



Towards end-to-end structure determination from x-ray diffraction data using deep learning



Gabe Guo¹✉, Judah Goldfeder¹, Ling Lan², Aniv Ray¹, Albert Hanming Yang¹, Boyuan Chen¹,
Simon J. L. Billinge¹✉ & Hod Lipson¹✉

Powder crystallography is the experimental science of determining the structure of molecules provided in crystalline-powder form, by analyzing their x-ray diffraction (XRD) patterns. Since many materials are readily available as crystalline powder, powder crystallography is of growing usefulness to many fields. However, powder crystallography does not have an analytically known solution, and therefore the structural inference typically involves a laborious process of iterative design, structural refinement, and domain knowledge of skilled experts. A key obstacle to fully automating the inference process computationally has been formulating the problem in an end-to-end quantitative form that is suitable for machine learning, while capturing the ambiguities around molecule orientation, symmetries, and reconstruction resolution. Here we present an ML approach for structure determination from powder diffraction data. It works by estimating the electron density in a unit cell using a variational coordinate-based deep neural network. We demonstrate the approach on computed powder x-ray diffraction (PXRD), along with partial chemical composition information, as input. When evaluated on theoretically simulated data for the cubic and trigonal crystal systems, the system achieves up to 93.4% average similarity (as measured by structural similarity index) with the ground truth on unseen materials, both with known and partially-known chemical composition information, showing great promise for successful structure solution even from degraded and incomplete input data. The approach does not presuppose a crystalline structure and the approach are readily extended to other situations such as nanomaterials and textured samples, paving the way to reconstruction of yet unresolved nanostructures.

Crystallography is the experimental science of determining the structure of crystals by analyzing x-ray, neutron or electron diffraction patterns^{1–3}. Powder crystallography is a sub-branch of crystallography that solves this problem when the measured sample consists of a large number of small, randomly oriented grains of the material^{4–7}. This problem is mathematically harder because of the loss of orientational information which must be recovered through inference during the structure reconstruction. It is useful when single crystals are difficult to obtain experimentally. However, it also is a good starting point for developing methods to determine the structure of nanomaterials and molecules in solution⁸, problems that currently have no robust solution.

The field of structure determination from powder diffraction⁹ has grown by adapting conventional crystallographic methods to the powder case. As with all crystallographic methods, these use inference and an iterative design approach to obtain structure candidates. The approach is a human-intensive activity requiring hands-on guidance by skilled experts. It involves first identifying the crystallographic coordinate system, a process called indexing, followed by finding the fractional coordinates of atoms in the unit cell from Bragg peak intensities^{1,9}. For PXRD data, the process sometimes works and sometimes does not, depending on the quality of the data and the complexity of the structure. It is not a straightforward process and requires considerable expertise.

¹Department of Computer Science, Columbia University, New York, NY, 10027, USA. ²Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, 10027, USA. ³Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, 27708, USA. ⁴Department of Mechanical Engineering, Columbia University, New York, NY, 10027, USA. ✉e-mail: gzg2104@columbia.edu; sb2896@columbia.edu; hl2891@columbia.edu

Recent work suggests that deep learning methods hold great potential to simplify the solution of complex inference problems with a straightforward end-to-end process. For instance, the protein-folding problem has recently been “solved” by end-to-end deep learning approaches like AlphaFold^{10,11} and RoseTTAFold¹². This is highly relevant, because protein folding is a sister problem to powder crystallography—both problems involve recovering the enigmatic shape of complex molecules from sparse and low-dimensional (i.e., 1-dimensional) inputs (amino acid sequences for the case of proteins and PXRD patterns for the powder crystallography case)¹³. Other examples of problems that have yielded to end-to-end learning are image classification¹⁴, autonomous vehicle driving¹⁵, and speech recognition¹⁶.

Machine and deep learning methods have been proposed to accelerate various stages of the powder crystallographic process. However, most of these works are conducted in a classification or feature regression paradigm: given an observation such as the XRD pattern, predict a property of the structure, such as space group symmetry, phase, unit cell parameters, or magnetism^{17–30}. There are some works that generate crystal structures, but their methodologies are not readily applicable to our problem because they (1) largely focus on unconditional (with respect to XRD pattern) generation cases in which there is no ground truth structure to reconstruct^{31–33}; (2) solve the easier single-crystal diffraction problem^{34,35}; (3) were designed only for specific classes of materials, such as proteins³⁶ and monometallic nanoparticles³⁷. Furthermore, the source code for many works in the deep learning for crystallography paradigm is not open-sourced, limiting their reproducibility^{23,32,34}.

Here, we propose an approach towards an end-to-end deep neural network that is able to determine a transformed three-dimensional electron density field directly from a 1-dimensional diffraction pattern. The actual electron density distribution may then be recovered with the inverse transform as we describe below.

The model we call *CrystalNet* is a variational³⁸ query-based multi-branch deep neural network (DNN) architecture (also known as a conditional implicit neural representation^{39–43}) that takes powder x-ray diffraction patterns and chemical composition information as input, and outputs a continuous function that is related to the 3D electron density distribution. We call this function the Cartesian mapped electron density (CMED) because we map the electron density from the crystallographic coordinate system of the structure to a Cartesian coordinate system. This distorts the resulting electron density but places it on a universal basis that allows the model to be seamlessly trained on structures from different crystal systems and with different unit cell parameters. The advantage of this representation for material structure is that it frees us from traditionally predefined properties such as the number of atoms and the crystallographic coordinate system. The actual electron density distribution may be recovered from the CMED through the inverse mapping, and if required, the discrete molecular structure can be straightforwardly decoded from this electron distribution if needed⁴⁴. After training, given a previously unseen diffraction pattern (and corresponding chemical composition information), *CrystalNet* can be queried to produce a 3D CMED map at any desired resolution. Due to our variational approach^{38,45}, *CrystalNet* can also be queried multiple times to produce different predictions, should the first guess be sub-satisfactory. The design, training and testing protocols are described in the Methods section.

The performance of the model are described here. We report preliminary results from the cubic and trigonal crystal systems using theoretically simulated data from the Materials Project⁴⁶. *CrystalNet* was able to reconstruct atomic structures from the cubic system almost perfectly. For the trigonal system, *CrystalNet* achieves success in most cases, with the infrequent failure modes providing insights for future work. We chose the cubic and trigonal systems for the initial tests as representative systems that are close to, and far from, respectively, the Cartesian coordinate system. They both have the property $a = b = c$, but in the trigonal case one of the lattice angles is 120° . As such these systems might be representative of best-case and not-as-good scenarios. Although other crystal systems were not explored fully in this study, the results on these two crystal systems give us

hope that our approach can be highly effective for the remaining five systems. We note that the model does not make use of any symmetry or chemical property information beyond composition and yet still shows success. This means that such information may be added as priors in future iterations when there is even greater information loss in the input signal, for example, due to very low symmetry structures or broad diffraction signals characteristic of nanomaterials.

We also conduct ablation studies by systematically reducing the input chemical composition information to gain insight into which information is most important for AI-enabled powder crystallography going forward. We find that while this information helps our model, for these high symmetry structures, crystal reconstruction is generally successful with only the XRD data and no compositional information at all.

Results

We evaluate *CrystalNet* by feeding in the XRD pattern, chemical composition, and queried coordinates as input. *CrystalNet* then processes this information with multiple branches and fuses it into one shared representation. Finally, via the charge density regressor, it outputs the predicted charge densities at each of the queried coordinates. See Fig. 1 for a schematic overview of how *CrystalNet* works.

Reconstruction

Table 1 shows reconstruction success metrics (SSIM, PSNR) on the cubic and trigonal crystal systems from powder XRD and chemical formula information. SSIM stands for structural similarity index, which measures the patchwise correspondence between two signals on a scale of 0 (worst) to 1 (best)⁴⁷. PSNR stands for peak signal-to-noise ratio, which measures the magnitude of the predicted charge density signal relative to the size of the errors in the prediction, where higher values are better, and ∞ indicates perfect reconstruction⁴⁸: typically, values of PSNR above 30 are considered high-fidelity reconstructions⁴⁹. More details are available in Methods.

To demonstrate the functionality of our methodology, Fig. 2 shows sampled reconstructions of two testing crystals viewed from various angles, given only chemical composition and powder XRD as input. Inspired by variational approaches, we achieve multiple reconstructions by sampling from the conditional latent distribution³⁸. We see that this stochasticity in output can be helpful if the initial guess is incorrect; in principle, we can resample to obtain a more reasonable prediction that matches the given XRD, as measured by the analytically solved forward process. Even for failure cases like Ge_7Ir_3 , we still see that sampling multiple times allows us to get a prediction that is closer to the ground truth. On average, over five latent space samples for the same given crystal (i.e., XRD and formula input), the standard deviation of SSIM is 0.017 in the cubic system and 0.018 in the trigonal system, while the standard deviation of PSNR is 2.78 in the cubic system and 0.68 in the trigonal system.

See Fig. 3a for success cases of cubic reconstruction, and Fig. 3b for failure cases of cubic reconstruction. Overall, reconstruction is very successful over a diverse range of crystal structures, judging from both visual and quantitative metrics.

Quantitatively (Table 1), we achieve great success, as evidenced by the 0.934 mean SSIM on the testing set. Indeed, value 1 (perfect reconstruction) is actually within one standard deviation (0.149) of this mean testing SSIM, indicating that many crystals had near-perfect reconstructions. The PSNR is also very high, above the typical success threshold⁴⁹ of 30, even if we go one standard deviation (12.7) below the mean PSNR (43.0) on the testing set.

From a qualitative perspective, we also see many good results (Fig. 3a). Encouragingly, we see that our method succeeds for structures (such as $\text{V}_3(\text{Co}_{10}\text{B}_3)_2$) with a high number of atoms in the unit cell, despite not knowing how many atoms are contained a priori. Its success also seems to be consistent across a variety of chemical compositions, e.g., it succeeds on both $\text{Gd}_2\text{Hf}_2\text{O}_7$ and ThCd_1 , which share no common elements. We also observe, as expected, that crystals containing similar elements—such as $\text{Zr}_3\text{Sb}_4\text{Pt}_3$ and $\text{Ce}_3\text{Sb}_4\text{Pt}_3$ —have similar structures, albeit with different average charge densities. Even the cubic failure modes (Fig. 3b) still give

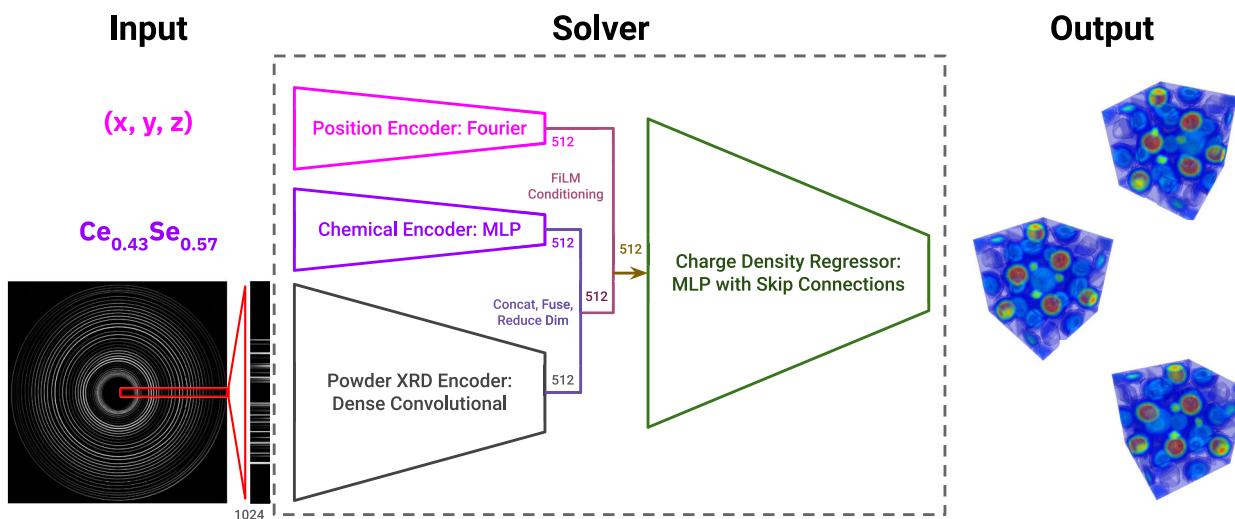


Fig. 1 | CrystalNet System Overview. As input, *CrystalNet* takes in 1D powder x-ray diffraction (PXRD) patterns that may be obtained by azimuthally integrating a 2D diffraction pattern such as the black square. It also takes in chemical composition ratios, and the atomic coordinates of known structures. In the Solver it

processes each input item with a specialized branch (pink, purple and black). It then fuses them into one unified latent vector of length 512, which is passed through the charge density regressor to produce a voxelized 3D charge density map at arbitrary resolution. See “Methods” for more details.

Table 1 | Reconstruction Performance from combined powder XRD and chemical formula

System	# Test	SSIM ($\mu \pm \sigma$)	PSNR ($\mu \pm \sigma$)
Cubic	500	0.934 ± 0.149	43.0 ± 12.7
Trigonal	237	0.741 ± 0.215	27.8 ± 8.1

PSNR is unbounded, and higher values are better.

SSIM ranges between [0, 1], where higher is better.

good guesses for rough structural outlines, even if the details are slightly incorrect. For instance, $\text{Cs}_3\text{H}_{12}\text{N}_4\text{F}_3$ has the predicted general structure close to the ground truth, but the charge density peaks are not as sharp, and the atomic boundaries are slightly blurred. Another example is $\text{Cr}_4\text{GaCuSe}_8$, which actually has a predicted structure reasonably close to the ground truth, except that the predicted structure is oriented upside-down and has some extraneous medium-charge locations. Indeed, the upside-down prediction is actually not that significant of an error, since material identity is invariant to rotation.

See Fig. 4a for success cases of trigonal reconstruction, and Fig. 4b for failure cases of trigonal reconstruction. Quantitatively and qualitatively, reconstruction on this system is also successful, although not as successful as the cubic system.

Looking at Table 1, we see that both the SSIM and PSNR are lower than that achieved on the cubic system. This is expected, as the trigonal system is less symmetric. That being said, SSIM levels are still decent, with average value 0.741 out of 1. Average PSNR levels are only slightly below the threshold for high-fidelity reconstruction (27.8 vs. 30).⁴⁹

Moving to qualitative analysis, similar to the cubic system outcomes, we are able to solve crystals with a high number of atoms in the unit cell (e.g., $\text{CrP}_6(\text{WO}_8)_3$), and crystals from diverse chemical makeups (e.g., LaZnCuP_2 , $\text{Ba}(\text{B}_2\text{Pt}_3)_2$). Additionally, in the trigonal success cases (Fig. 4a), we see that the model is able to successfully solve crystal structures with considerably lower symmetry than the examples in the cubic system. For instance, $\text{Rb}_3\text{Na}(\text{RuO}_4)_2$ and $\text{Mn}_8\text{Nb}_3\text{Al}$ are considerably less symmetric than any of the examples displayed for the cubic system, yet our method was still able to achieve high-fidelity reconstructions of both.

Furthermore, due to the CMED representation placing all crystals in a unit cell with orthogonal inter-axial angles (as opposed to the non-

orthogonal inter-axial angles of the trigonal system), we observe slight atomic distortion in both the ground truth and predicted structures, e.g., $\text{Ba}(\text{B}_2\text{Pt}_3)_2$ has ellipsoid rather than spherical site shapes. This is expected, and more detail about the CMED representation is available in Methods.

We see that the failure cases (Fig. 4b) are a bit more apparent for the trigonal system than for the cubic system. Indeed, some of the predictions do not contain useful information, e.g., $\text{Si}_5\text{P}_6\text{O}_{25}$. Noticeably, the model appears to have difficulty predicting the high charge density regions, such as in $\text{Pr}_6\text{Mn}(\text{SiS}_7)_2$. That being said, some failures (such as NaBiF_6) still contain reasonable information about the structure, which can be used as a first step in an iterative structural refinement process. It is also notable that a lot of the failure cases exhibit difficulty with orientation. For instance, NaBiF_6 , Rb_2PtC_2 , and $\text{Rb}(\text{V}_3\text{S}_4)_2$ have reconstructions that would be considered more reasonable, were they rotated differently.

Data ablation

We conduct ablation studies on the chemical formula information, since in reality, this data is known to varying degrees during the crystallographic process. We try three ablations: (1) Eliminate elemental ratio information, with a 1 in the composition vector if the element is contained in the material, and 0 otherwise; (2) Randomly drop one element from the ratio-free composition information, i.e., flip 1 to 0 for a singular randomly selected element (at least one element must be known, so we do not drop elements if the material contains a singular element); (3) No elemental information at all, leaving only XRD. In all these experiments, full XRD information was retained in all these ablation studies. See Table 2 for the results of the ablation studies.

See Fig. 5a for visualizations. As expected, as we ablate information about the chemical composition, the quantitative reconstruction performance, as measured by SSIM and PSNR, declines on the cubic system (Table 2). That being said, the visual and quantitative results indicate that even with heavy degradation in the elemental composition information inputted, we still achieve very reasonable reconstructions. Indeed, there is virtually no difference between the *Baseline* (powder + full composition info) and the *No Ratio* versions of the model, as measured by SSIM and PSNR. And, even though there is a ten-point PSNR gap between *Baseline* and *No Formula*, *No Formula* still has a mean PSNR of 30.0, which is higher than that of any of the trigonal versions.

See Fig. 5b for visualizations. The trend of decreasing performance with decreasing degrees of chemical composition information still generally

holds for the trigonal system (Table 2). Yet, similar to the cubic model, the trigonal model still works even under this heavy degradation in information.

Surprisingly, different from the cubic system, removing the formula altogether from the trigonal reconstruction model's input actually performs slightly better (as measured by SSIM and PSNR) than randomly dropping

one element from the composition information. For instance, the *No Formula* reconstruction for ErNi_3 is more successful than the *Drop One* (Er) reconstruction in Fig. 5b. It is also interesting to note that even the *No Formula* version of the cubic model performs better than the *Baseline* (full information) version of the trigonal model: this indicates that (at least using our model design), the cubic system is easier to solve than the trigonal system.

Discussion

This is a successful attempt at large-scale reconstruction of crystals in the cubic and trigonal systems. This is significant because it can pave the way for fully automated solutions to crystal structures from powder XRD data, potentially speeding up materials discovery and analysis by orders of magnitude. Furthermore, even if the structure initially predicted by our method is not correct, it can still be used as a first guess in the iterative refinement process, or we can even re-sample from the latent space to generate another candidate (since we use a variational approach).

Of particular interest is our CMED representation (described further in Methods). By mapping all structures onto a universal coordinate system, we are able to train the same model architecture on structures from different crystal systems and unit cell parameters. This is advantageous (especially in comparison to approaches that predict coordinates of discrete atoms), because this representation does not require a priori knowledge of properties that are required by other methods, such as the number of atoms or lattice vectors. However, because it re-maps structures onto another coordinate system, the CMED inherently distorts atoms, in size and shape.

All the experiments conducted were on simulated powder x-ray diffraction patterns. Furthermore, many of the materials in the Materials Project are theoretical materials that have never been synthesized⁴⁶. This still provides us valid data pairs to train and evaluate our model, since generating XRD from crystal structure is an analytically solved problem¹. However, this also means that much of the data is free from defects we would find in experimental data, e.g., peak broadening, missing peaks⁴⁸. Thus, while we

Fig. 2 | Multi-view variational reconstructions. We display CMED reconstructions of previously unseen crystals at multiple viewing angles. Given chemical composition and powder XRD as input, we generate a distribution over latent codes, which we then sample from to obtain multiple crystal reconstructions. This allows us to obtain better guesses, if the first prediction is sub-optimal.

Fig. 3 | Cubic system reconstructions. **a** shows the CMED plots for the success cases. **b** shows the failure cases. Ground truth and *CrystalNet* prediction alternate left-to-right. Formulas are under ground truth images, and the corresponding powder XRD peak inputs are under predictions. Powder XRD peak inputs are visualized as relative intensity maps, with the diffraction angle (horizontal direction) increasing from $0^\circ \rightarrow 180^\circ$.

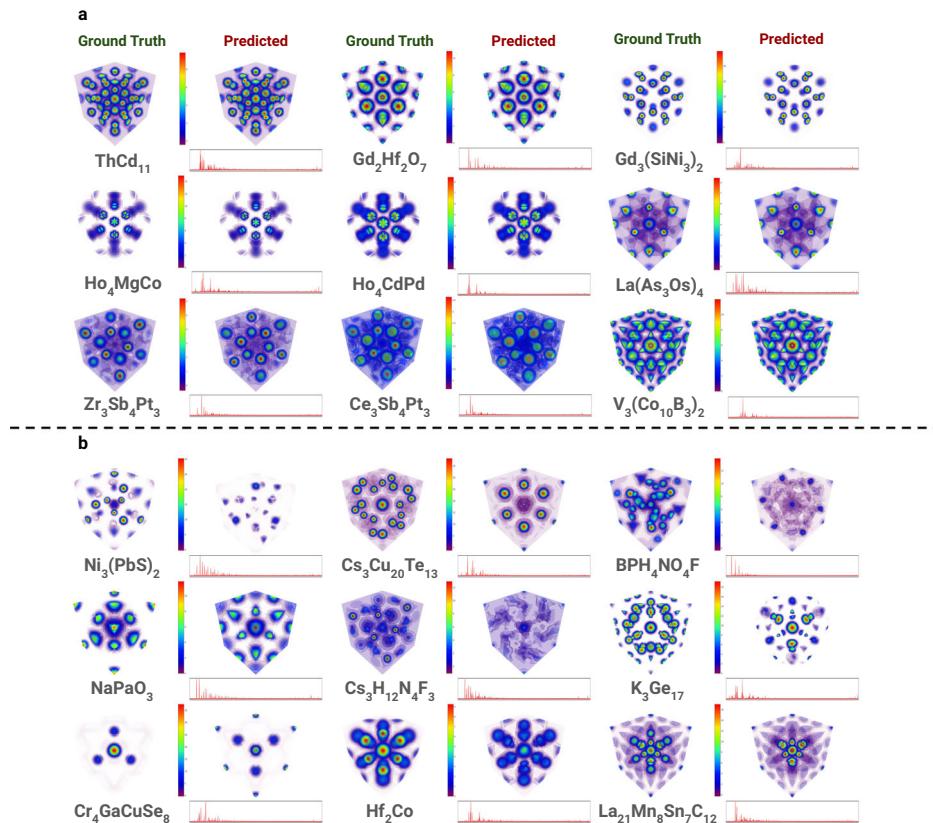


Fig. 4 | Trigonal system reconstructions. **a** shows the CMED plots for the success cases. **b** shows the failure cases. Ground truth and *CrystalNet* prediction alternate left-to-right. Formulas are under ground truth images, and the corresponding powder XRD peak inputs are under predictions. Powder XRD peak inputs are visualized as relative intensity maps, with the diffraction angle (horizontal direction) increasing from $0^\circ \rightarrow 180^\circ$.

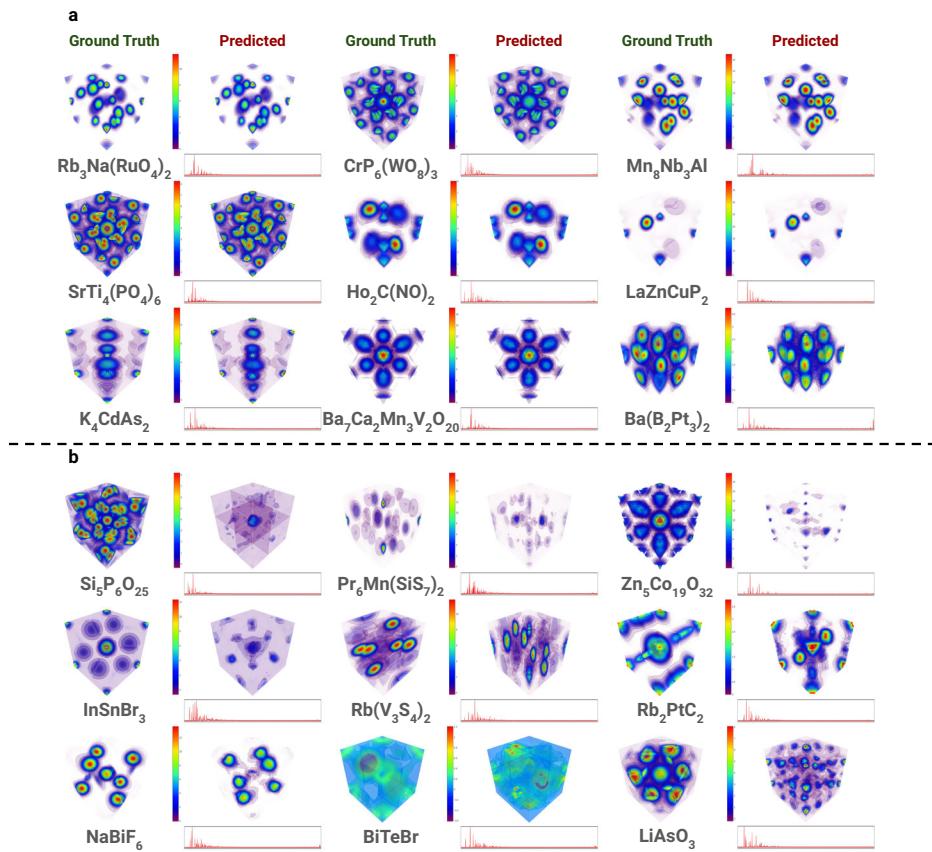


Table 2 | Ablation Performance at various levels of input information: (1) Baseline: Powder XRD + full chemical composition information; (2) No Ratios: Powder XRD + elements contained, without any information about their ratios; (3) Drop Element: Powder XRD + elements contained, with one element randomly dropped; (4) No Formula: Powder XRD information only

System	SSIM ($\mu \pm \sigma$)	PSNR ($\mu \pm \sigma$)
Cubic		
Baseline	0.914 ± 0.171	40.9 ± 12.8
No Ratios	0.916 ± 0.161	40.3 ± 12.5
Drop Element	0.887 ± 0.170	32.5 ± 9.0
No Formula	0.868 ± 0.174	30.0 ± 8.6
Trigonal		
Baseline	0.732 ± 0.207	26.9 ± 7.2
No Ratios	0.718 ± 0.209	26.8 ± 7.2
Drop Element	0.695 ± 0.203	25.4 ± 6.1
No Formula	0.703 ± 0.207	25.5 ± 6.1

have shown that deep learning methods, in principle, can work to solve the structure problem, there will still need to be future work to overcome this simulation-to-real gap.

Furthermore, we solved only the two most symmetrical crystal systems, out of seven total¹. Based on our preliminary explorations on the other five systems, we hope that this method, with appropriate tweaks, could be applicable to them. Indeed, due to our CMED representation, the data format should be exactly the same: empirical chemical formula and PXRD pattern as input, voxelized electron density grid as target. Yet, future exploration needs to be done to adapt our approach to these other systems,

such as addressing the unequal lattice vector lengths and different symmetry operations.

Additionally, solving crystal structures can be a one-to-many problem, in the cases of degraded XRD and/or chemical composition data. Although the variational approach allows us to have variation in the output via resampling from the latent space (see “Methods”), we seek more principled ways to model the uncertainty in our predictions.

Also, our representation of chemical composition information only tells the model which elements are contained, but it does not encode information about the chemical properties. In future works, we can perhaps incorporate some prior chemical knowledge, e.g., atomic mass, period, group.

Finally, as seen by some outputs that were reasonable but oriented incorrectly, future work should either propose a reliable method for enforcing canonical poses or design a model that can learn on and output multiple orientations of the same structure.

Methods

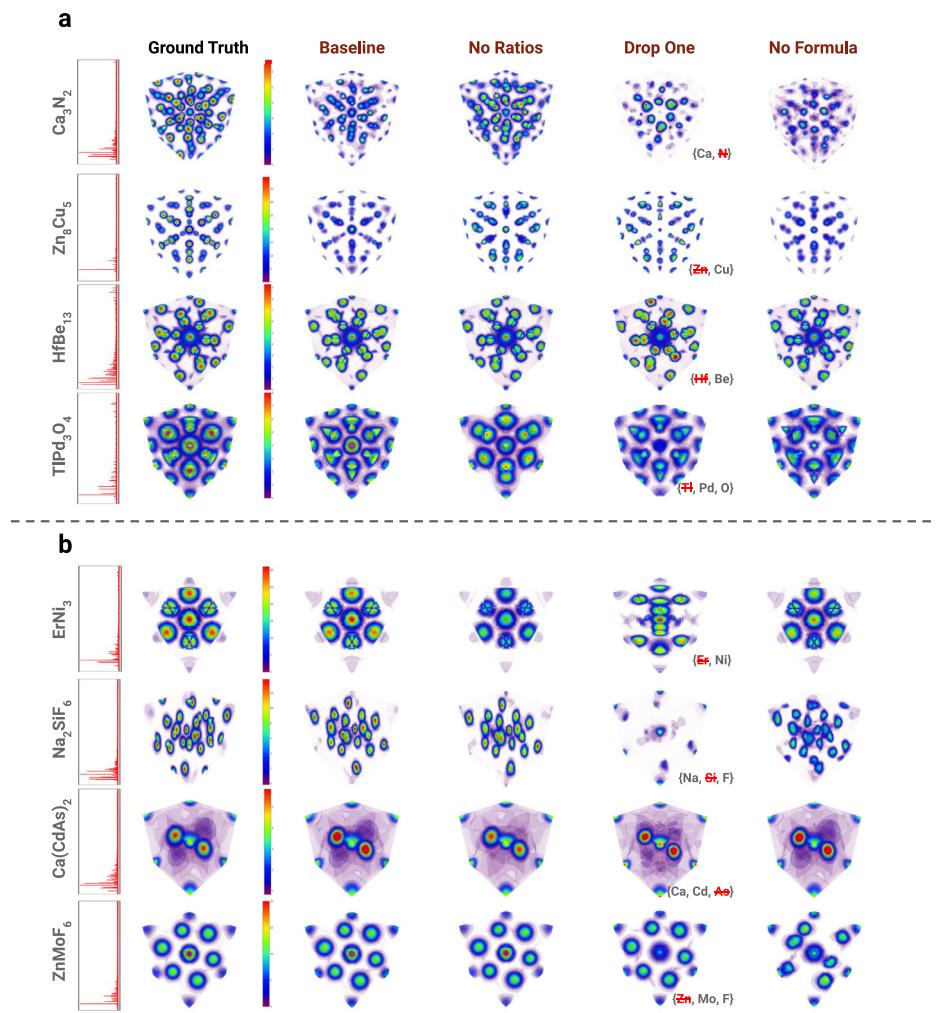
Dataset

We get our data from the Materials Project⁴⁶, which has publicly available standard data on over 150,000 inorganic compounds, largely for materials in the Inorganic Crystal Structure Database (ICSD)⁵⁰. Some of the material properties are experimentally observed, while others are calculated with Density Functional Theory (DFT)^{51,52}.

We ensure there is no train-test leakage in the dataset, as follows. Our criteria for whether two molecules are “duplicates” is that they have the same (1) chemical formula; and (2) spacegroup. We go through our datasets and find all the molecules that have the same formula-spacegroup combination. Out of the molecules that share the same formula-spacegroup combination, we remove all but one of them from our dataset.

We use data from the cubic and trigonal crystal systems, which constitute two out of the seven total crystal systems⁵³. We only experiment on

Fig. 5 | Ablation studies. **a** shows cubic system ablations. **b** shows trigonal system ablations. Left-to-right in each panel: ground truth, full information (XRD + formula) prediction, excluding elemental ratios (XRD + elements contained) prediction, randomly dropped element (XRD + all but one element contained) prediction, XRD-only prediction.



these two systems in this preliminary study because the intra-crystal axial lengths are equal (i.e., $a = b = c$), which eliminates the need to predict the axial lengths (whether implicitly as an intermediate calculation, or explicitly as the model's output), and allows us to focus on predicting charge densities. See Table 3 for the numbers of crystals used in our experiments.

We run separate experiments for the cubic and trigonal systems, i.e., we train and test one version of our model only on cubic crystals, and we train and test another version of our model only on trigonal crystals. In practice, to determine the structure of a material, we would run each version of the model (where each version is trained to solve one specific crystal system) on the XRD and partial chemistry information, then take the most plausible structure from the given outputs. This does not add significant burden to the end user of our method, since there are only seven total crystal systems, and inference time for our model is less than a minute per structure.

We use the theoretically calculated powder x-ray diffraction patterns from the Materials Project API. The diffraction angle ranges used were between 0° and 180° . More detail is available in the references^{54,55}.

The simulated patterns are generated using the MoK α wavelength of 0.711 Å. Depending on the atom types present in the compounds, the amplitude of the powder XRD patterns may vary drastically. This variation can be inherently problematic for most machine learning algorithms⁵⁶. To solve this issue, we normalize the peak intensities so that the highest peak intensity is set to 1. While this normalization process does reduce some of the information related to specific atom species, it retains the relative differences between them. Consequently, when the chemical formula is provided, or even if only partial information about the atom

species is available, we can still reconstruct the structure with the correct atom types.

We reiterate that the simulated patterns we use are of higher quality than those collected in experimental settings, due to the lack of noise, e.g., peak broadening, missing/extraneous peaks. Thus, we note that performance is expected to fall off for real data compared to the simulated diffraction patterns. Tests on real data will be the subject of a future study.

We also incorporate the chemical composition, that is, the molar ratios of the elements contained in the material. We include this because chemical composition is often known, at least to some degree. We also test the robustness of the model by ablating this information to various degrees in our experiments.

For the training, validation, and testing data, we use electron density maps from Materials Project DFT calculations⁵⁷⁻⁵⁹. These are in a crystallographic basis, which depends discontinuously on the crystal system and details of the unit cell size and shape as we move from one material to another. We resample the electron densities within the unit cell onto a grid that has 50 voxels along each axis, with the locations of the voxels expressed in fractional coordinates. We use PyRho (a library from the Materials Project)⁵⁹ to do this via Fourier interpolation. The charge densities are further normalized to be expressed in $e^- \cdot \text{Å}^3$. This will give different spatial resolutions for different structures, but has the advantage that it gives a representation that is a uniformly shaped array for all materials.

We call this quantity the Cartesian mapped electron density (CMED). The result of the normalization and resampling is a grid of $50 \times 50 \times 50$ voxels. For visualization we can project this onto a Cartesian coordinate

system with orthonormal basis vectors. The CMED is distorted from the real electron density by the procedure, but it allows us to visualize all structures, from all crystal systems and unit cells, on the same coordinate system. However, more importantly, it allows in principle a single ML model to be trained on structures from all the different space groups and crystal systems.

We stress the importance of CMED's uniformly shaped array for all materials. In previous attempts, we attempted to predict electron densities in raw Cartesian space (e.g., electron density queries were at exact Angstrom positions). The issue with that is that the output domain was not well-bounded, so we needed to train with large maximum (x, y, z) coordinates. While this was reasonable for crystals with very large unit cells, it did not work so well on crystals with small unit cells, as the training objective was too sparse. In contrast, the CMED representation maps every unit cell to a space

where the coordinates are well-bounded, which makes training much more tractable.

To get from the CMED predicted by our model to an undistorted electron density the inverse mapping must be carried out. If the unit cell of the unknown structure is indexed and the lattice parameters are known, this is straightforwardly done by plotting the voxels in the same order in the other basis.

In practice, we seek an end-to-end procedure that can discover the unit cell parameters as part of the automated process. This has not been done in the current paper, but we believe it will be straightforward. Indeed, there is already evidence that such information can be obtained straightforwardly by ML^{25,60,61}.

Neural network design

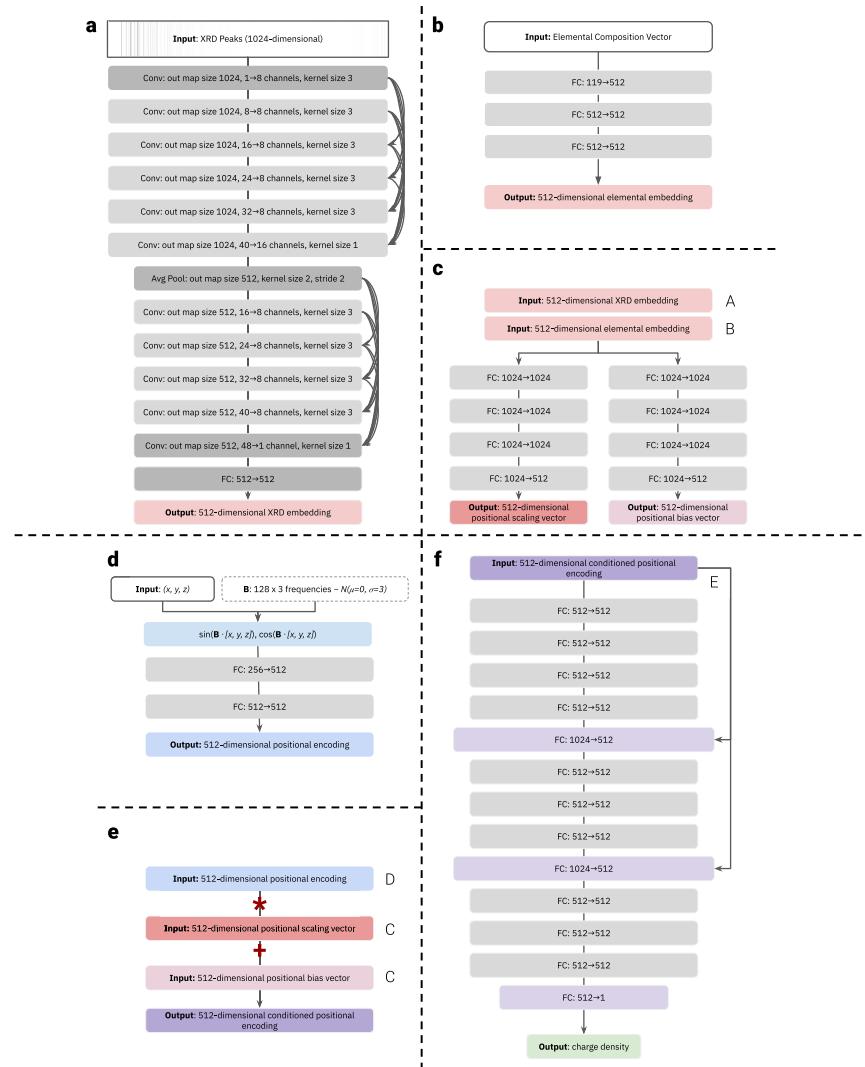
See Fig. 6 for the layer-by-layer neural network architecture. See Fig. 7 for a mid-level system diagram that shows how the components interact. The XRD, chemical composition, and spatial positions are inputted into the model and processed by separate branches. The XRD and chemical composition embeddings are fused with each other via concatenation. Then, they are fused with the spatial position embedding via FiLM⁶². That fused representation is then passed to the charge regressor, which predicts the charge density at the queried spatial positions. In total, our model has 14,775,187 parameters.

Table 3 | Number of crystals

Crystal System	Train	Val	Test
Trigonal	9177	241	237
Cubic	16378	250	500

All samples are obtained from the Materials Project⁴⁶. Training data consists of a mix of stable and unstable crystals, while validation and testing data is purely stable. We conduct a separate experiment on each crystal system.

Fig. 6 | CrystalNet architecture. **a** shows the architecture of the x-ray diffraction encoder, **b** shows the architecture of the elemental composition encoder, **c** shows the architecture of the feature fusion network, **d** shows the architecture of the positional encoder, **e** shows the architecture of the conditioning network, and **f** shows the architecture of the final charge density regressor.



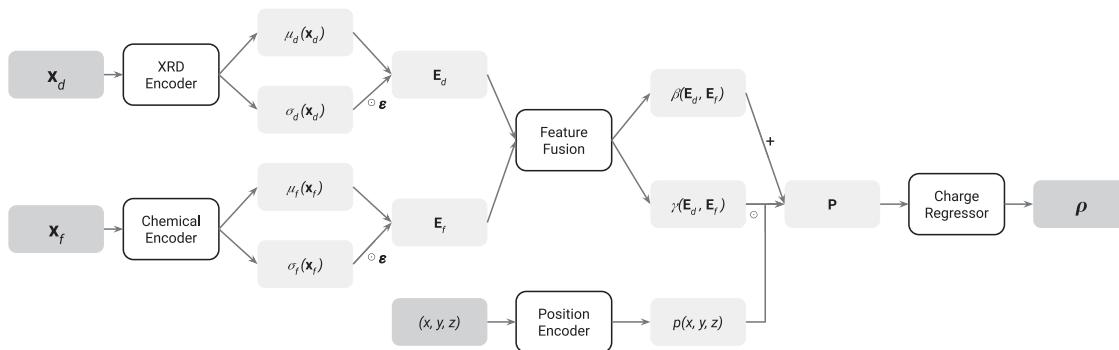


Fig. 7 | CrystalNet system diagram. XRD and chemical input are passed through encoders, which output latent feature distributions. Sampling from those latent distributions, we fuse the features with the positional (x, y, z) query, and pass the

fused information through the charge density regressor to get our final charge density prediction.

We adopt a variational approach^{38,45} for powder XRD and formula embedding prediction. Particularly, rather than deterministically predicting the embeddings, we predict the means and standard deviations of the embedding distributions, which are modeled as multivariate Gaussian distributions. Thus, we have

$$E(\mathbf{x}) \sim N(\mu(\mathbf{x}), \sigma^2(\mathbf{x})) \quad (1)$$

where E is a sample from the distribution of formula- or XRD-conditioned embeddings, \mathbf{x} is the corresponding formula or XRD input, μ is the neural network function that regresses the mean, and σ is the neural network function that regresses the standard deviation. We use the reparameterization

$$E(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon * \sigma(\mathbf{x}) \quad (2)$$

where ϵ is unit Gaussian noise, to make the process differentiable³⁸.

We justify this variational approach with the following reasons: (1) Crystallographic inference, i.e., predicting molecular structure given XRD and formula, can be a one-to-many problem, so a non-deterministic approach is appropriate for modeling these multiple outputs. (2) Crystallography is an iterative design process. The variational approach allows us to resample candidate structures, if the first prediction is not appropriate. (3) Variational approaches allow the model to learn a smoother latent space, which may generalize better to out-of-training-distribution inputs³⁸.

We note that although Fig. 6 depicts the powder XRD (Panel a) and formula (Panel b) encoders as deterministic networks, this is only for the sake of simplicity in the illustration. In reality, we have two versions of each network, one for regressing μ , and the other for regressing σ , which are then combined to produce the actual embedding, according to Eq. (2).

The powder XRD encoder is shown in Fig. 6, Panel a; in short, the goal of this branch is to extract relevant information from the sparse XRD pattern. The inputs are the extracted peaks \mathbf{x}_d from the x-ray diffraction patterns, which are normalized such that the highest peak is at intensity 1. They are represented as vectors with 1024 pixels of resolution, where the value at each pixel represents the intensity of the diffraction pattern at that location. The outputs are 512-dimensional embeddings $E_d = E_d(\mathbf{x}_d)$.

The architecture is an adaptation of the DenseNet architecture for vector (rather than image matrix) inputs, with the most important design characteristic being the densely connected concatenations between convolutional feature maps⁵³. The convolutional feature maps provide the important inductive bias of translational invariance, since (at least in early stages of processing) we wish to extract low-level features from all XRD peaks, regardless of where they are located, in essentially the same way. The dense connections promote integration of low-level and high-level features that may both be important to solving the task. Every convolutional layer (except the last one) is followed by LayerNorm⁶⁴ and ReLU; the final linear

layer is followed by BatchNorm⁵⁶. We re-emphasize that technically, we have two versions of the XRD encoder under our variational framework: one for regressing $\mu_d(\mathbf{x}_d)$, and one for regressing $\sigma_d(\mathbf{x}_d)$, to construct $E_d(\mathbf{x}_d)$ as defined in Eq. (2).

The formula encoder is shown in Fig. 6, Panel b; this branch is intended to extract relevant information about the chemistry that complements the information contained in the XRD peaks. The input is the empirical formula, represented as a 118-dimensional vector \mathbf{x}_f , where each index of the vector refers to the normalized amount (as defined by number of atoms) of the element with that atomic number that is contained in the formula. (For instance, if the formula was H2O, we would first normalize that to H0.66O0.33. The resultant vector would contain 0.66 at index 1, the atomic number of hydrogen; 0.33 at index 8, the atomic number of oxygen; and 0 everywhere else.) The output is a 512-dimensional embedding $E_f = E_f(\mathbf{x}_f)$.

The architecture is a simple MLP, in which every linear layer is followed by BatchNorm (which improves stability and convergence speed)⁵⁶ and ReLU. The only exception is that the last layer does not use ReLU. We reiterate that we use a variational framework for regressing $E_f(\mathbf{x}_f)$, which technically necessitates two versions of the encoder, one for $\mu_f(\mathbf{x}_f)$, and one for $\sigma_f(\mathbf{x}_f)$.

The feature fusion network is shown in Fig. 6, Panel c; this network is designed to integrate the XRD and chemical information into one unified representation. The inputs are the concatenated embeddings from the XRD encoders and formula encoders, such that we have a 1024-dimensional combined embedding. This combined embedding then gets passed through two MLPs with four linear layers each, and BatchNorm⁵⁶ and ReLU following every linear layer. The outputs are two 512-dimensional embeddings, one for multiplicative interactions (labeled $y(E_d, E_f)$), the other for additive interactions (labeled $\beta(E_d, E_f)$) with the positional encoding (described in next section).

The positional encoder is shown in Fig. 6, Panel D; its function is to convert the positional information into a format that can meaningfully interact with the aforementioned XRD and chemical information. It takes in the (x, y, z) coordinates as input. The inputted coordinates are normalized and centered, such that $-0.5 \leq x, y, z \leq +0.5$. The output is a 512-dimensional positional embedding.

This approach of querying specific coordinates as compared to directly predicting a voxel grid is advantageous, because in principle, it allows us to represent electron density maps with arbitrary precision. (That being said, in our work, the maximum resolution is the 50^3 grid).

To process the input, we use modified random Fourier features⁴¹, according to the formula:

$$p([x, y, z]) = \text{MLP}([\sin(\mathbf{B} \cdot [x, y, z]^T), \cos(\mathbf{B} \cdot [x, y, z]^T)]) \quad (3)$$

We generate the frequency matrix $\mathbf{B} \in \mathbb{R}^{m \times 3}$, where each $B_{ij} \sim N(0, \sigma^2)$. (We set $m = 128$, $\sigma = 3$.) Then, we calculate

$\mathbf{x}' = \mathbf{B} \cdot [x, y, z]^T$, which represents linear combinations of each of the coordinates. Then, we calculate $\sin(x')$ and $\cos(x')$ and concatenate them to get a $2m$ -dimensional pseudo-Fourier series representation. We employ this coordinate transformation for two reasons: (1) it approximates a high-dimensional Fourier series of the charge density map, which allows the model to capture high-frequency features (shown via Neural Tangent Kernel theory^{41,65}); (2) the cosine (periodic even function) and sin (periodic odd function) parameterizations allow us to encode the many inherent symmetries¹ of crystals. Finally, we pass the $2m$ -dimensional pseudo-Fourier series representation through two linear layers, with a BatchNorm⁵⁶ and ReLU in between; to get our positional encoding $p([x, y, z])$.

The feature conditioner is shown in Fig. 6, Panel E. This part of the architecture combines the information from all the previous branches: XRD, chemical, and positional; such that it can be fed into the final charge density regressor. It takes as input the multiplicative embeddings $\gamma(\mathbf{E}_d, \mathbf{E}_f)$, additive embeddings $\beta(\mathbf{E}_d, \mathbf{E}_f)$, and positional encoding $p([x, y, z])$. It outputs \mathbf{P} , the 512-dimensional feature-conditioned positional encoding.

The feature-conditioned positional encoding \mathbf{P} is calculated as:

$$\mathbf{P} = \gamma(\mathbf{E}_d, \mathbf{E}_f) * p([x, y, z]) + \beta(\mathbf{E}_d, \mathbf{E}_f) \quad (4)$$

This is known as the feature-wise linear modulation (FiLM)⁶². It is effective because it allows us to have both multiplicative and additive interactions during feature conditioning, which increases expressivity. (In contrast, traditional concatenation-based approaches to feature conditioning are shown to only simulate additive interactions⁶⁶).

The charge density regressor is shown in Fig. 6, Panel f; it is responsible for predicting the final structure of the crystal. The input is \mathbf{P} , the feature-conditioned positional encoding. The output is the charge density at the corresponding (x, y, z) coordinates that \mathbf{P} was generated from. One major advantage of designing the network to continuously output electron density at arbitrary query points (as opposed to outputting a set of discrete atomic coordinates, for instance) is that we can predict structures without needing to know a priori how many atoms are contained in the material.

The architecture is a MLP with BatchNorm⁵⁶ and ReLU after every layer, except for the final layer. It also uses skip connections to encourage feature reuse, inspired by DeepSDF⁴³ and NeRF⁴⁰.

Training process

We minimize L1 Loss on the predicted charge densities, averaged over the entire batch:

$$L_1(\rho_{\text{pred}}, \rho_{\text{gt}}) = |\rho_{\text{pred}} - \rho_{\text{gt}}| \quad (5)$$

Minimizing this loss encourages the predicted output to match the ground truth output. We call this the reconstruction loss.

We also simultaneously minimize a KL-Divergence Loss on the predicted mean $\mu_d(\mathbf{x}_d)$, $\mu_f(\mathbf{x}_f)$ and standard deviation $\sigma_d(\mathbf{x}_d)$, $\sigma_f(\mathbf{x}_f)$ of the distribution of embedding vectors $E_d(\mathbf{x}_d)$, $E_f(\mathbf{x}_f)$ ^{38,45,67}, similar to that in β -VAE⁴⁵:

$$\begin{aligned} \beta L_{\text{KL}}(\mu(\mathbf{x}), \sigma(\mathbf{x})) &= \frac{\beta}{N} \sum_{i=1}^N (1 + \log(\sigma_i(\mathbf{x})^2) - \mu_i(\mathbf{x})^2 - \sigma_i(\mathbf{x})^2) \\ &\propto D_{\text{KL}}(q_\phi(E(\mathbf{x})|\mathbf{x})||p(E(\mathbf{x}))) \end{aligned} \quad (6)$$

where $E(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon * \sigma(\mathbf{x})$, $N = |E(\mathbf{x})| = |\mu(\mathbf{x})| = |\sigma(\mathbf{x})| = 512$ is the dimensionality of the embedding vector, $q_\phi(E(\mathbf{x})|\mathbf{x})$ is the conditional distribution of the embedding vectors given the inputted XRD pattern or chemical formula, and β is a weighting parameter given to the loss. This closed form is possible because we parameterize $p(E(\mathbf{x}))$ as $N(0, \mathbf{I})$, following Kingma and Welling, who include a derivation in their paper³⁸. Intuitively, minimizing this loss encourages the XRD and formula embedding vectors to match multivariate Gaussian distributions, which not only smoothens the

latent space, but encourages variation in the outputs, such that we can conduct an iterative refinement process in this one-to-many problem.

Thus, the total loss to be minimized is the sum of Equations (5) and (6):

$$L = L_1(\rho_{\text{pred}}, \rho_{\text{gt}}) + \beta L_{\text{KL}}(\mu_d(\mathbf{x}_d), \sigma_d(\mathbf{x}_d)) + \beta L_{\text{KL}}(\mu_f(\mathbf{x}_f), \sigma_f(\mathbf{x}_f)) \quad (7)$$

The β term tweaks the balance between the reconstruction and the KL terms. Empirically, we set $\beta = 0.05$.

We considered incorporating XRD adherence into our loss function⁶⁸, but we ultimately did not. This choice was made because it is not straightforward to compute the diffraction pattern from the CMED representation directly without carrying out an inverse transform, and we wanted to use a more direct objective for reconstruction performance, like L1 Loss.

We train our model to minimize the total loss from Equation (7) for 1500 epochs, with 128 crystals per batch at a resolution of 10^3 sampled charge densities per crystal. The charge densities are sampled via stratified bin sampling, where $x, y, z \sim \text{Uniform}[\frac{i}{S}, \frac{i+1}{S}]$ (we set $S = 10$) – this probabilistically allows us, over the course of the optimization procedure, to capture fine-resolution details of the electron density field, despite processor memory limits for individual batches⁴⁰.

We use the Adam⁶⁹ optimizer. We follow a cosine annealing schedule with warm restarts⁷⁰, in which the learning rate decays from 10^{-3} to 10^{-6} , then increases back to 10^{-3} and decays again to 10^{-6} over another cycle that has double the number of epochs: this helps the optimization procedure break out of local minima. The initial cycle length is 100 epochs, and increases to 200, 400, and 800 on the subsequent cycles, to constitute the 1500 total epochs.

As data augmentation, we randomly add small Gaussian perturbations from $N(0, 0.001^2 \mathbf{I})$ to the inputted XRDs and chemical formula ratios (the perturbed input undergoes a ReLU, since we cannot have negative peaks or ratios). We also randomly shift the XRD patterns by less than 0.6° .

We save the version of the model that has the highest SSIM⁴⁷ score on the validation set at the end of each epoch, where the model is given two guesses for each structure, and the rotation (24 ways) of the predicted structure that gives the highest SSIM score with the ground truth is used.

Evaluation setup

We run through the testing dataset, and give the model 5 tries (via sampling from the latent space in the variational framework) to predict each crystal structure. (We give the model multiple tries because crystallography is typically an iterative refinement process, so we consider our model successful if it can give a good guess.) For each guess, we rotate the predicted crystal 24 ways (in multiples of 90° about the x, y, z unit cell axes) and take the best SSIM⁴⁷ and PSNR⁴⁸ over all these rotations, as compared to the ground truth crystal. Finally, we report the best results over all rotations of all guesses of each crystal structure.

At evaluation time, we sample evenly to get a $50 \times 50 \times 50$ charge density map (i.e., 3D grid) for each crystal. We then use 3D SSIM⁴⁷ and PSNR⁴⁸ as our evaluation metrics on the resultant 3D grid.

SSIM ranges from 0 to 1, where higher is better. It compares the structural similarity of the ground truth charge density map with the predicted charge density map. SSIM is calculated over patches of the 3D structure with a sliding cubic window of side length 7, and then averaged over all such patches. The patch-wise formula is:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

where \mathbf{x}, \mathbf{y} are the spatially corresponding patches of the ground truth and predicted electron density maps, μ_x, μ_y are the mean charge densities (i.e., intensity) in those patches, σ_x, σ_y are the standard deviations of the charge densities (i.e., contrast) in those patches, σ_{xy} is the covariance between the position-wise charge densities in those patches, and C_1, C_2 are small constants for numerical stability.

PSNR stands for peak signal-to-noise ratio, where higher is better. Its value is theoretically infinite. Typically, values above 30 are considered good⁴⁹. PSNR is calculated as follows⁷¹:

$$PSNR(\mathbf{X}_{gt}, \mathbf{X}_{pred}) = 10 * \log_{10} \left(\frac{(\max(\mathbf{X}_{gt}) - \min(\mathbf{X}_{gt}))^2}{MSE(\mathbf{X}_{gt}, \mathbf{X}_{pred})} \right) \quad (9)$$

where \mathbf{X}_{gt} , \mathbf{X}_{pred} are the ground truth and predicted charge density maps, respectively.

Formula ablation experiment

To reduce the computational burden of these formula ablation studies, we make a few modifications. In the optimization loop, we only train for 700 total epochs, use 8³ samples per crystal, and have only 1 sample from the latent space per validation crystal. Additionally, in testing, we give the model 3 tries (instead of 5) to predict each crystal via variational sampling. We can make these modifications because the purpose of these ablation experiments is to compare the predictive ability of the model at varying degrees of chemical composition information, rather than to optimize the model to perfect predictive ability. For fair comparison, we also recalculate the baseline model performance according to these pared-down protocols.

Data availability

Please visit the Materials Project website⁴⁶ to obtain the data: <https://nextgen.materialsproject.org/materials>.

Code availability

Code is available at <https://github.com/gabeguo/deep-crystallography-public>.

Received: 30 January 2024; Accepted: 20 August 2024;

Published online: 07 September 2024

References

1. Giacovazzo, C. *Fundamentals of crystallography*, vol. 7 (Oxford University Press, USA, 2002).
2. Hammond, C. *The basics of crystallography and diffraction*, vol. 21 (International Union of Crystallography texts on crystallography, 2015).
3. Lipson, H. & Beevers, C. The crystal structure of the alums. *Proc. R. Soc. Lond. Ser. A-Math. Phys. Sci.* **148**, 664–680 (1935).
4. Dinnebier, R. E. & Billinge, S. J. *Powder diffraction: theory and practice* (Royal Society of Chemistry, 2008).
5. Daniel, V. & Lipson, H. S. An x-ray study of the dissociation of an alloy of copper, iron and nickel. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* **181**, 368–378 (1943).
6. Lipson, H. S. & Stokes, A. The structure of graphite. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* **181**, 101–105 (1942).
7. Lipson, H. *The study of metals and alloys by X-ray powder diffraction methods* (University College Cardiff Press Cardiff, 1984).
8. Billinge, S. J. & Levin, I. The problem with determining atomic structure at the nanoscale. *Science* **316**, 561–565 (2007).
9. David, W. I. F. & Shankland, K. Structure determination from powder diffraction data. *Acta Crystallogr A Found. Crystallogr* **64**, 52–64 (2008).
10. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
11. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using alphafold2. *Nat. Commun.* **13**, 1265 (2022).
12. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
13. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
14. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* **25** (2012).
15. Bojarski, M. et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316* (2016).
16. Amodei, D. et al. Deep speech 2: end-to-end speech recognition in english and mandarin. In *International conference on machine learning*, 173–182 (PMLR, 2016).
17. Liu, C.-H., Tao, Y., Hsu, D., Du, Q. & Billinge, S. J. L. Using a machine learning approach to determine the space group of a structure from the atomic pair distribution function. *Acta Cryst. A* **75**, 633–643 (2019).
18. Oviedo, F. et al. Fast and interpretable classification of small x-ray diffraction datasets using data augmentation and deep neural networks. *npj Comput. Mater.* **5**, 60 (2019).
19. Suzuki, Y. et al. Symmetry prediction and knowledge discovery from x-ray diffraction patterns using an interpretable machine learning approach. *Sci. Rep.* **10**, 21790 (2020).
20. Park, W. B. et al. Classification of crystal structure using a convolutional neural network. *IUCrJ* **4**, 486–494 (2017).
21. Lee, J.-W., Park, W. B., Lee, J. H., Singh, S. P. & Sohn, K.-S. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic xrd powder patterns. *Nat. Commun.* **11**, 86 (2020).
22. Aguiar, J., Gong, M. L., Unocic, R., Tasdizen, T. & Miller, B. Decoding crystallography from high-resolution electron imaging and diffraction datasets with deep learning. *Sci. Adv.* **5**, eaaw1949 (2019).
23. Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 2775 (2018).
24. Tiong, L. C. O., Kim, J., Han, S. S. & Kim, D. Identification of crystal symmetry from noisy diffraction patterns by a shape analysis and deep learning. *npj Comput. Mater.* **6**, 196 (2020).
25. Garcia-Cardona, C. et al. Learning to predict material structure from neutron scattering data. In *2019 IEEE International Conference on Big Data (Big Data)*, 4490–4497 (IEEE, 2019).
26. Merker, H. A. et al. Machine learning magnetism classifiers from atomic coordinates. *iScience* **25**, 10 (2022).
27. Maffettone, P. M. et al. Crystallography companion agent for high-throughput materials discovery. *Nat. Comput. Sci.* **1**, 290–297 (2021).
28. Szymanski, N. J., Bartel, C. J., Zeng, Y., Tu, Q. & Ceder, G. Probabilistic deep learning approach to automate the interpretation of multi-phase diffraction spectra. *Chem. Mater.* **33**, 4204–4215 (2021).
29. Uryu, H. et al. Deep learning enables rapid identification of a new quasicrystal from multiphase powder diffraction patterns. *Adv. Sci.* **11**, 2304546 (2024).
30. Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023).
31. Merchant, A. et al. Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
32. Yang, M. et al. Scalable diffusion for materials generation. In *International Conference on Learning Representations* (2024).
33. Hernández-García, A. et al. Crystal-gflownet: sampling materials with desirable properties and constraints. In *AI for Accelerated Materials Design-NeurIPS 2023 Workshop* (2023).
34. Pan, T., Jin, S., Miller, M. D., Kyrilidis, A. & Phillips, G. N. A deep learning solution for crystallographic structure determination. *IUCrJ* **10**, 487–496 (2023).
35. Pan, T. et al. "CrysFormer: Protein structure determination via Patterson maps, deep learning, and partial structure attention." *Structural Dynamics* **11**.4 (2024).
36. Barbarin-Bocahu, I. & Graille, M. The x-ray crystallography phase problem solved thanks to alphafold and rosettafold models: a case-study report. *Acta Crystallogr. Sect. D: Struct. Biol.* **78**, 517–531 (2022).

37. Kjær, E. T. S. et al. DeepStruc: towards structure solution from pair distribution function data using deep generative models. *Digit. Discov.* **2**, 69–80 (2023).

38. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In *International Conference on Learning Representations* (2014).

39. Yu, A., Ye, V., Tancik, M. & Kanazawa, A. Pixelnerf: neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4578–4587 (2021).

40. Mildenhall, B. et al. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **65**, 99–106 (2021).

41. Tancik, M. et al. Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Inf. Process. Syst.* **33**, 7537–7547 (2020).

42. Sitzmann, V., Zollhöfer, M. & Wetzstein, G. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Adv. Neural Inform. Process. Syst.* **32** (2019).

43. Park, J. J., Florence, P., Straub, J., Newcombe, R. & Lovegrove, S. DeepSDF: learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 165–174 (2019).

44. Hoffmann, J. et al. Data-driven approach to encoding and decoding 3-d crystal structures. *arXiv preprint arXiv:1909.00949* Preprint at <https://arxiv.org/abs/1909.00949>, <https://github.com/hoffmannjordan/Encoding-Decoding-3D-Crystals> (2019).

45. Higgins, I. et al. beta-vae: learning basic visual concepts with a constrained variational framework. In *International conference on learning representations* (2016).

46. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

47. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

48. Hore, A. & Ziou, D. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, 2366–2369 (IEEE, 2010).

49. Bull, D. & Zhang, F. *Intelligent image and video compression: communicating pictures* (Academic Press, 2021).

50. Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the inorganic crystal structure database (icsd): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B: Struct. Sci.* **58**, 364–369 (2002).

51. Hafner, J. Ab-initio simulations of materials using vasp: density-functional theory and beyond. *J. Comput. Chem.* **29**, 2044–2078 (2008).

52. The materials project workshop. https://workshop.materialsproject.org/lessons/01_website_walkthrough/website_walkthrough/.

53. Raja, P. & Barron, A. R. *Physical methods in chemistry and nano science* (Rice University, 2019).

54. De Graef, M. & McHenry, M. E. *Structure of materials: an introduction to crystallography, diffraction and symmetry* (Cambridge University Press, 2012).

55. Diffraction patterns: how diffraction patterns are calculated on the materials project (mp) website. <https://docs.materialsproject.org/methodology/materials-methodology/diffraction-patterns>.

56. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, 448–456 (PMLR, 2015).

57. Charge density: obtaining the charge density shown on the materials project (mp) website. <https://docs.materialsproject.org/methodology/materials-methodology/charge-density>.

58. Chgcar. <https://www.vasp.at/wiki/index.php/CHGCAR>.

59. Shen, J.-X. et al. A representation-independent electronic charge density database for crystalline materials. *Sci. Data* **9**, 661 (2022).

60. Chitturi, S. R. et al. Automated prediction of lattice parameters from x-ray powder diffraction patterns. *J. Appl. Crystallogr.* **54**, 1799–1810 (2021).

61. Guccione, P., Diacono, D., Toso, S. & Caliandro, R. Towards the extraction of the crystal cell parameters from pair distribution function profiles. *IUCrJ* **10**, 610–623 (2023).

62. Perez, E., Strub, F., De Vries, H., Dumoulin, V. & Courville, A. Film: visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 32 (2018).

63. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).

64. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. In *Advances in NIPS 2016 Deep Learning Symposium* Preprint at <https://arxiv.org/abs/1607.06450> (2016).

65. Jacot, A., Gabriel, F. & Hongler, C. Neural tangent kernel: convergence and generalization in neural networks. *Adv. Neural Inform. Process. Syst.* **31** (2018).

66. Dumoulin, V. et al. Feature-wise transformations. *Distill* <https://distill.pub/2018/feature-wise-transformations> (2018).

67. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).

68. Lee, J., Oba, J., Ohba, N. & Kajita, S. Creation of crystal structure reproducing x-ray diffraction pattern without using database. *npj Comput. Mater.* **9**, 135 (2023).

69. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *International Conference on Learning Representations* (2015).

70. Loshchilov, I. & Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations* (2017).

71. Van der Walt, S. et al. Scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).

Acknowledgements

Work in the Lipson group was supported by U.S. National Science Foundation under AI Institute for Dynamical Systems grant 2112085. Work in the Billinge group was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (DOE-BES) under contract No. DE-SC0024141.

Author contributions

H.L. and S.J.L.B. proposed the research. G.G., B.C., and H.L. designed the system architecture. G.G., S.J.L.B., and H.L. designed the experiment systems. G.G. wrote the code. J.G., A.H.Y., and A.R. ensured the quality of the code. B.C. and L.L. conducted initial explorations. J.G., L.L., and A.R. provided helpful insights. S.J.L.B., H.L., B.C., J.G., L.L., and G.G. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-024-01401-8>.

Correspondence and requests for materials should be addressed to Gabe Guo, Simon J. L. Billinge or Hod Lipson.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024