Supplementary Information

Supplementary Information for:

The Loss Landscape of Powder X-Ray Diffraction-Based Structure Optimization Is Too Rough for Gradient Descent

Authors: Nofit Segal, Akshay Subramanian, Mingda Li, Benjamin Kurt Miller, Rafael Gomez Bombarelli

1.1 Selected Structures

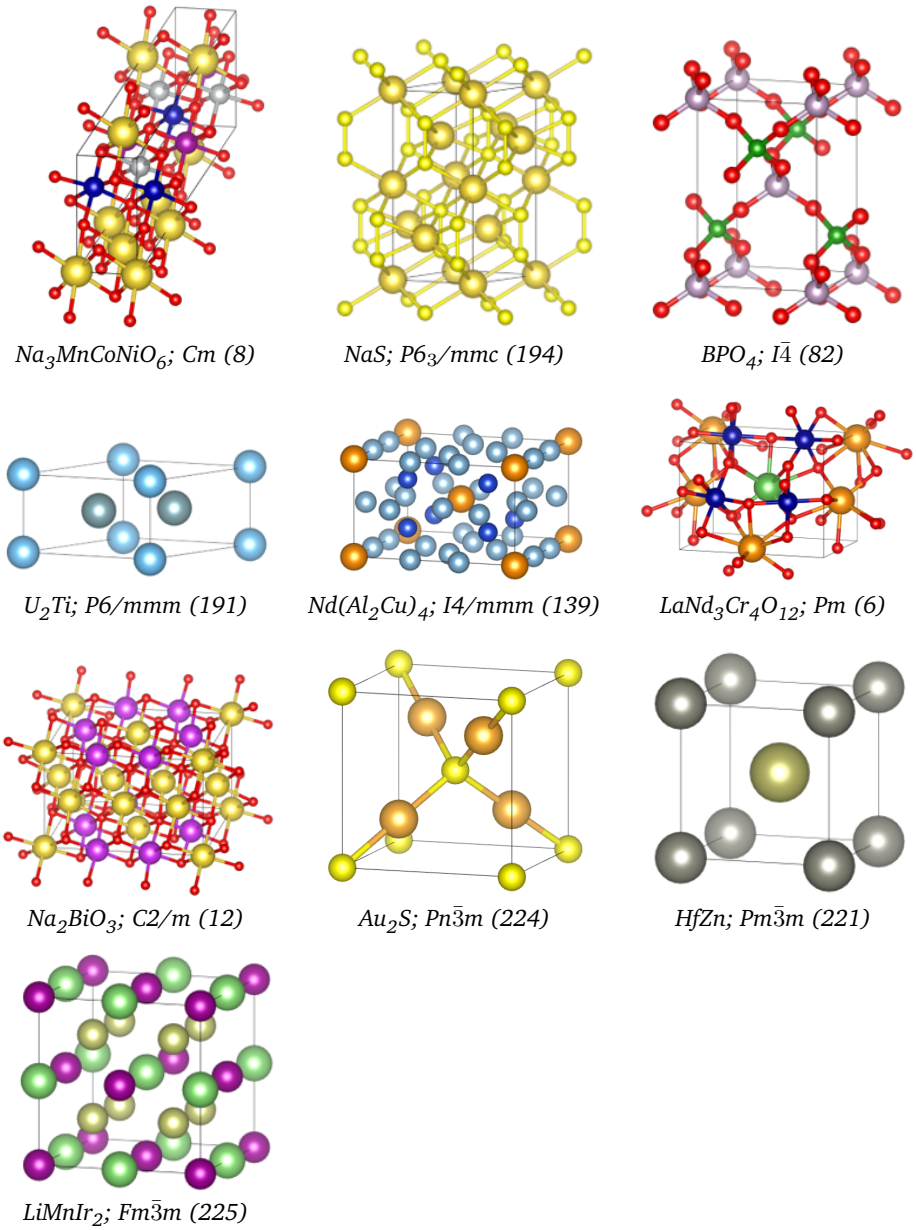


Fig. S1 Crystal structures used in this study. Each structure is labeled with its chemical formula, space group symbol, and space group number.

1.2 XRD Representation

We compute diffraction patterns from the structure factor contributions of each atomic site, in a differentiable way. Lattice parameters are converted to a real-space cell, the reciprocal lattice is derived, and all allowed Miller indices within the maximum scattering vector are generated. For each (hkl), reciprocal distances and diffraction angles are calculated, elemental scattering factors are retrieved and weighted by site occupancies, and intensities are obtained by squaring the modulus of the summed structure factor with a Lorentz–polarization correction.

We calculate XRD peak profiles using the Pseudo-Voigt approximation, which models the peak shape as a linear combination of Gaussian and Lorentzian components:

$$pV(x) = \eta\, G(x) + (1-\eta)\, L(x)$$

where $G(x)$ is the Gaussian function, $L(x)$ is the Lorentzian function, and $\eta \in [0,1]$ is the mixing parameter controlling the relative contributions.

1.2.1 Gaussian and Lorentzian peak shapes

For a peak centered at $2\theta_0$, the Gaussian and Lorentzian components are given by:

$$G(x) = \exp\left[-\frac{4\ln 2\,(x-2\theta_0)^2}{H_G^2}\right],$$

$$L(x) = \frac{1}{1+\frac{4(x-2\theta_0)^2}{H_L^2}},$$

where H_G and H_L are the full widths at half maximum (FWHM) for the Gaussian and Lorentzian profiles, respectively.

1.2.2 Caglioti parameters

In practice, peak broadening in XRD is described by the Caglioti relation:

$$H^2(2\theta) = U\tan^2\theta + V\tan\theta + W,$$

where U, V, and W are the Caglioti parameters. This equation gives the squared FWHM as a function of diffraction angle, and is applied separately for the Gaussian and Lorentzian widths, i.e., $H_G(2\theta)$ and $H_L(2\theta)$. The parameters account for instrumental and sample-dependent broadening effects.

1.2.3 Final pattern representation

We compute the total XRD pattern by summing $pV(x)$ contributions from all Bragg reflections over $2\theta \in [0^\circ, 90^\circ]$, and then discretize the intensity into bins of width 0.01°. We adopt Caglioti parameterss $U = 0.1, V = 0.01, W = 0.1$ and $\eta = 0.1$. This produces a fixed-length xrd vector, $\mathbf{x}$, of size 9000 for each structure.

1.3 Symmetry Projectors by Crystal Family

For each crystal family, the symmetry projector $\mathcal{P}$ maps the given lattice parameters $(a,b,c,\alpha,\beta,\gamma)$ to the symmetrized parameters consistent with the family:

$$\mathcal{P}_{\text{cubic}}(a,b,c,\alpha,\beta,\gamma) = \big(\bar{a},\bar{a},\bar{a},\ 90^\circ,\ 90^\circ,\ 90^\circ\big), \qquad \bar{a} = \frac{a+b+c}{3}$$

$$\mathcal{P}_{\text{hexagonal}}(a,b,c,\alpha,\beta,\gamma) = \big(\bar{a},\bar{a},c,\ 90^\circ,\ 90^\circ,\ 120^\circ\big), \qquad \bar{a} = \frac{a+b}{2}$$

$$\mathcal{P}_{\text{tetragonal}}(a,b,c,\alpha,\beta,\gamma) = \big(\bar{a},\bar{a},c,\ 90^\circ,\ 90^\circ,\ 90^\circ\big), \qquad \bar{a} = \frac{a+b}{2}$$

$$\mathcal{P}_{\text{orthorhombic}}(a,b,c,\alpha,\beta,\gamma) = \big(a,b,c,\ 90^\circ,\ 90^\circ,\ 90^\circ\big)$$

$$\mathcal{P}_{\text{monoclinic}}(a,b,c,\alpha,\beta,\gamma) = \big(a,b,c,\ 90^\circ,\ \beta,\ 90^\circ\big)$$

$$\mathcal{P}_{\text{triclinic}}(a,b,c,\alpha,\beta,\gamma) = \big(a,b,c,\ \alpha,\ \beta,\ \gamma\big)$$

1.4 Implementation Details

Each optimization run was performed for a maximum of 30,000 iterations, with batch sizes of 16. For lattice distortions, the learning rates explored were between 0.0005 and 0.01. For coordinate distortions, the learning rates explored were between 1e-04 to 1e-06. The XRD parameters U, V, and W (see Appendix Section 1.2) were kept fixed at $[0.1, 0.01, 0.1]$, the same values used for generating the ground truth XRD patterns used for reference. Optimization was terminated early if the convergence tolerance of 10^{-4} in the loss function between 1000 steps was reached.

For lattice distorted structures, the lattice lengths and angles were optimized jointly while atomic coordinates were held fixed. For coordinate distortions, only coordinates were optimized. The loss functions considered included negative cosine similarity, negative entropy similarity, negative cross correlation, and mean squared error (MSE). In the MSE case, to ensure numerical stability, reflections were subject to physics-based clipping within Miller indices up to $[h, k, l]_{\max} = 6$.

For comparison with relaxation-based methods, structures were relaxed using CHGNet. For distorted coordinates, relaxation was done with the BFGS optimizer. For the distorted lattice case, relaxation was done with the FrechetCellFilter and BFGS optimizer. The relaxation process was terminated once the maximum atomic force ($f_{\max}$) fell below 0.02 eV/Å.

1.5 Correlations between AMD and XRD Similarity

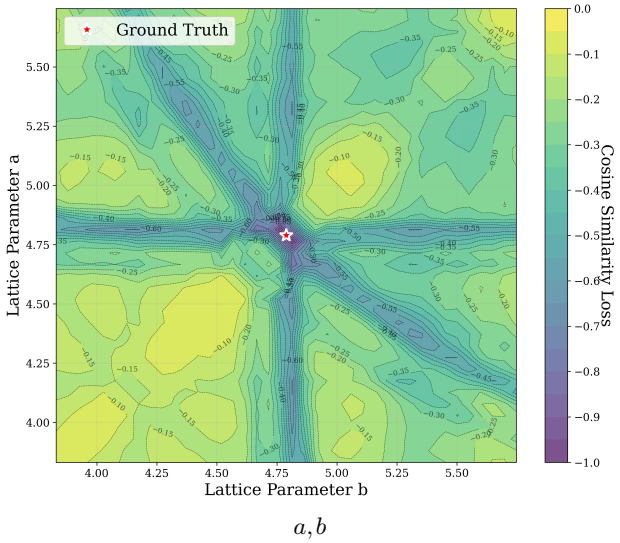
Table S1 Pearson Correlation Between Average Minimum Distances (AMD) and XRD Similarity Metrics for the Different Noise Types and Levels.

Noise Type	Noise Level	Cosine Similarity	MSE	Entropy Similarity	Cross Correlation	Cosine Similarity + Constraints
Lattice	0.01	0.76	0.59	0.31	0.58	**0.79**
	0.05	0.43	0.4	0.53	0.32	**0.83**
	0.1	0.36	0.27	0.39	0.29	**0.76**
	0.15	0.16	0.1	0.23	0.22	**0.72**
	0.2	0.05	0.1	0.13	0.13	**0.65**
Coordinates	0.001	-0.64	**0.23**	-0.6		-
	0.005	-0.72	**0.27**	-0.55		-
	0.01	-0.64	**0.46**	-0.64	-0.82	-
	0.05	-0.02	**0.51**	-0.27	-0.18	-
	0.1	**0.52**	0.48	0.43	0.41	-

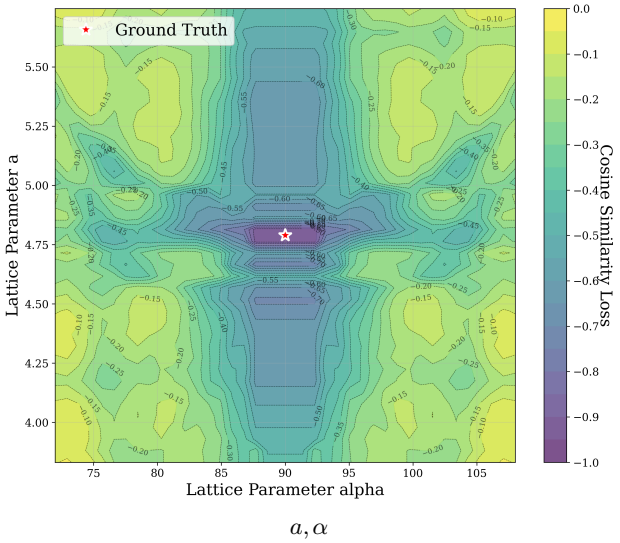
Table S2 Spearman Correlation Between Average Minimum Distances (AMD) and XRD Similarity Metrics for the Different Noise Types and Levels.

Noise Type	Noise Level	Cosine Similarity	MSE	Entropy Similarity	Cross Correlation	Cosine Similarity + Constraints
Lattice	0.01	**0.93**	0.61	0.65	0.61	0.65
	0.05	0.47	0.43	0.58	0.35	**0.89**
	0.1	0.39	0.29	0.39	0.26	**0.77**
	0.15	0.18	0.1	0.25	0.2	**0.67**
	0.2	0.3	0.13	0.12	0.13	**0.55**
Coordinates	0.001	-0.67	**0.74**	-0.64		-
	0.005	-0.64	**0.78**	-0.51		-
	0.01	-0.55	**0.84**	-0.62	-0.79	-
	0.05	-0.06	**0.63**	-0.46	-0.27	-
	0.1	0.5	**0.52**	0.28	0.26	-

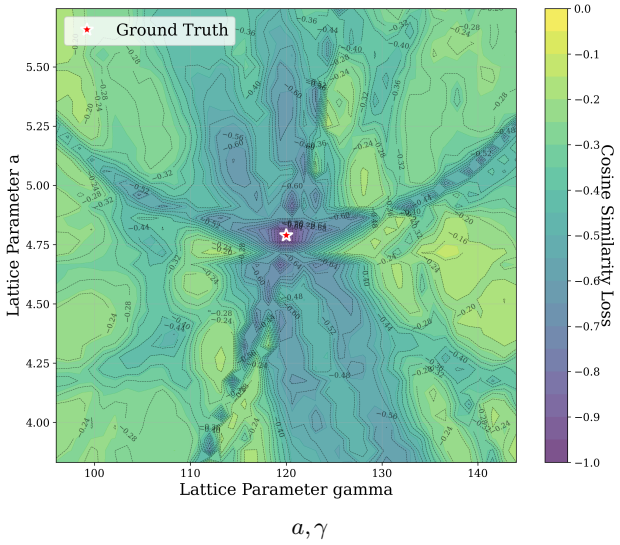
1.6 2D Landscape of XRD-Based Optimization



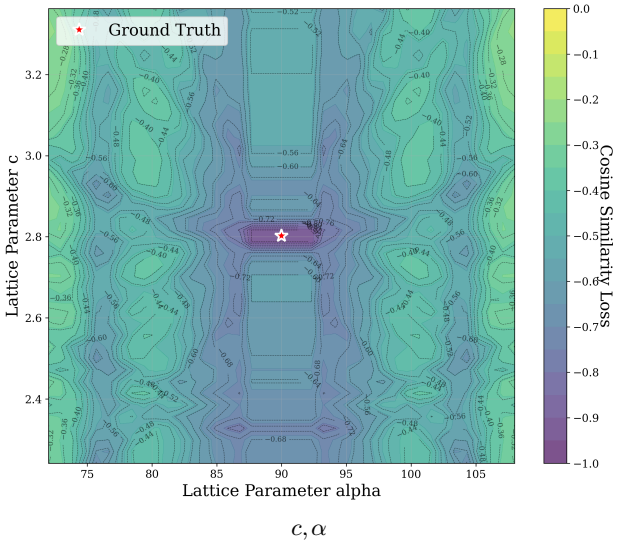
a, b



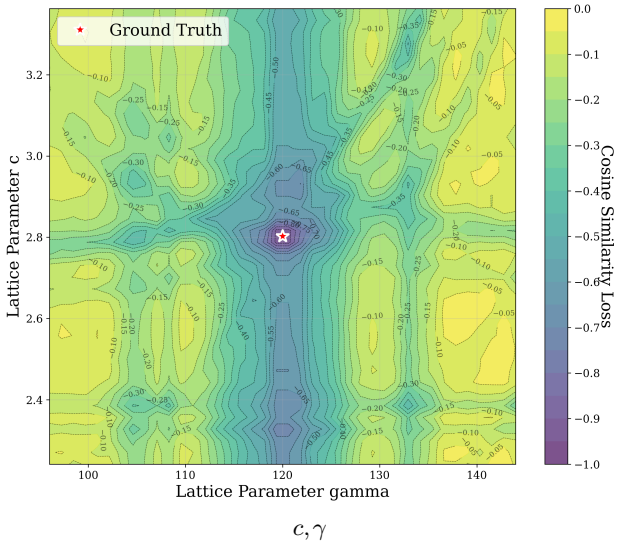
a, α



a, γ

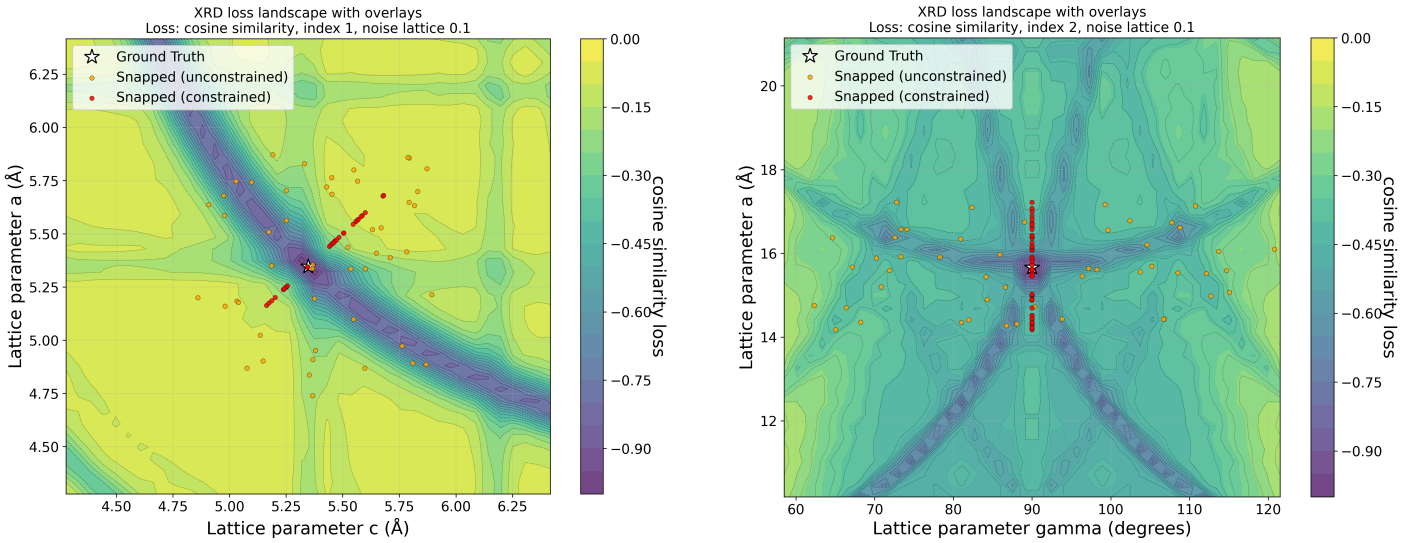


c, α



c, γ

Fig. S2 All 2D landscape slices of XRD cosine similarity of U_2Ti. Each slice is labeled with the distorted parameters. Due to the hexagonal symmetry of the structure, some slices are redundant and thus obscured.



(a) **(b)**

Fig. S3 2D landscape of XRD cosine similarity (CS) loss as a function of lattice parameters, with overlaid endpoints of gradient descent trajectories for both unconstrained and symmetry-constrained optimization. (a) Lattice parameters a and b [Å] of cubic Au_2S are perturbed. (b) Lattice parameters a [Å] and γ [°] of monoclinic $Na_3MnCoNiO_6$ are perturbed. Each slice is defined by varying two lattice parameters while fixing the remaining parameters at their ground-truth values. Trajectory endpoints are projected onto the same two-dimensional subspace. Notably, endpoints do not consistently align with the local minima visible in the 2D slices, reflecting the fact that optimization occurs in the full six-dimensional space. Consequently, failures may correspond to basins that are not apparent in the projected landscape. These results illustrate that while symmetry constraints reduce the effective search space, the XRD similarity landscape remains non-convex even within the constrained manifold.

1.7 XRD Similarity vs. Geometrical Agreement

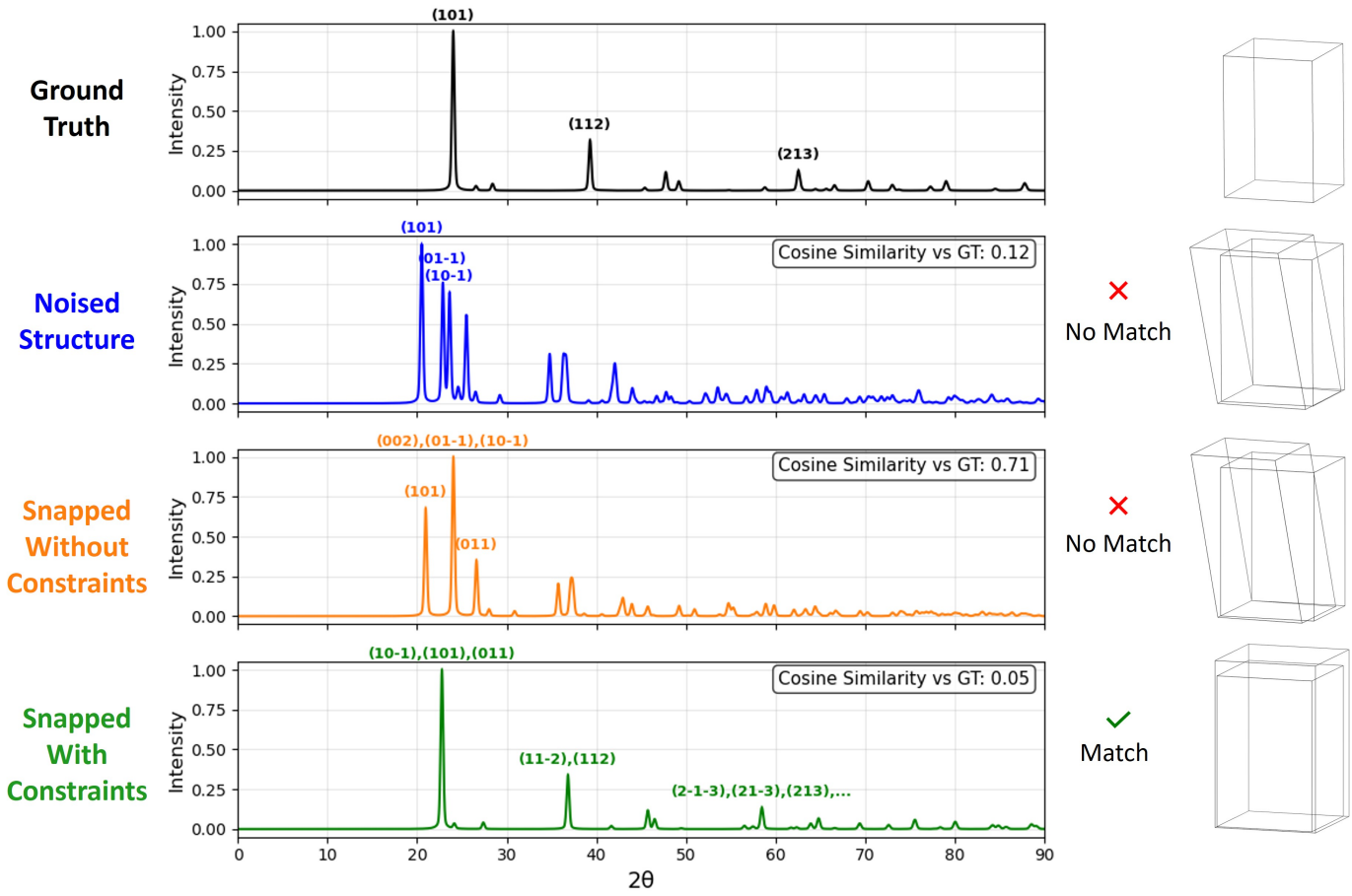


Fig. S4 Lattice and XRD patterns of BPO_4. Each row shows the lattice parameters, corresponding XRD pattern, and unit cell relative to the ground truth. From top to bottom: ground truth; distorted lattice structure with 0.1 noise level; result of XRD-based GD optimization without constraints; and result of XRD-based GD optimization with symmetry-based constraints. For each case, the cosine similarity to the ground truth pattern and the structure match status are reported.

1.8 2D Landscape of Energy-Based Optimization

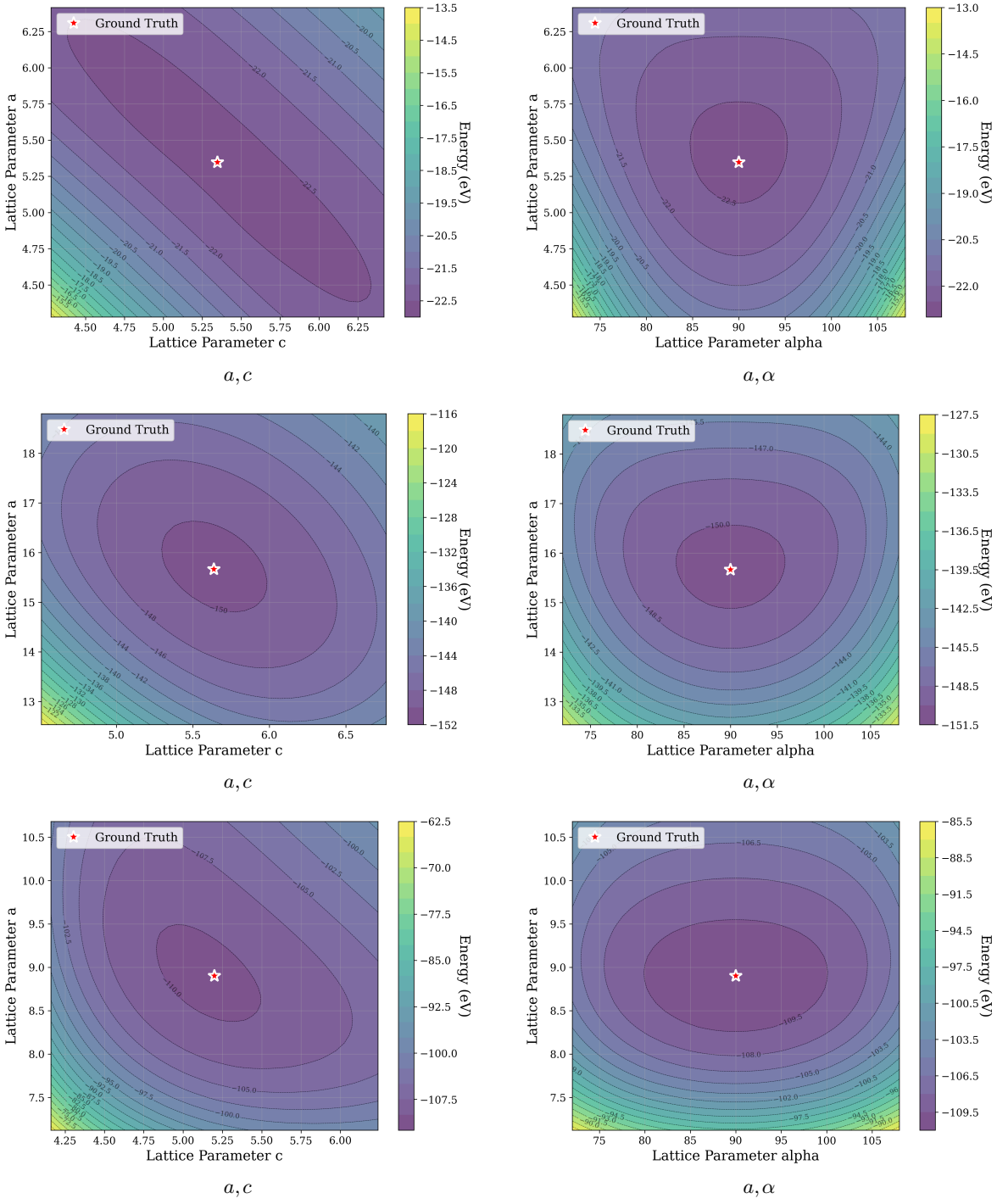


Fig. S5 2D landscape slices of Energy Relaxation for different structures from our study. (Top) Cubic Au_2S. (Middle) Monoclinic $Na_3MnCoNiO_6$. (Bottom) Tetragonal $Nd(Al_2Cu)_4$. Each slice is labeled with the distorted lattice parameters.