Supplementary Information for

Quantifying Symmetry Breaking as a Design Variable for Giant Altermagnetic Spin Splitting

Kichan Chun and Gunn Kim*

Department of Physics, Sejong University, Seoul 05006, Republic of Korea
Hybrid Materials Research Center, Sejong University, Seoul 05006, Republic of Korea
*E-mail: gunnkim@sejong.ac.kr

1 Complete Descriptor Set

Table S1 lists all 52 descriptors used in the XGBoost model. MSBI, MPF, and the p/d electron ratio are defined in the main text. Two additional descriptors not defined there are specified below.

Asymmetry inhomogeneity (σ_{inhom}). Let $D_{\mathrm{min}} = \min(D_d, D_I)$ denote the smaller of the two motif deviations defined in the main text, and let s_{opt} denote the optimal one-to-one Hungarian assignment that realizes D_{min}. Under this assignment, denote the matched ligand vectors on the reference motif by $\{\mathbf{q}_j\}_{j=1}^{N}$ and those on the compared motif by $\{\mathbf{p}_{\mathrm{opt},\, s_{\mathrm{opt}}(j)}\}_{j=1}^{N}$, where $\mathbf{p}_{\mathrm{opt}}$ equals $\mathbf{p}$ when $D_d \leq D_I$ and $-\mathbf{p}$ otherwise. The asymmetry inhomogeneity is the normalized standard deviation of these residuals:

$$\sigma_{\mathrm{inhom}} = \frac{1}{l_0}\sqrt{\frac{1}{N}\sum_{j=1}^{N}(d_j - \bar{d})^2}, \tag{S1}$$

where $d_j = \|\mathbf{q}_j - \mathbf{p}_{\mathrm{opt},\, s_{\mathrm{opt}}(j)}\|$ is the residual magnitude for the j-th matched ligand pair, $\bar{d} = N^{-1}\sum_j d_j$, and l_0 is the mean M–X bond length defined in the main text. A small σ_{inhom} indicates that the symmetry breaking is uniformly distributed across all ligand sites; a large value signals a localized distortion.

Relative motif orientation (θ_{align}). θ_{align} is the minimal rotation angle required to superimpose the ligand coordinates of $\mathcal{M}_1$ onto those of $\mathcal{M}_2$. The optimal atom pairing is determined by the Hungarian algorithm [1] applied to the same ligand-vector sets used for MSBI; the rotation angle is then extracted from the corresponding Kabsch superposition [2]. Unlike MSBI, which deliberately avoids Kabsch alignment to preserve the fixed-frame rotational information (main text, MSBI definition in Methods), θ_{align} is itself defined as the minimal Kabsch rotation angle and therefore serves as a complementary scalar that explicitly quantifies this residual orientation.

Table S1. Complete list of the 52 structural, chemical, and electronic descriptors. Motifs $\mathcal{M}_1$ and $\mathcal{M}_2$ denote the coordination polyhedra of the two antiparallel magnetic sublattices, with ligand-vector sets $\mathcal{A}$ and $\mathcal{B}$. M and X refer to the magnetic and non-magnetic species. All geometric quantities are computed from the atomic coordinates of each dataset entry (fully relaxed parents; strained coordinates for augmentation variants); no electronic-structure-derived inputs enter the feature space. Site-resolved quantities are averaged over the two magnetic sites unless stated otherwise.

Category	Symbol	Description
Symmetry	MSBI	Motif Symmetry Breaking Index. Product of D_d and D_I as defined in the main text, quantifying the degree to which the paired motifs violate both identity and inversion relations.
	σ_{inhom}	Asymmetry inhomogeneity [Equation (S1)]. Normalized standard deviation of atom-pairing residuals, measuring whether the symmetry breaking is uniformly distributed or localized.
Bond	MX_{max}	Maximum M–X bond length in the motif.
	MX_{min}	Minimum M–X bond length.
	MX_{avg}	Mean M–X bond length.
	MX_{std}	Standard deviation of M–X bond lengths.
	MX_{spread}	Radial span, $MX_{\mathrm{max}} - MX_{\mathrm{min}}$.
	MX_{CV}	Coefficient of variation, $MX_{\mathrm{std}}/MX_{\mathrm{avg}}$.
	MX_{elong}	Axial elongation proxy, $MX_{\mathrm{max}} - MX_{\mathrm{avg}}$.
Angle	XMX_{max}	Maximum central bond angle ($\angle X$–M–X).
	XMX_{min}	Minimum central bond angle.
	XMX_{avg}	Mean central bond angle.
	XMX_{std}	Standard deviation of central bond angles.
	XMX_{spread}	Angular span, $XMX_{\mathrm{max}} - XMX_{\mathrm{min}}$.
	XXX_{max}	Maximum peripheral angle ($\angle X$–X–X).
	XXX_{min}	Minimum peripheral angle.
	XXX_{std}	Standard deviation of peripheral angles.
	XXX_{spread}	Angular span, $XXX_{\mathrm{max}} - XXX_{\mathrm{min}}$.
Topology	N_X	Coordination number determined by CrystalNN.
	D_{motif}	Motif dimensionality: 2 (planar) or 3 (polyhedral).
	V_{motif}	Convex-hull volume (area for 2D motifs).
	XX_{long}	Longest ligand–ligand distance within the motif.
	XX_{short}	Shortest ligand–ligand distance within the motif.

Continued on next page

Table S1 — continued from previous page

Category	Symbol	Description
	η_{shape}	Shape anisotropy, $XX_{\mathrm{long}}/XX_{\mathrm{short}}$.
Distortion	σ_{shape}	Normalized standard deviation of ligand–centroid distances.
	δ_{off}	Off-center displacement of M from the ligand centroid, normalized by the mean ligand–centroid radius.
Chemical	Z_M	Atomic number of the magnetic species.
	Z_X	Atomic number of the ligand species.
	χ_M	Pauling electronegativity of M.
	χ_X	Pauling electronegativity of X.
	ΔZ	Atomic-number difference, $Z_M - Z_X$.
	$\lvert\Delta Z\rvert$	Absolute atomic-number difference.
	$\Delta\chi$	Electronegativity difference, $\chi_M - \chi_X$.
	$\lvert\Delta\chi\rvert$	Absolute electronegativity difference; proxy for bond polarity.
Electronic	$n_{d,M}$	Number of d electrons of M (PBE pseudopotential valence; Table S2).
	$n_{p,X}$	Number of p electrons of X (PBE pseudopotential valence; Table S2).
	n_{unpaired}	Number of unpaired d electrons of M, from a formal free-ion high-spin d-count; does not resolve ligand-field-driven low-/high-spin states.
	μ_{spin}	Formal spin-only moment, $\sqrt{n_{\mathrm{unpaired}}(n_{\mathrm{unpaired}}+2)}\,\mu_B$, evaluated from n_{unpaired} above; a composition-level proxy rather than a DFT-computed moment.
	p/d electron ratio	$n_{p,X}/n_{d,M}$; composition-level proxy for p–d hybridization propensity as defined in the main text.
Global	$d_{CC}^{(1)}$	1st nearest motif-centroid distance (periodic images included).
	$d_{CC}^{(2)}$	2nd nearest motif-centroid distance.
	$d_{CC}^{(3)}$	3rd nearest motif-centroid distance.
	$d_{MM}^{(1)}$	1st nearest magnetic-neighbor distance.
	$d_{MM}^{(2)}$	2nd nearest magnetic-neighbor distance.
	$d_{MM}^{(3)}$	3rd nearest magnetic-neighbor distance.
	V_{cell}	Unit-cell volume.

Continued on next page

Table S1 — continued from previous page

Category	Symbol	Description
	MPF	Motif Packing Fraction, $V_{\text{motif}}/V_{\text{cell}}$ (area ratio for 2D motifs) as defined in the main text.
	θ_{align}	Relative motif orientation from Hungarian matching and Kabsch superposition (see text above).
Hybrid	$\Gamma_{\chi\text{-elong}}$	Chemical–geometric coupling, $\lvert\Delta\chi\rvert \cdot MX_{\text{elong}}$.
	$\Delta^{(1)}_{\text{sub}}$	1st-shell sublattice displacement, $d^{(1)}_{MM} - d^{(1)}_{CC}$.
	$\Delta^{(2)}_{\text{sub}}$	2nd-shell sublattice displacement, $d^{(2)}_{MM} - d^{(2)}_{CC}$.
	$\Delta^{(3)}_{\text{sub}}$	3rd-shell sublattice displacement, $d^{(3)}_{MM} - d^{(3)}_{CC}$.

2 Valence Electron Counts for the p/d Ratio

Table S2 lists the d- and p-electron counts used to evaluate the p/d ratio defined in the main text, together with the PBE PAW potential (POTCAR) used for each element in the DFT calculations. Each count is the number of d (or p) valence electrons in the corresponding POTCAR valence configuration; for example, `Sc_sv` ($3s^2\,3p^6\,3d^2\,4s^1$) gives $n_{d,M} = 2$, and `Ga_d` ($3d^{10}\,4s^2\,4p^1$) gives $n_{p,X} = 1$.

Table S2. Valence d- and p-electron counts used in this work, with the PBE PAW potential (POTCAR) used for each element; $n_{d,M}$ and $n_{p,X}$ are the d and p valence-electron counts of the listed POTCAR configuration. A dash indicates an element not included in the present candidate set (Tc). (a) Magnetic species ($n_{d,M}$). (b) Non-magnetic species ($n_{p,X}$).

(a) Magnetic species

Element	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
POTCAR	`Sc_sv`	`Ti_sv`	`V_sv`	`Cr_pv`	`Mn_pv`	`Fe`	`Co`	`Ni`	`Cu`	`Zn`
$n_{d,M}$	2	3	4	5	6	7	8	9	10	10
Element	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
POTCAR	`Y_sv`	`Zr_sv`	`Nb_sv`	`Mo_sv`	–	`Ru_pv`	`Rh_pv`	`Pd`	`Ag`	`Cd`
$n_{d,M}$	2	3	4	5	–	7	8	9	10	10

(b) Non-magnetic species

Element	B	C	N	O	F
POTCAR	`B`	`C`	`N`	`O`	`F`
$n_{p,X}$	1	2	3	4	5
Element	Al	Si	P	S	Cl
POTCAR	`Al`	`Si`	`P`	`S`	`Cl`
$n_{p,X}$	1	2	3	4	5
Element	Ga	Ge	As	Se	Br
POTCAR	`Ga_d`	`Ge_d`	`As`	`Se`	`Br`
$n_{p,X}$	1	2	3	4	5
Element	In	Sn	Sb	Te	I
POTCAR	`In_d`	`Sn_d`	`Sb`	`Te`	`I`
$n_{p,X}$	1	2	3	4	5
Element	Tl	Pb	Bi		
POTCAR	`Tl_d`	`Pb_d`	`Bi_d`		
$n_{p,X}$	1	2	3		

3 Dataset Overview

The dataset comprises 3,851 binary altermagnetic entries derived from 323 parent prototypes generated using MatterGen [3] and labeled by spin-polarized DFT. The 3,851 entries comprise the 323 fully relaxed parents together with their symmetry-screened strained variants (four strain modes $\times$ four scaling factors per parent; Methods); the distributions shown below are therefore reported over the full augmented set. Each structure contains exactly two magnetic atoms per unit cell in a compensated antiparallel configuration and satisfies the three screening criteria used in the main text: near-zero total magnetization ($< 0.005\ \mu_B$), negligible mean spin splitting at Γ (< 0.01 eV), and finite non-relativistic SSE (≥ 0.001 eV).

The SSE distribution [Figure S1(a)] is heavily right-tailed, with the majority of entries below 0.75 eV; entries above 0.5 eV account for 28% of the dataset. A $\log(1+\mathrm{SSE})$ transformation (natural logarithm; `log1p`) was applied to mitigate this skew during training. Among the magnetic species [Figure S1(b)], Mn, Fe, and Cr together comprise 50% of all samples. Most compounds adopt

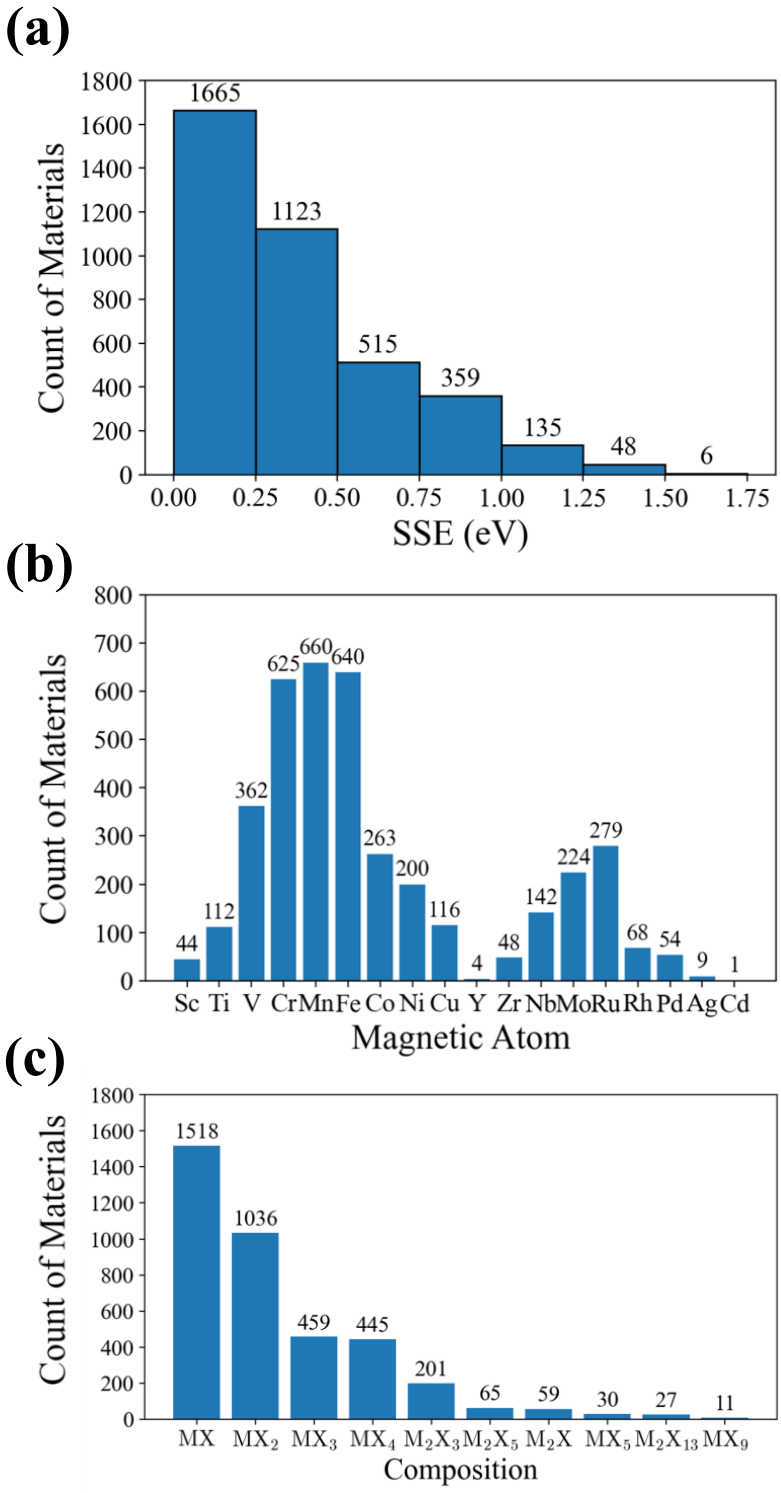

Figure S1. Statistical distribution of the dataset. (a) Histogram of SSE. (b) Distribution by magnetic atom. (c) Distribution by chemical composition.

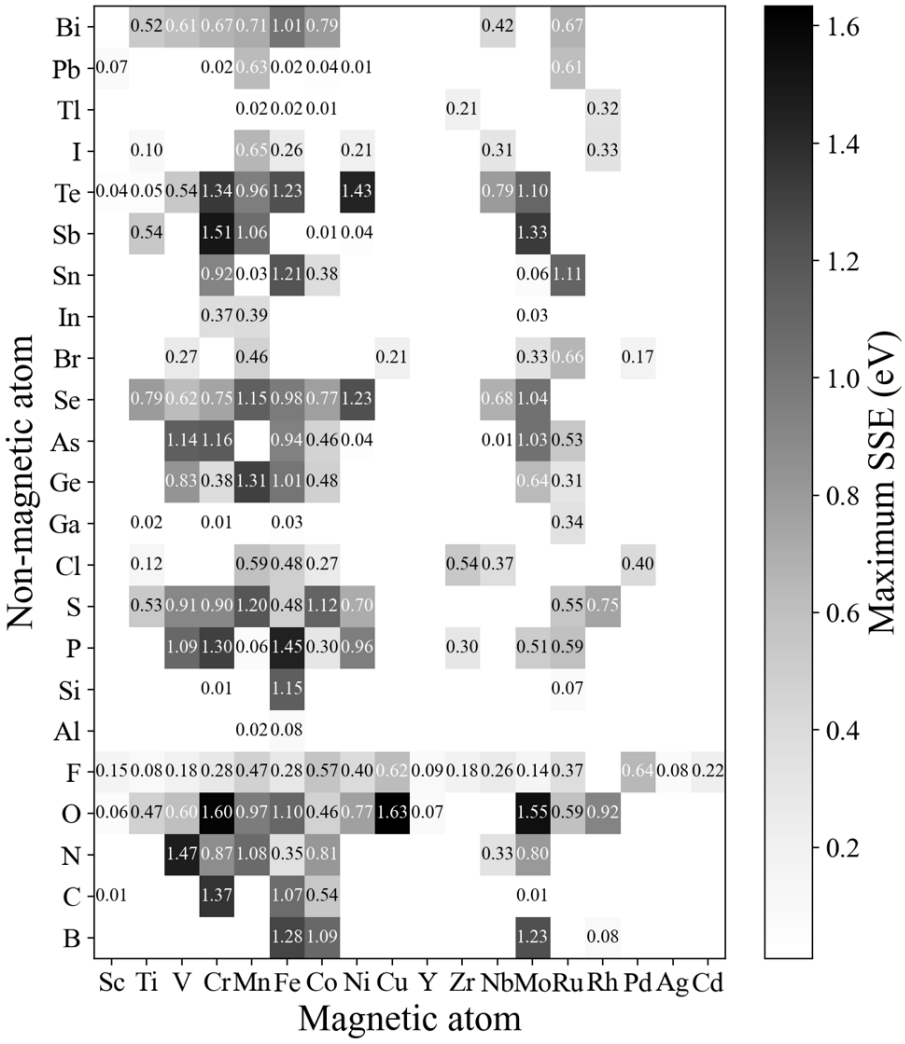

Figure S2. Heatmap of the maximum DFT-calculated SSE (eV) for each M–X elemental combination.

MX or MX_2 stoichiometries (66% of entries), with MX_3 and MX_4 contributing an additional 23% [Figure S1(c)].

The SSE heatmap (Figure S2) shows that pnictogen and chalcogen ligands (N, P, O, S, Se) consistently yield large SSE, whereas halogens (F, Cl, Br, I) produce small values, consistent with their predominantly ionic bonding character and weak p–d covalency. Several M–X combinations exceed 1 eV, notably V–N, Cr–(O, P, Sb), Fe–(B, P), Ni–Te, Cu–O, and Mo–(B, O, Sb). Overall, the dataset spans a broad range of magnetic species, ligand chemistries, stoichiometries, and SSE values, providing a chemically diverse reference set for surrogate training.

4 Machine-Learning Workflow

XGBoost (version 2.1.4) [4] was trained on `log1p`-transformed SSE using the 52-descriptor set (Table S1), with sample weights inversely proportional to parent-prototype group size to mitigate augmentation bias. Generalization was evaluated via 10-fold GroupKFold cross-validation grouped by parent prototype, ensuring that all strain-derived variants of a given parent remain within the same fold. Groups were pre-sorted by median SSE and interleaved across folds to balance the skewed target distribution.

Within each outer fold, early stopping (patience 100, maximum 5,000 rounds) used a held-out validation set drawn from an inner group-aware split. Hyperparameters (Table S3) were optimized over 200 TPE trials [5] via Optuna [6], each scored by the mean out-of-fold R^2 over three independent GroupKFold runs. Using the optimized parameters, three additional 10-fold GroupKFold runs with distinct group orderings were performed; per-sample predictions were averaged to yield a single out-of-fold estimate. The global R^2 between these averaged predictions and the true SSE values is 0.70.

The final inference model was trained on the entire dataset using the median best-iteration count from cross-validation. SHAP values [7] were averaged over 100 models independently retrained with unique random seeds; values computed in `log1p` space were rescaled to SSE units (eV) via local linearization of the inverse transform. The ranking stability of the top-three descriptors across all 100 seeds confirms that the identified hierarchy is robust against stochastic variation.

Table S3. XGBoost hyperparameter search space (Optuna TPE) and optimized values.

Hyperparameter	Search Range	Optimized Value
`learning_rate`	[0.01, 0.08]	0.0210
`max_depth`	[4, 9] (integer)	7
`min_child_weight`	[1, 7] (integer)	1
`subsample`	[0.6, 1.0]	0.7560
`colsample_bytree`	[0.6, 1.0]	0.7283
`reg_lambda` (L2)	$[10^{-2}, 10.0]$ (log)	0.4606
`reg_alpha` (L1)	$[10^{-3}, 1.0]$ (log)	0.0524

5 Surrogate Model Performance

The surrogate achieves $R^2 = 0.70$ and MAE = 123.1 meV across grouped cross-validation folds (Figure S3). Data points cluster along the 1:1 line across the full SSE range, confirming that the descriptor set encodes sufficient structural and compositional information for candidate ranking prior to DFT validation.

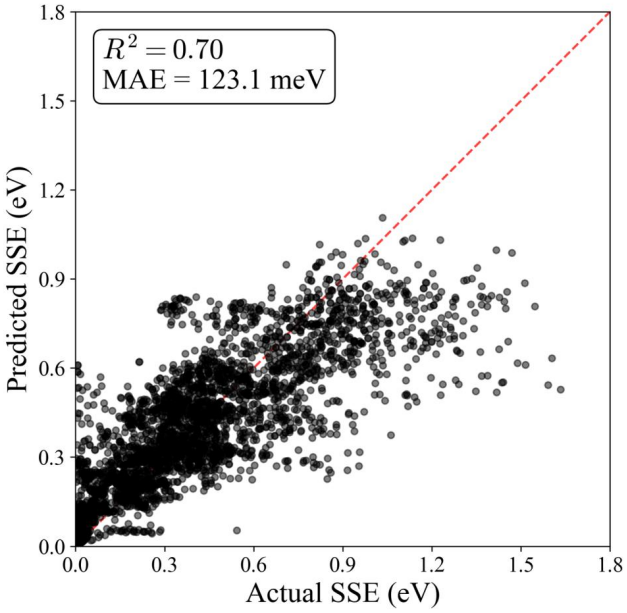

Figure S3. Parity plot of predicted versus DFT-calculated SSE under grouped cross-validation. The dashed line marks perfect agreement. Deviations are largest in the high-SSE region ($\gtrsim 1$ eV), where training data are sparse.

6 Bayesian Optimization Details

6.1 Search Space

Each BO trial constructs a candidate descriptor vector through three routes, so that the optimizer explores a chemically informed region of descriptor space rather than arbitrary feature combinations.

Composition. The atomic numbers (Z_M, Z_X) are sampled from discrete sets:

$$Z_M \in \mathcal{S}_{\text{mag}}, \qquad Z_X \in \mathcal{S}_{\text{nonmag}}, \tag{S2}$$

where $\mathcal{S}_{\text{mag}}$ spans the $3d$ transition metals and the $4d$ series excluding Tc (Sc–Zn; Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Cd), and $\mathcal{S}_{\text{nonmag}}$ spans p-block elements from B–F to Tl–Bi. Once (Z_M, Z_X) is fixed, all composition-derived descriptors are evaluated deterministically:

$$\Delta\chi = \chi_M - \chi_X, \qquad \Delta Z = Z_M - Z_X, \qquad p/d = \frac{n_{p,X}}{n_{d,M}}, \tag{S3}$$

together with $|\Delta\chi|$, $|\Delta Z|$, n_{unpaired}, and μ_{spin}.

Local motif geometry. The coordination number $N_X \in \{4, 6\}$ is treated as a discrete decision variable. For $N_X = 4$, the ligand cage is initialized as either tetrahedral or square-planar, whereas for $N_X = 6$ an octahedral template is used. The BO sampler then perturbs this template through spherical-coordinate parameters that control radial distances, polar angles, azimuthal angles, and symmetry-breaking distortions such as stretches, tilts, and jitter. These operations produce a concrete set of ligand coordinates relative to the magnetic center, from which all intra-motif descriptors—including bond statistics, angle statistics, V_{motif}, η_{shape}, σ_{shape}, δ_{off}, MX_{elong}, D_{motif},

and $\Gamma_{\chi\text{-elong}}$—are computed algorithmically rather than sampled. The virtual motifs at this stage do not carry a resolved planar/polyhedral classification, so the packing fraction MPF is evaluated for all coordination classes using the volume-based proxy $V_{\text{motif}}/V_{\text{cell}}$; the area-based definition for planar motifs is applied only at the labeled-dataset and DFT-validation stages (Methods). As a consequence, BO-stage and DFT-stage MPF values are not directly comparable for square-planar candidates (see Section 7).

Global packing. A small number of coarse-grained parameters are used to approximate inter-motif coupling without generating a full periodic crystal during optimization. The unit-cell volume V_{cell} is sampled within the empirical range of the reference dataset, which in turn determines MPF through $V_{\text{motif}}/V_{\text{cell}}$. Inter-motif centroid distances $d_{CC}^{(k)}$, magnetic-neighbor distances $d_{MM}^{(k)}$, and the associated sublattice displacements $\Delta_{\text{sub}}^{(k)} = d_{MM}^{(k)} - d_{CC}^{(k)}$ for $k = 1, 2, 3$ are also sampled within dataset bounds, as is the alignment angle θ_{align}. MSBI and σ_{inhom}, which are inter-motif quantities that cannot be derived from a single virtual motif, are treated in the same way as bounded scalar proxies. The BO stage therefore explores a bounded descriptor manifold rather than a one-to-one realizable crystal manifold; physical realizability is imposed only after prototype matching and independent DFT relaxation.

6.2 Convergence

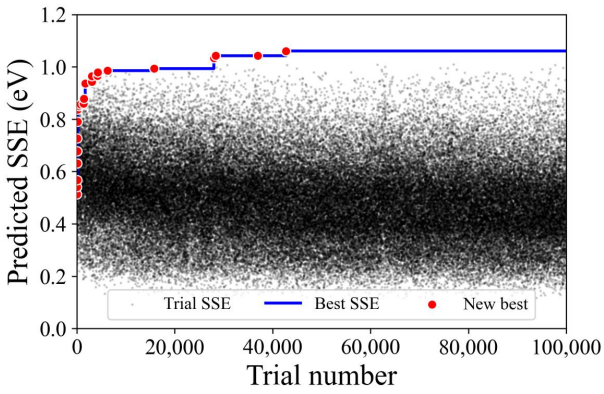

Figure S4. Convergence of the Bayesian optimization. Gray dots show the surrogate-predicted SSE for each of the 10^5 TPE trials; the blue curve tracks the cumulative best SSE, and red circles mark the trials at which a new optimum was found. The optimizer reaches $\sim 95\%$ of its final cumulative best value within the first $\sim$ 5,000 trials and plateaus near 1.06 eV after approximately 50,000 trials.

The cumulative best SSE rises rapidly during the initial exploration phase and plateaus near 1.06 eV after roughly 50,000 trials (Figure S4). The continued scatter of individual trial SSE values across the full 0–1.1 eV range reflects the intended exploration–exploitation balance of TPE and suggests that the observed optimum is not simply a local artifact of early sampling.

6.3 Prototype Matching Methodology

To convert each BO-optimized descriptor vector into a periodic crystal structure amenable to DFT validation, it is matched to the nearest prototype in the reference dataset. Let $\mathbf{f}_j^{\text{opt}}$ and $\mathbf{F}_i$ denote

the full feature vectors of the j-th optimized candidate and the i-th reference structure, respectively. Matching is restricted to the structural partition of the feature vector,

$$\mathbf{f}_j^{\mathrm{opt}} = \left(\mathbf{f}_j^{\mathrm{struc}}, \mathbf{f}_j^{\mathrm{chem}}\right), \qquad \mathbf{F}_i = \left(\mathbf{F}_i^{\mathrm{struc}}, \mathbf{F}_i^{\mathrm{chem}}\right), \tag{S4}$$

where $\mathbf{f}^{\mathrm{struc}}$ contains geometric and symmetry descriptors and $\mathbf{f}^{\mathrm{chem}}$ contains chemical and electronic descriptors. Excluding the chemical partition ensures that the similarity metric reflects structural resemblance independent of composition, while the BO procedure itself already determines the substituted chemical identities.

Each structural descriptor k is z-score-normalized using the dataset mean μ_k and standard deviation σ_k:

$$\tilde{f}_{j,k}^{\mathrm{struc}} = \frac{f_{j,k}^{\mathrm{struc}} - \mu_k}{\sigma_k}, \qquad \tilde{F}_{i,k}^{\mathrm{struc}} = \frac{F_{i,k}^{\mathrm{struc}} - \mu_k}{\sigma_k}. \tag{S5}$$

Two complementary metrics are then computed in this normalized space:

$$c_{ij} = \frac{\tilde{\mathbf{f}}_j^{\mathrm{struc}} \cdot \tilde{\mathbf{F}}_i^{\mathrm{struc}}}{\|\tilde{\mathbf{f}}_j^{\mathrm{struc}}\| \, \|\tilde{\mathbf{F}}_i^{\mathrm{struc}}\|}, \qquad d_{ij} = \left\| \tilde{\mathbf{f}}_j^{\mathrm{struc}} - \tilde{\mathbf{F}}_i^{\mathrm{struc}} \right\|. \tag{S6}$$

The cosine similarity c_{ij} measures alignment of the structural-descriptor pattern, whereas the Euclidean distance d_{ij} penalizes deviations in absolute feature magnitudes. The nearest prototype for each optimized candidate j is selected by lexicographic ranking,

$$i_j^* = \arg\max_i \left(c_{ij}, -d_{ij}\right), \tag{S7}$$

which first maximizes cosine similarity and then uses Euclidean distance only as a tiebreaker. After prototype selection, the BO-identified species are substituted onto the prototype sites, and the resulting periodic structure is independently relaxed with spin-polarized DFT for validation.

Because cosine similarity is the primary criterion and Euclidean distance enters only to break ties, a matched prototype can exhibit high cosine similarity yet a large absolute distance. This is precisely the situation for the octahedral candidates reported in the main text, whose distances ($d \approx 21$–24) substantially exceed the median pairwise Euclidean distance among the 323 parent prototypes ($d \approx 7.8$, computed as the median over all parent–parent pairs in the normalized structural-descriptor space). The optimizer drives the inter-motif descriptors—most notably MSBI—toward the upper edge of their sampled ranges (MSBI ≈ 1.42 for all three octahedral candidates), so the matched prototype shares the same structural pattern ($c \approx 0.89$–0.91) while differing in absolute magnitude. The large d therefore signals descriptor-magnitude extrapolation by the optimizer, not a poor archetype match; the subsequent DFT relaxation relaxes these descriptors back into the favorable regime (e.g., MSBI $1.42 \rightarrow 0.68$; Table S4), where the quantitative validation is obtained.

6.4 Coordination-Class Coverage

Of the 323 parent prototypes in the dataset, 7 (2.2%) exhibit square-planar ($N_X = 4$) coordination and 204 (63%) are octahedral ($N_X = 6$); the remaining prototypes adopt other coordination environments (e.g., tetrahedral and lower-symmetry polyhedra) that are present in the training set but lie outside the $N_X \in \{4, 6\}$ classes targeted by the Bayesian-optimization search. The square-planar/octahedral imbalance is reflected in the prototype-matching quality of the BO candidates reported in the main text: the four-coordinate candidates achieve cosine similarities of $c \approx 0.55$–0.69, substantially lower than the $c \approx 0.89$–0.91 obtained for the six-coordinate candidates. The lower c

values do not imply poor BO performance, but rather the limited structural diversity available within the square-planar subset of the training data. Expanding the square-planar subset, for example by targeted generation of CuO-type and PtS-type structures, would likely improve matching fidelity and is therefore a natural direction for future dataset augmentation.

7 DFT Validation of Remaining BO Candidates

Figures S5 and S6 present the DFT validation results for the BO-selected candidates not shown in the main text. Table S4 compares the BO-optimized and DFT-relaxed descriptor values for all magnetically stable candidates. The descriptor values drift during relaxation, most notably MSBI for the octahedral candidates, which decreases by roughly a factor of two (e.g., from 1.42 to 0.68); this reflects the fact that the Bayesian-optimization stage samples the inter-motif descriptors (MSBI, σ_{inhom}, and the packing proxies) within dataset bounds rather than constructing them from a fully realized crystal. For the square-planar candidates, the apparent increase in MPF (e.g., from 0.22 to 0.53) is largely definitional rather than structural: the Bayesian-optimization stage evaluates MPF using the volume-based proxy $V_{\mathrm{motif}}/V_{\mathrm{cell}}$ for all candidates, whereas the DFT-relaxed value uses the area-based definition for planar motifs (Methods). Because the area- and volume-based MPF are not directly comparable in magnitude, the BO and DFT MPF columns for the square-planar candidates should be read within their respective stages rather than as a relaxation-induced change; the favorable-regime assessment (MPF above its empirical threshold) is therefore made against the appropriate per-coordination definition. The optimizer thus selects a promising chemistry, coordination class, and descriptor regime, while the independent DFT relaxation—which preserves the favorable regime (MSBI > 0.4, MPF above threshold, $p/d < 1$) for all six candidates—provides the quantitative validation.

Table S4. Descriptor drift from BO optimization to DFT relaxation. BO columns report the surrogate-optimized descriptor values; DFT columns report the values recomputed from independently relaxed structures. For square-planar ($N_X = 4$) candidates, the BO-stage MPF is volume-based and the DFT-stage MPF is area-based (Methods); the two are not directly comparable, so the MPF columns for these rows are interpreted within each stage rather than as a relaxation-induced change. All candidates remain within the favorable descriptor regime (MSBI > 0.4, MPF above its empirical threshold, $p/d < 1$) after full structural relaxation.

		MSBI		MPF		p/d		SSE (eV)	
N_X	Composition	BO	DFT	BO	DFT	BO	DFT	Pred.	DFT
4	Fe–S	1.044	0.871	0.222	0.533	0.571	0.571	1.061	1.297
4	Co–O	1.404	0.812	0.273	0.521	0.500	0.500	1.033	0.972
4	Cr–S	0.661	0.668	0.255	0.528	0.800	0.800	1.009	0.552
6	Ni–S	1.421	0.681	0.427	0.333	0.444	0.444	0.893	0.823
6	Co–S	1.422	0.689	0.321	0.333	0.500	0.500	0.874	1.103
6	Fe–As	1.424	0.690	0.295	0.333	0.429	0.429	0.867	1.089

7.1 Square-Planar Candidates ($N_X = 4$)

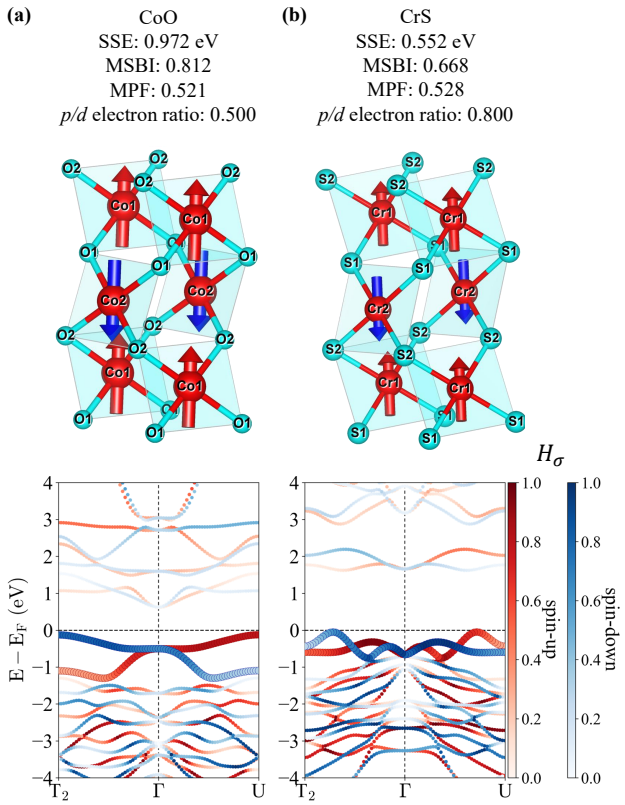

Figure S5. DFT validation of square-planar BO candidates. (a) CoO (SSE = 0.972 eV) and (b) CrS (SSE = 0.552 eV). Top panels show the relaxed crystal structures with spin order, and bottom panels show spin-resolved p–d hybridization fatbands. Band opacity scales with H_σ; the band pair used for SSE evaluation is highlighted.

CoO [Figure S5(a)] yields SSE = 0.972 eV, in close agreement with the surrogate prediction of 1.033 eV. The fatbands display strong spin-asymmetric p–d hybridization near E_F, consistent with the covalency-driven mechanism identified in the main text. CrS [Figure S5(b)] yields SSE = 0.552 eV, substantially below the surrogate prediction of 1.009 eV. The fatbands display weaker spin-asymmetric hybridization compared with CoO, consistent with the lower SSE; this, combined with the limited structural diversity of square-planar prototypes in the training set, likely accounts for the discrepancy. As with square-planar FeS in the main text, these square-planar CoO and CrS candidates are best read as motif-level predictions for the square-planar coordination environment rather than as asserted ground-state polymorphs of the respective compositions.

7.2 Octahedral Candidates ($N_X = 6$)

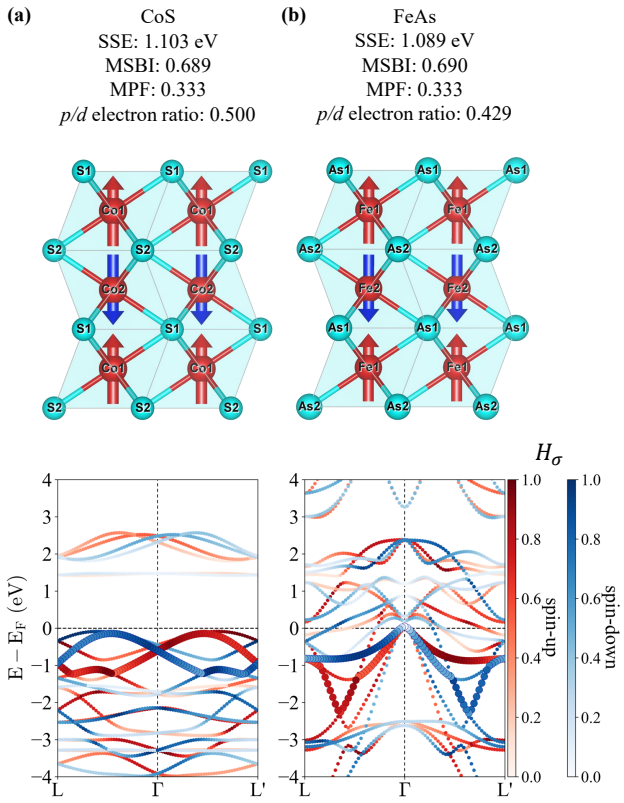

Figure S6. DFT validation of octahedral BO candidates. (a) CoS (SSE = 1.103 eV) and (b) FeAs (SSE = 1.089 eV). Top panels show the relaxed crystal structures with spin order, and bottom panels show spin-resolved p–d hybridization fatbands. Band opacity scales with H_σ; the band pair used for SSE evaluation is highlighted.

CoS [Figure S6(a)] and FeAs [Figure S6(b)] both exceed their surrogate predictions (0.874 and 0.867 eV, respectively), yielding DFT-validated SSE of 1.103 and 1.089 eV. Both structures exhibit the hallmark spin-asymmetric p–d hybridization pattern near E_F, confirming that the covalency-driven amplification mechanism generalizes across NiAs-type compositions. NiAs-type CoS has been independently reported as a g-wave altermagnet from correlated first-principles calculations [8]. That study evaluates the spin splitting along off-nodal high-symmetry lines rather than over the full Brillouin-zone mesh used here, so it corroborates the g-wave symmetry assignment while its reported magnitude is not directly comparable to our full-mesh SSE of 1.103 eV. As with the square-planar candidates above, the FeAs result is best read as a motif-level prediction for the NiAs-type Fe–As environment rather than an assertion about the bulk phase. The conservative surrogate behavior in this coordination class—where DFT values consistently meet or exceed predictions—is consistent with the high prototype density and matching fidelity ($c \approx 0.89$–0.91) available for octahedral motifs.

8 Synthesis Routes for NiAs-type α-NiS

Table S5 summarizes established synthesis routes for NiAs-type α-NiS, the experimentally accessible candidate identified in this work. These reports establish access to the NiAs polymorph, while an

independent first-principles and tight-binding chemical-bonding analysis has established α-NiS as an altermagnet [9]; the g-wave designation follows from its magnetic point group. No epitaxial thin-film growth has been reported to date; however, epitaxial films of isostructural CrSb (NiAs-type) have recently been realized by both MBE and sputtering with ARPES-confirmed altermagnetic band splitting [10], suggesting that a similar approach is feasible for α-NiS.

Table S5. Established synthesis routes for NiAs-type α-NiS.

Method	Form	Key conditions	Ref.
Chemical-vapor transport	Single crystal	I_2 transport, 700–800 °C	[11]
Solvothermal + annealing	Hierarchical particles	$Ni(NO_3)_2$ + thiourea, 200 °C 9 h; Ar anneal 400 °C	[12]
Confined sulfidation	NPs in carbon rods	Ni/C precursor + thiourea in glycerol, 280 °C 1 h	[13]
Solvothermal	Spheres and wires	$NiSO_4$ + thiourea in EG/H_2O(2:1), 180 °C 12 h	[14]
Spray pyrolysis	Thin film	$NiCl_2$ + thiourea, 250 °C, glass	[15]

References

[1] Kuhn, H. W. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2**, 83–97 (1955).

[2] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Foundations of Crystallography* **32**, 922–923 (1976).

[3] Zeni, C. *et al.* A generative model for inorganic materials design. *Nature* **639**, 624–632 (2025).

[4] Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).

[5] Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems* **24** (2011).

[6] Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (2019).

[7] Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **30** (2017).

[8] Davoudi Tanha, S., Modarresi, M., Roknabadi, M., Hu, T. & Mogulkoc, A. Altermagnetic phase in cobalt sulfide: Effects of local correlation and crystal structure. *Physical Review Materials* **9**, 074406 (2025).

[9] Mandal, A., Das, A. & Nanda, B. Deterministic role of chemical bonding in the formation of altermagnetism: Reflection from the correlated electron system nis. *Physical Review B* **112**, 014420 (2025).

[10] Reimers, S. *et al.* Direct observation of altermagnetic band splitting in crsb thin films. *Nature Communications* **15**, 2116 (2024).

[11] Gore, T. E. Growth of single crystals of crowningshieldite (α-nis) by chemical-vapour transport. *Journal of Crystal Growth* **648**, 127880 (2024).

[12] Tan, Y. *et al.* Solvothermal synthesis of hierarchical α-nis particles as battery-type electrode materials for hybrid supercapacitors. *Journal of Alloys and Compounds* **806**, 1068–1076 (2019).

[13] Sun, C. *et al.* Phase-controlled synthesis of α-nis nanoparticles confined in carbon nanorods for high performance supercapacitors. *Scientific Reports* **4**, 7054 (2014).

[14] Liu, S., Shi, Q., Tong, J., Li, S. & Li, M. Controlled synthesis of spherical α-nis and urchin-like β-nis microstructures. *Journal of Experimental Nanoscience* **9**, 475–481 (2014).

[15] Boughalmi, R. *et al.* Metallic behavior of nis thin film under the structural, optical, electrical and ab initio investigation frameworks. *Materials Chemistry and Physics* **163**, 99–106 (2015).