>) (%)	11.74	9.01
	MAE of lattice constants (<i>abc</i>) (Å)	0.67	-
	MAPE of lattice angles ($\alpha\beta\gamma$) (%)	2.23	5.07
	MAE of lattice angles ($\alpha\beta\gamma$) (deg)	2.09	-
	MAE of site fractional coordinates (a.u.)	0.024	0.045
TLB-1	MAE of h_{form} (eV/atom)	0.162	0.051
	R^2 score of h_{form}	0.93	-
TLB-2	MAE of M_s (T)	0.07	N/A
	R^2 score of M_s	0.87	N/A

The accuracy of reproducing the constituent elements is 98.64%. The mean absolute error (MAE) of the lattice constants is <0.7 Å, and that of the lattice angles is only 2°. The site fractional coordinates are reconstructed with an MAE of 0.024. The lattice angles and site fractional coordinates are reconstructed much more accurately with IRCR compared to FTCP. Only the mean absolute percentage error (MAPE) of lattice constant reconstruction is marginally lower compared to FTCP. Overall, the reconstruction performance of the encoder–decoder combination is good and comparable to or better than what has been achieved so far.

The MAE of h_{form} prediction by TLB-1 is 0.162 eV/atom, and the corresponding R^2 score is 0.93. The performance of the TLB for h_{form} prediction in ref 26 is somewhat better than our model with an MAE of 0.051 eV/atom. One reason for this is perhaps that we have included all materials, both stable and unstable, in the training process. In ref 26, only materials lying up to 80 meV/atom of the convex hull are included. For the M_s prediction by TLB-2, the MAE is 0.07 T, and the R^2 score is 0.87. The performance is much better than other models for M_s prediction.^{29,30,42}

We also show the parity plots (actual versus predicted property values) of h_{form} and M_s predictions in Figure 2a,b. The parity plot of h_{form} demonstrates a balanced prediction over the entire range of values. The plot of M_s shows an overall underestimation of M_s but a slight overestimation at low M_s values. Similar issues with the performance of magnetic moment predictors have been reported in the literature.^{29,30,42} Compared to these works, the performance of the TLB-2 is better.

Next, we investigate the structure of the latent space to analyze how it is organized according to property. We take two approaches for this. First, for an easy visualization, we obtain a two-dimensional representation using principal component analysis (PCA), an unsupervised learning model for dimensionality reduction.⁴³ We plot the first two principal components in Figure 2c–e. We make two important

observations: (1) the points in the latent space are densely packed, centered around the origin, creating a continuous latent space. (2) There is a continuous change of both h_{form} and M_s values, indicating a property gradient in the latent space. In particular, Figure 2e suggests that the points in the latent space tend to form four different clusters depending on the two properties. These indicate that the latent space of MagGen is a property informed continuous space.

Although PCA is successful in providing key insights into the organization of the latent space, there must be significant information loss due to dimensionality reduction (from 256 to 2). Hence, in the second approach, we analyze the structure of the 256-dimensional latent space directly using the ideas of graph theory. We believe this is the first attempt of employing graph theoretic tools for efficient materials design using any GM.

A simple graph is a pair $G = (V, E)$ comprising the sets V and E . V is a set of vertices or nodes. $E \subset \{\{x, y\} | x, y \in V \text{ and } x \neq y\}$ is a set of edges. In our case, V is the set of all μ 's to which the training set materials get mapped. Two nodes are connected by an edge if their similarity measure, as indicated by the Tanimoto coefficient,^{44,45} crosses a threshold. The Tanimoto coefficient between two nodes v_1 and v_2 , with latent space vectors \vec{z}_1 and \vec{z}_2 , is defined as in eq 3.

$$S_{v_1, v_2} = \frac{\vec{z}_1 \cdot \vec{z}_2}{z_1^2 + z_2^2 - \vec{z}_1 \cdot \vec{z}_2} \quad (3)$$

Because of property embedding, we expect to find more edges between materials in the same class compared with materials across classes. However, since the model training is a statistical exercise, it is possible to find materials with dissimilar properties having high similarity. An edge connecting them would be incorrect from a materials perspective. We consider an edge to be correct if it connects two nodes representing materials in the same class; otherwise, it is incorrect. We analyze the structure of the resulting graph for a range of similarity thresholds between 0.3 and 0.8 using two quantities: (1) edge density, the number of edges per node, and (2) the fraction of correct edges, the ratio of the number of correct edges to the total number of edges. The plots are given in panels a and b of Figure 3, respectively. Edge density is higher at lower threshold values, as expected. The graph becomes increasingly sparse with higher threshold values. Figure 3b makes it evident that even for small values of the threshold, the fraction of correct edges is quite high. At a threshold of 0.5, this value is 90%. If we consider only the class of stable magnetic materials, the fraction of correct edges is >80%. This clearly demonstrates that the latent space of MagGen is organized according to the property values.

With the above insight, in order to generate new stable ferromagnetic materials with $h_{\text{form}} \leq 0$ eV/atom and $M_s \geq 0.5$ T, we sample points from the region of the latent space where the materials satisfying the design targets lie. By applying the decoder, we obtain the corresponding IRCRs, which are then converted to materials. The postprocessing steps for converting IRCR to a material is discussed in Section S4 in the SI. The two TLBs predict the h_{form} and M_s values directly from each sampled point.

Several methods have been used in the literature to sample points from the latent space, such as local perturbation (Lp),²⁶ spherical linear interpolation (Slerp),^{17,26,46} and global perturbation (Gp).²⁶ In terms of validity of the generated

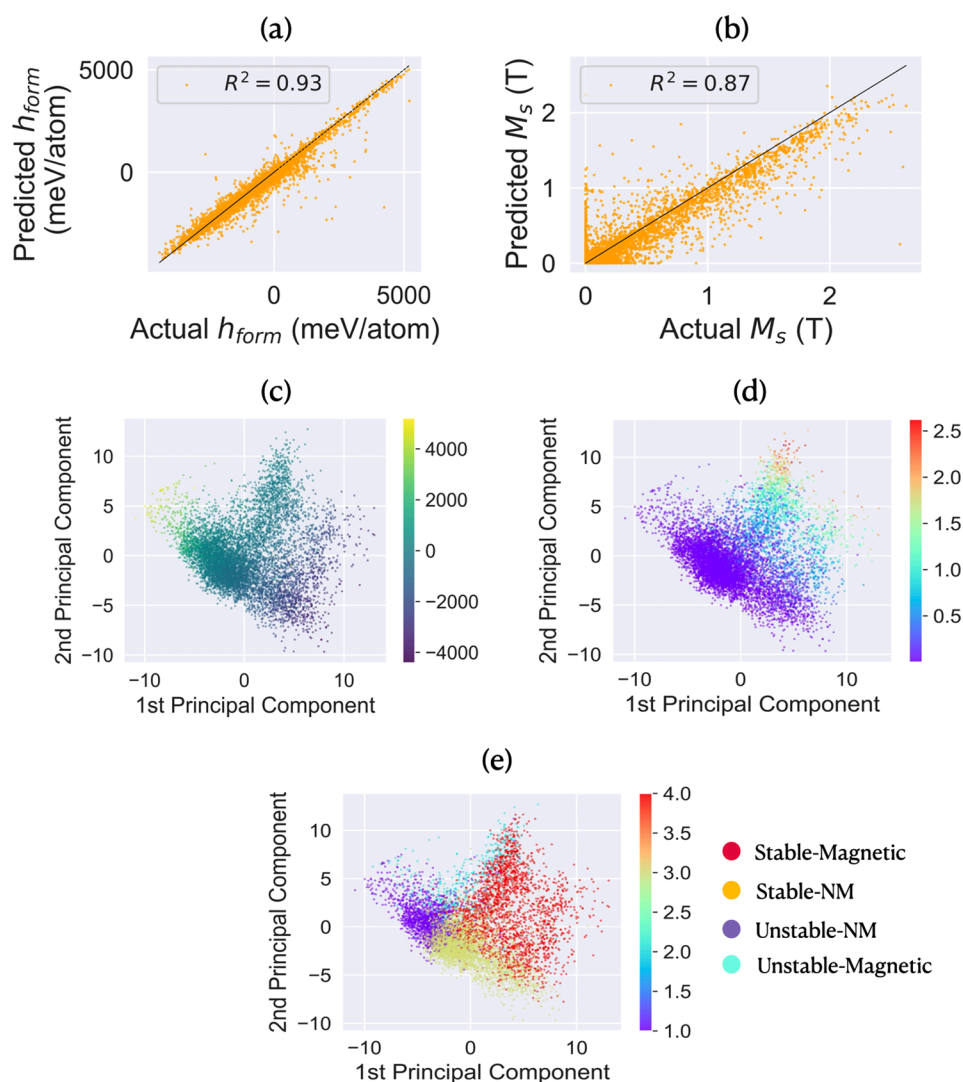


Figure 2. (a, b) Parity plots of h_{form} and M_s predictions from two target-learning branches. (c–e) Visualization of two-dimensional representation of the latent space, obtained by PCA. Color schemes are based on h_{form} values in meV/atom, M_s values in T, and four material classes, respectively.

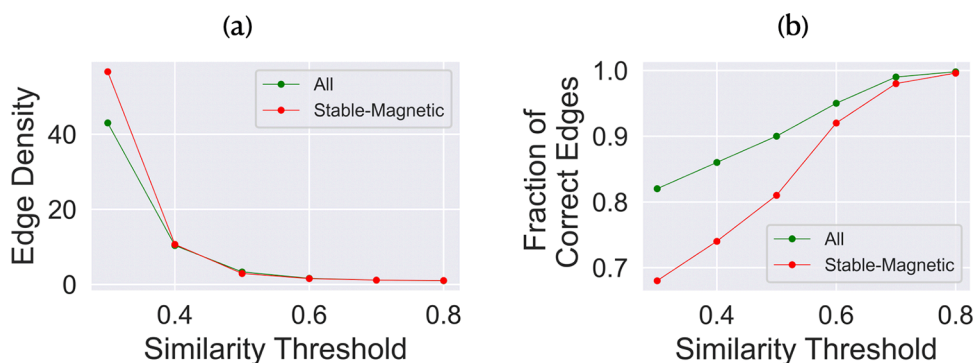


Figure 3. (a) Plot of edge density versus similarity threshold and (b) plot of fraction of correct edges versus similarity threshold for all materials and stable magnetic materials.

crystal structures in the setting of property driven materials design, Lp works the best.²⁶ In our work, we used Lp as the sampling strategy. A schematic diagram of new material generation using Lp is shown in Figure 1c. Here the latent vector (μ) of a parent material is locally perturbed by adding to it a scaled Gaussian noise, sampled from a standard normal distribution. The scale factor controls the trade-off between

exploitation and exploration. The generated materials are expected to display high reconstruction accuracy with low novelty for smaller values of the scale factor. For larger values, materials with high novelty are expected, but they could have low reconstruction accuracy. The effect of the scale factor on material generation is discussed in Section S5 in the SI. Here, we have used a scale factor of 1.0.

We also check the relevance of the graph analysis on the efficiency of material generation in terms of satisfying the design targets. First, we generated 15 351 materials using Lp around the points representing stable magnetic materials with $M_s \geq 0.5$ T in the training set without any regard to how these materials are connected to other materials in the graph. Out of these, 11 806 or 76.9% of materials are predicted to have $h_{\text{form}} \leq 0$ eV/atom and $M_s \geq 0.5$ T by the TLBs.

Next, we selected parent materials from the training set that not only satisfy the design targets but also have more than 80% correct edges at a similarity threshold of 0.4. We generated a total of 3183 materials using Lp around the points representing the selected parent materials. Out of these, 3064 materials are predicted to have $h_{\text{form}} \leq 0$ eV/atom and $M_s \geq 0.5$ T by the TLBs. Thus, an impressive 96.2% of the generated materials are predicted to satisfy the design targets.

While property embedding alone ensures that $\sim 77\%$ of the generated materials satisfy design targets, that number increases dramatically when the connectivity of the parent materials is also taken into account. As a comparison, Zhao et al.²⁴ generated 2 126 042 cubic materials using CubicGAN and screened them by a crystal graph convolutional neural network (CGCNN)¹¹ to predict h_{form} . In the absence of property embedding in the latent space, only 108 897, i.e., only 5%, of the generated materials were predicted to have $h_{\text{form}} \leq 0$ eV/atom.

Finally, we validate the generated materials via DFT. We randomly select 100 materials out of 14 870, predicted to both satisfy the design targets and have stoichiometry that is different from their parent materials. We choose this smaller set since performing complete lattice structure optimization of 14 870 materials is beyond our computational resources.

First, we optimized the lattice shape and size and the positions of the atoms of the selected materials. We then calculated h_{form} and M_s in the converged structures. Details are given in Section S6 in the SI.

To assess the performance of MagGen in new materials generation, we use the same three metrics that were used in ref 26: (1) validity rate, defined to be the fraction of materials for which structural relaxation is successful in DFT; (2) success rate, i.e., the fraction of materials satisfying the design targets as calculated via DFT; and (3) improvement over random success rate, where random success rate is the probability that a material randomly chosen from the training data set satisfies the design target properties. This is approximated by the fraction of materials in the data set that satisfy the design targets. Improvement is given by (success rate – random success rate)/random success rate.

Out of 100 materials, structural relaxation converged for 87 materials, giving a validity rate of 87%. Among these 87 materials, 60 materials have $h_{\text{form}} \leq 0$ eV/atom, and 68 have $M_s \geq 0.5$ T. Overall, 46 materials satisfy both design targets. Thus, the success rate for generating materials with $h_{\text{form}} \leq 0$ eV/atom is 60% and that for generating magnetic materials with $M_s \geq 0.5$ T is 68%. Overall, the success rate of generating materials that satisfy both design targets is 46%. In the entire data set, only 16.54% of materials have $h_{\text{form}} \leq 0$ eV/atom and $M_s \geq 0.5$ T, so the improvement over random success rate is 178%.

Ren et al.²⁶ obtained validity rates of 42.9–96.4% and success rates of 7.1–38.9% for different material design strategies. A success rate of 38.9% was obtained for a single-target design task ($h_{\text{form}} = -0.5$ eV/atom), and a success rate of

7.1% was obtained for multitarget design tasks. The success rate was found to decrease significantly as the number of target properties increased.

The validity rate of CubicGAN²⁴ was only 33.83%. For PGCGM,⁴⁷ the validity and success rates were 93.45% and 39.6%, respectively, for generating materials with the sole target of $h_{\text{form}} \leq 0$ eV/atom.

Clearly, MagGen is a significant improvement over the previously reported GMs in a number of points. First, it is trained only on real space features (IRCR), which can be directly taken from material cif files and do not require any computations for their construction. Second, by embedding both target properties simultaneously in the latent space, the success rate improves significantly in a multitarget task, making the materials generation process practically more appealing. Third, a similarity analysis based on graph structure in the latent space makes the material generation process more efficient. Out of the 46 materials, 2 of the most interesting materials (reasons discussed below) are listed in Table 2, along with their DFT-calculated properties. A complete list of these 46 materials is given in Tables S3 and S4.

Table 2. Composition, h_{form} , M_s , E_{hull} , and K_1 of the Two Candidate Stable Magnetic Materials from MagGen, Verified by DFT Calculations, That Are Promising for RE-Free Permanent Magnet Applications

composition	h_{form} (eV/atom)	M_s (T)	E_{hull} (eV/atom)	K_1 (MJ/m ³)
Fe ₃ O ₂ F ₂	−1.881	1.09	0.076	2.15
MnNi ₂ O ₂	−1.034	1.23	0.248	1.99

We wish to add that while calculating the success rate, we also considered the materials that have M_s in the range of $0.43 \leq M_s \leq 0.5$ T for successful materials generation. The lower bound is obtained by subtracting the MAE of M_s prediction by TLB-2 from our threshold of 0.5 T. This is to explicitly ensure that materials that are within the range of uncertainty of ML prediction are also counted.

In order to have confidence in the ability of MagGen to explore the materials space, we studied how novel the generated materials are. We searched for the 46 new magnetic materials in the MP,⁴¹ OQMD,⁴⁸ and ICSD⁴⁹ databases. Only 18 entries had matching stoichiometry, but the space groups were different (Table S5).

We calculated E_{hull} for the 46 candidate materials for a more complete assessment of their thermodynamic stability and experimental synthesizability using the pymatgen⁵⁰ PhaseDiagram module, as well as formation energies of the materials with the same phase as in MP. We found 20 materials having $E_{\text{hull}} \leq 0.25$ eV/atom and six more that are within 0.3 eV/atom of the hull. Under appropriate conditions of synthesis, materials that are 300 meV/atom above the hull (e.g., 1T-MoS₂) are routinely synthesized. Therefore, these materials are likely to be experimentally synthesized.

In search of candidate materials for permanent magnets, we calculate K_1 for the materials that have $M_s \geq 1$ T and $E_{\text{hull}} \leq 0.25$ eV/atom (Table S3). Among 12 such materials, 10 materials turn out to have uniaxial anisotropy, and 2 materials, namely Fe₃O₂F₂ and MnNi₂O₂, have $K_1 > 1$ MJ/m³. Thus, although not explicitly embedded in the latent space, MagGen is able to generate permanent magnets with large uniaxial anisotropy that are also RE-free.

There are two minor, but common, shortcomings that MagGen suffers from. The discovered materials belong to the space groups ranging from 1 to 8, suggesting bias toward generating low-symmetry structures. In fact, other generative frameworks also suffer from a similar bias.^{17,26,51,52} Zhao et al.⁴⁷ used symmetry oriented constraints to generate high-symmetry structures. However, no TLBs were used to infuse property information into the latent space. The generation of high-symmetry structures in property driven materials design is an open question. Recently, diffusion-based models⁵³ have shown promise for generating high-symmetry crystal structures.

IRCR does not have rotational, translational, permutational, or supercell invariance, argued to be requirements for materials representations for ML.⁵⁴ Most common representations used for generative models do not satisfy full invariance. The issue can be addressed by data augmentation.²² Recently developed diffusion-based models^{51–53,55} ensure full invariance via equivariant graph neural networks. Although it is not clearly understood, experience shows that even without the full invariance properties, representations do lead to successful materials generation, such as in voxel image representation,¹⁷ FTCP,²⁶ and IRCR in this work.

On a technical note, we wish to highlight that the performance of MagGen improves rapidly with increasing training data size (Figure S2). Hence, the success rate can be increased significantly once it is trained on bigger training data.

In summary, we developed a generative deep learning model to inversely design stable magnetic materials. DFT calculations were performed to validate the ML predictions. We demonstrated that real space features (IRCR) are sufficient for the inverse design purpose, eliminating the need for complicated reciprocal space features. We created a material latent space that simultaneously encodes information about two properties. Graph theory was used to analyze the structure of the latent space. This helped us in an ultra-efficient sampling of the space, whereby we discovered 46 stable magnetic materials by performing DFT on only 100 materials, giving a very high success rate of 46%. A total of 20 materials were found to be within 0.25 eV/atom of their respective convex hulls, suggesting the possibility of experimental synthesis. Two of these materials were also found to have uniaxial AE in excess of 1 MJ/m³, making them attractive candidates for rare earth-free permanent magnets.

■ ASSOCIATED CONTENT

Data Availability Statement

The data associated with the present work is available at <https://doi.org/10.6084/m9.figshare.25287742.v1>.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpclett.4c00068>.

Details of the training data set, IRCR, VAE architecture and training, DFT calculations, and the generated materials (PDF)

Transparent Peer Review report available (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Prasenjit Sen — Harish-Chandra Research Institute, a CI of Homi Bhabha National Institute, Jhansi, Prayagraj 211019, India; Department of Physics, Indian Institute of Science

Education and Research, Tirupati, Tirupati 517507, India; orcid.org/0000-0002-6334-0091; Email: prasenup@gmail.com

Authors

Sourav Mal — Harish-Chandra Research Institute, a CI of Homi Bhabha National Institute, Jhansi, Prayagraj 211019, India

Gaurav Seal — Indian Institute of Science Education and Research Thiruvananthapuram, Thiruvananthapuram, Kerala 695551, India

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpclett.4c00068>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was funded by the DAE, Govt. of India through institutional funding to HRI. All of the computations were done on the cluster computing facility at HRI (<https://www.hri.res.in/cluster/>). S.M. acknowledges Jyotirmaya Shivottam and Subhankar Mishra for the many insightful discussions.

■ REFERENCES

- (1) Gómez-Bombarelli, R.; et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15*, 1120–1127.
- (2) Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Aspuru-Guzik, A. What is high-throughput virtual screening? A perspective from organic materials discovery. *Annu. Rev. Mater. Res.* **2015**, *45*, 195–216.
- (3) Shinde, A.; Suram, S. K.; Yan, Q.; Zhou, L.; Singh, A. K.; Yu, J.; Persson, K. A.; Neaton, J. B.; Gregoire, J. M. Discovery of manganese-based solar fuel photoanodes via integration of electronic structure calculations, pourbaix stability modeling, and high-throughput experiments. *ACS Energy Lett.* **2017**, *2*, 2307–2312.
- (4) Wu, Y.; Lazic, P.; Hautier, G.; Persson, K.; Ceder, G. First principles high throughput screening of oxynitrides for water-splitting photocatalysts. *Energy Environ. Sci.* **2013**, *6*, 157–168.
- (5) Kirklin, S.; Chan, M. K. Y.; Trahey, L.; Thackeray, M. M.; Wolverton, C. High-throughput screening of high-capacity electrodes for hybrid Li-ion–Li–O₂ cells. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22073–22082.
- (6) Kirklin, S.; Meredig, B.; Wolverton, C. High-throughput computational screening of new Li-ion battery anode materials. *Adv. Energy Mater.* **2013**, *3*, 252–262.
- (7) Gorai, P.; Toberer, E. S.; Stevanović, V. Computational identification of promising thermoelectric materials among known quasi-2D binary compounds. *J. Mater. Chem. A* **2016**, *4*, 11110–11116.
- (8) Li, X.; Zhang, Z.; Yao, Y.; Zhang, H. High throughput screening for two-dimensional topological insulators. *2D Materials* **2018**, *5*, 045023.
- (9) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohler, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G.; Marzari, N. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **2018**, *13*, 246–252.
- (10) Yeo, B. C.; Kim, D.; Kim, H.; Han, S. S. High-throughput screening to investigate the relationship between the selectivity and working capacity of porous materials for propylene/propane adsorptive separation. *J. Phys. Chem. C* **2016**, *120*, 24224–24230.
- (11) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.

- (12) Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **2019**, *31*, 3564–3572.
- (13) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.
- (14) Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol. Pharmaceutics* **2017**, *14*, 3098–3104.
- (15) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **2018**, *4*, 120–131.
- (16) Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced adversarial neural computer for de novo molecular design. *J. Chem. Inf. Model.* **2018**, *58*, 1194–1204.
- (17) Noh, J.; Kim, J.; Stein, H. S.; Sanchez-Lengeling, B.; Gregoire, J. M.; Aspuru-Guzik, A.; Jung, Y. Inverse design of solid-state materials via a continuous representation. *Matter* **2019**, *1*, 1370–1384.
- (18) Long, T.; Fortunato, N. M.; Opahle, I.; Zhang, Y.; Samathrakris, I.; Shen, C.; Gutfleisch, O.; Zhang, H. Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. *npj Comput. Mater.* **2021**, *7*, 66.
- (19) Kim, B.; Lee, S.; Kim, J. Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **2020**, *6*, eaax9324.
- (20) Court, C. J.; Yildirim, B.; Jain, A.; Cole, J. M. 3-D inorganic crystal structure generation and property prediction via representation learning. *J. Chem. Inf. Model.* **2020**, *60*, 4518–4535.
- (21) Court, C. J.; Jain, A.; Cole, J. M. Inverse design of materials that exhibit the magnetocaloric effect by text-mining of the scientific literature and generative deep learning. *Chem. Mater.* **2021**, *33*, 7217–7231.
- (22) Kim, S.; Noh, J.; Gu, G. H.; Aspuru-Guzik, A.; Jung, Y. Generative adversarial networks for crystal structure prediction. *ACS Cent. Sci.* **2020**, *6*, 1412–1420.
- (23) Noura, A.; Sokolovska, N.; Crivello, J.-C. CrystalGAN: learning to discover crystallographic structures with generative adversarial networks. *arXiv* **2019**, 1.
- (24) Zhao, Y.; Al-Fahdi, M.; Hu, M.; Siriwardane, E. M. D.; Song, Y.; Nasiri, A.; Hu, J. High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Adv. Sci.* **2021**, *8*, 2100566.
- (25) Choubisa, H.; Askerka, M.; Ryczko, K.; Voznyy, O.; Mills, K.; Tamblyn, I.; Sargent, E. H. Crystal site feature embedding enables exploration of large chemical spaces. *Matter* **2020**, *3*, 433–448.
- (26) Ren, Z.; Tian, S. I. P.; Noh, J.; Oviedo, F.; Xing, G.; Li, J.; Liang, Q.; Zhu, R.; Aberle, A. G.; Sun, S.; et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter* **2022**, *5*, 314–335.
- (27) Coey, J. Permanent magnets: plugging the gap. *Scr. Mater.* **2012**, *67*, 524–529.
- (28) Nieves, P.; et al. Database of novel magnetic materials for high-performance permanent magnet development. *Comput. Mater. Sci.* **2019**, *168*, 188–202.
- (29) Mal, S.; Sen, P. Leveraging available data for efficient exploration of materials space using Machine Learning: a case study for identifying rare earth-free permanent magnets. *J. Magn. Magn. Mater.* **2024**, *589*, 171590.
- (30) Dutta, A.; Sen, P. Machine learning assisted hierarchical filtering: a strategy for designing magnets with large moment and anisotropy energy. *J. Mater. Chem. C* **2022**, *10*, 3404–3417.
- (31) Liao, T.; Xia, W.; Sakurai, M.; Wang, R.; Zhang, C.; Sun, H.; Ho, K.-M.; Wang, C.-Z.; Chelikowsky, J. R. Magnetic iron-cobalt silicides discovered using machine-learning. *Phys. Rev. Mater.* **2023**, *7*, 034410.
- (32) Halder, A.; Rom, S.; Ghosh, A.; Saha-Dasgupta, T. Prediction of the properties of the rare-earth magnets $\text{Ce}_2\text{Fe}_{17-x}\text{Co}_x\text{CN}$: a combined machine-learning and ab initio study. *Phys. Rev. Applied* **2020**, *14*, 034024.
- (33) Sakurai, M.; et al. Discovering rare-earth-free magnetic materials through the development of a database. *Phys. Rev. Mater.* **2020**, *4*, 114408.
- (34) Vishina, A.; Vekilova, O. Y.; Björkman, T.; Bergman, A.; Herper, H. C.; Eriksson, O. High-throughput and data-mining approach to predict new rare-earth free permanent magnets. *Phys. Rev. B* **2020**, *101*, 094407.
- (35) Vishina, A.; Eriksson, O.; Herper, H. C. Stable and metastable rare-earth-free permanent magnets from a database of predicted crystal structures. *Acta Mater.* **2023**, *261*, 119348.
- (36) Zhao, X.; Wang, C.-Z.; Yao, Y.; Ho, K.-M. Large magnetic anisotropy predicted for rare-earth-free $\text{Fe}_{16-x}\text{Co}_x\text{N}_2$ alloys. *Phys. Rev. B* **2016**, *94*, 224424.
- (37) Zhou, H.; Yan, S.; Wu, L.; Wan, X.; Wang, D. High-throughput search for potential permanent magnet materials. *Phys. Rev. Mater.* **2023**, *7*, 044405.
- (38) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2022**, 1.
- (39) Kingma, D. P.; Welling, M. An introduction to variational autoencoders. *FNT in Machine Learning* **2019**, *12*, 307–392.
- (40) Elrashidy, A.; Della-Giustina, J.; Yan, J.-A. Accelerated data-driven discovery and screening of two-dimensional magnets using graph neural networks. *arXiv* **2024**, 1.
- (41) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: the materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
- (42) Rhone, T. D.; Chen, W.; Desai, S.; Torrisi, S. B.; Larson, D. T.; Yacoby, A.; Kaxiras, E. Data-driven studies of magnetic two-dimensional materials. *Sci. Rep.* **2020**, *10*, 15795.
- (43) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. In *An introduction to statistical learning: with applications in R*; Springer, 2013.
- (44) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20.
- (45) Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* **2015**, *27*, 735–743.
- (46) White, T. Sampling Generative Networks: Notes on a Few Effective Techniques. *arXiv* **2016**, 1.
- (47) Zhao, Y.; Siriwardane, E. M. D.; Wu, Z.; Fu, N.; Al-Fahdi, M.; Hu, M.; Hu, J. Physics guided deep learning for generative design of crystal materials with symmetry constraints. *npj Comput. Mater.* **2023**, *9*, 38.
- (48) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (49) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **2019**, *52*, 918–925.
- (50) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (51) Xie, T.; Fu, X.; Ganea, O.; Barzilay, R.; Jaakkola, T. S. Crystal diffusion variational autoencoder for periodic material generation. *arXiv* **2021**, 1.
- (52) Wines, D.; Xie, T.; Choudhary, K. Inverse design of next-generation superconductors using data-driven deep generative models. *J. Phys. Chem. Lett.* **2023**, *14*, 6630–6638.

(53) Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Shysheya, S.; Crabbé, J.; Sun, L.; Smith, J. MatterGen: a generative model for inorganic materials design. *arXiv* **2023**, 1.

(54) Nelson, J.; Sanvito, S. Predicting the Curie temperature of ferromagnets using machine learning. *Phys. Rev. Mater.* **2019**, 3, 104405.

(55) Lyngby, P.; Thygesen, K. S. Data-driven discovery of 2D materials by deep generative models. *npj Comput. Mater.* **2022**, 8, 232.