Scalar-pathway fidelity improves physical accuracy in short-range equivariant interatomic potentials — Supplementary Information

Jia Bi[1], Alin Marin Elena[2*], Samuel Pinilla[1,3*]

[1]Science and Technology Facilities Council, Harwell Campus, Didcot, OX11 0QX, Oxfordshire, United Kingdom.
[2]Science and Technology Facilities Council, Keckwick Lane, Daresbury, WA4 4AD, Cheshire, United Kingdom.
[3]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0QX, Oxfordshire, United Kingdom.

*Corresponding author(s). E-mail(s): alin-marin.elena@stfc.ac.uk; samuel.pinilla@diamond.ac.uk;
Contributing authors: Jia.Bi@stfc.ac.uk;

Contents

S1 Per-molecule molecular benchmark results

This section reports the full per-molecule energy and force MAE values that underlie the molecular benchmark summary in Section 2.2 of the main text. Two complementary tables are provided. Table S1 contains controlled in-house comparisons among MACE-family variants under matched hyperparameters and training budgets. Table S2 contains the external literature reference values for context only. As discussed in the main text, the two tables follow different protocols and should not be interpreted as directly comparable: literature numbers are reproduced from the original publications and follow their respective training, evaluation, and reporting conventions.

The within-backbone trends are consistent across both benchmarks: PAN and PGS each reduce force MAE relative to baseline MACE for every flexible or moderately anharmonic molecule, with the largest absolute force-error reductions observed for aspirin, malonaldehyde, salicylic acid, and paracetamol. Near-harmonic molecules (benzene and toluene) show small or negligible changes, consistent with the scalar pathway already being sufficiently expressive for those landscapes. The combined PAN+PGS variant gives the lowest force MAE across all ten molecules in both rMD17 and MD17, although for toluene at $N_{\text{train}} = 950$ the margin over MACE+PGS is within the reported seed-level variability.

The literature reference column places our MACE+PAN+PGS results in the external context. While the in-house gain over baseline MACE serves as the controlled comparison and the basis for the conclusions in the main text, our MACE+PAN+PGS values also remain competitive with the strongest equivariant baselines reported in the literature.

S2 Full stress benchmark

Section 2.2 of the main text reports MACE+PAN+PGS stress accuracy on the two systems with available stress labels (Ag and the LiF/Li–F short-range ionic subset). Here we provide the full breakdown across the four MACE-family variants. Stress accuracy is not reported for Si because the public `Silicon_MLIP_datasets` release used in this work does not provide stress labels; the stress weight in the training loss was therefore set to zero for Si.

The relative reduction of MACE+PAN+PGS over baseline MACE is 27% on Ag and 28% on the LiF/Li–F subset, tracking the corresponding force-error reductions reported in Table 1 of the main text. The same relative ordering of variants seen in energies and forces persists for stress.

S3 Short-range physics: distortion descriptor and ablations

This section provides definitions and extended ablation studies that support the short-range and mechanistic analyses in the main text (Section 2.3).

Table S1 Per-molecule rMD17 and MD17 results for MACE-family variants. Energy (E) and force (F) MAE are reported for rMD17 ($N_{\text{train}} = 1{,}000$) and MD17 ($N_{\text{train}} = 950$) under matched MACE backbone, hyperparameters, training budget, and data splits; only the scalar-pathway modules differ. Values are mean ± s.d. over five seeds. Energies are in meV, and forces are in meV/Å. Bold marks the uniquely best in-house result; ties are unbolded.

(a) rMD17 ($N_{\text{train}} = 1{,}000$)

Molecule	Metric	MACE	MACE+PAN	MACE+PGS	MACE+PAN+PGS
Aspirin	E	2.21 ± 0.09	2.11 ± 0.08	2.02 ± 0.07	**1.91 ± 0.06**
	F	6.62 ± 0.20	6.31 ± 0.18	5.92 ± 0.16	**5.59 ± 0.14**
Azobenzene	E	1.21 ± 0.05	1.13 ± 0.05	1.10 ± 0.05	**1.02 ± 0.04**
	F	3.02 ± 0.10	2.81 ± 0.10	2.72 ± 0.09	**2.51 ± 0.08**
Benzene	E	0.41 ± 0.02	0.41 ± 0.02	0.40 ± 0.02	0.40 ± 0.02
	F	0.31 ± 0.02	0.31 ± 0.02	0.30 ± 0.02	0.30 ± 0.02
Ethanol	E	0.41 ± 0.02	0.40 ± 0.02	0.39 ± 0.02	**0.38 ± 0.02**
	F	2.11 ± 0.07	2.06 ± 0.07	1.92 ± 0.06	**1.88 ± 0.06**
Malonaldehyde	E	0.81 ± 0.04	0.79 ± 0.04	0.72 ± 0.03	**0.69 ± 0.03**
	F	4.11 ± 0.13	3.92 ± 0.12	3.61 ± 0.11	**3.41 ± 0.10**
Naphthalene	E	0.51 ± 0.03	0.49 ± 0.03	0.46 ± 0.02	**0.43 ± 0.02**
	F	1.62 ± 0.06	1.55 ± 0.06	1.41 ± 0.05	**1.32 ± 0.05**
Paracetamol	E	1.32 ± 0.06	1.22 ± 0.05	1.15 ± 0.05	**1.08 ± 0.04**
	F	4.81 ± 0.15	4.61 ± 0.14	4.32 ± 0.13	**4.11 ± 0.12**
Salicylic acid	E	0.91 ± 0.04	0.85 ± 0.04	0.81 ± 0.04	**0.78 ± 0.03**
	F	3.12 ± 0.10	2.98 ± 0.10	2.81 ± 0.09	**2.62 ± 0.08**
Toluene	E	0.51 ± 0.03	0.48 ± 0.02	0.46 ± 0.02	**0.44 ± 0.02**
	F	1.51 ± 0.05	1.48 ± 0.05	1.41 ± 0.05	**1.36 ± 0.04**
Uracil	E	0.51 ± 0.03	0.48 ± 0.02	0.46 ± 0.02	**0.42 ± 0.02**
	F	2.12 ± 0.07	2.01 ± 0.07	1.88 ± 0.06	**1.78 ± 0.06**

(b) MD17 ($N_{\text{train}} = 950$, original *ab initio* labels)

Molecule	Metric	MACE	MACE+PAN	MACE+PGS	MACE+PAN+PGS
Aspirin	E	17.0 ± 0.5	16.5 ± 0.5	15.8 ± 0.5	**15.2 ± 0.4**
	F	43.9 ± 1.2	42.0 ± 1.1	40.5 ± 1.0	**38.8 ± 1.0**
Azobenzene	E	5.41 ± 0.18	5.02 ± 0.16	5.21 ± 0.17	**4.81 ± 0.15**
	F	17.7 ± 0.5	16.2 ± 0.5	17.0 ± 0.5	**15.8 ± 0.5**
Benzene	E	0.71 ± 0.03	0.71 ± 0.03	0.70 ± 0.03	0.70 ± 0.03
	F	2.72 ± 0.06	2.71 ± 0.06	2.70 ± 0.06	**2.61 ± 0.05**
Ethanol	E	6.71 ± 0.22	6.71 ± 0.22	6.21 ± 0.20	**6.11 ± 0.20**
	F	32.6 ± 0.9	32.1 ± 0.9	30.5 ± 0.8	**30.2 ± 0.8**
Malonaldehyde	E	10.0 ± 0.3	9.80 ± 0.32	9.01 ± 0.30	**8.81 ± 0.29**
	F	43.3 ± 1.2	42.5 ± 1.1	39.5 ± 1.0	**38.9 ± 1.0**
Naphthalene	E	2.11 ± 0.08	2.10 ± 0.08	2.02 ± 0.07	**2.01 ± 0.07**
	F	9.21 ± 0.27	9.11 ± 0.27	9.02 ± 0.27	**8.91 ± 0.26**
Paracetamol	E	9.71 ± 0.30	9.21 ± 0.29	9.02 ± 0.28	**8.61 ± 0.27**
	F	31.5 ± 0.9	30.2 ± 0.8	29.5 ± 0.8	**28.5 ± 0.7**
Salicylic acid	E	6.51 ± 0.21	6.31 ± 0.20	5.92 ± 0.19	**5.81 ± 0.18**
	F	28.4 ± 0.8	27.8 ± 0.7	26.5 ± 0.7	**26.1 ± 0.7**
Toluene	E	3.11 ± 0.11	3.10 ± 0.11	3.02 ± 0.10	**3.01 ± 0.10**
	F	12.1 ± 0.4	12.0 ± 0.4	11.8 ± 0.4	**11.7 ± 0.4**
Uracil	E	4.41 ± 0.14	4.32 ± 0.14	4.21 ± 0.13	**4.12 ± 0.13**
	F	25.9 ± 0.7	25.2 ± 0.7	24.8 ± 0.7	**24.3 ± 0.6**

Table S2 External literature reference values for rMD17 (force MAE, meV/Å). Values are reproduced from the original publications: MACE [1], Allegro [2], BOTNet [3], NequIP [4], GemNet [5], ACE [6], FCHL [7], GAP [8], ANI [9], PaiNN [10], DimeNet [11] and NewtonNet [12]. These literature baselines follow different training, evaluation, and reporting protocols and are included only as external reference points; they should not be interpreted as controlled like-for-like comparisons with the in-house MACE-family variants reported in the main text and in Table S1. Entries marked "–" were not reported in the original publication. The right-most column reports our MACE+PAN+PGS result from Table S1(a) for direct visual reference.

Molecule	MACE	Allegro	BOT Net	NequIP	Gem Net	ACE	FCHL	GAP	ANI	PaiNN	Dime Net	Newton Net	MACE+PAN+ PGS (ours)
Aspirin	6.6	7.3	8.5	8.2	9.5	17.9	20.9	44.9	40.6	16.1	21.6	15.1	**5.59**
Azobenzene	3.0	2.6	3.3	2.9	–	10.9	10.8	24.5	35.4	–	–	5.9	**2.51**
Benzene	0.3	0.2	0.3	0.3	0.5	0.5	2.6	6.0	10.0	–	8.1	–	0.30
Ethanol	2.1	2.1	3.2	2.8	3.6	7.3	6.2	18.1	13.4	10.0	10.0	9.1	**1.88**
Malonaldehyde	4.1	3.6	5.8	5.1	6.6	11.1	10.3	26.4	24.5	13.8	16.6	14.0	**3.41**
Naphthalene	1.6	0.9	1.8	1.3	1.9	5.1	6.5	16.5	29.2	3.6	9.3	3.6	**1.32**
Paracetamol	4.8	4.9	5.8	5.9	–	12.7	12.3	28.9	30.4	–	–	11.4	**4.11**
Salicylic acid	3.1	2.9	4.3	4.0	5.3	9.3	9.5	24.7	29.7	9.1	16.2	8.5	**2.62**
Toluene	1.5	1.8	1.9	1.6	2.2	6.5	8.8	17.8	24.3	4.4	9.4	3.8	1.36
Uracil	2.1	1.8	3.2	3.1	3.8	6.6	4.2	17.6	21.4	6.1	13.1	6.4	**1.78**

Table S3 Full stress MAE breakdown for MACE-family variants. Stress MAE in meV/Å^3 on the test split for Ag (author-generated) and the short-range ionic LiF/Li–F subset (MPtrj-derived). The reported stress uses the same volume normalization and sign convention as the DFT stress labels used in training and evaluation. All entries are mean ± standard deviation over five training seeds. PAN alone reduces stress error in heterogeneous configurations, PGS alone reduces error predominantly in the short-range region, and the combined PAN+PGS variant captures both effects.

System	**MACE**	**MACE+PAN**	**MACE+PGS**	**MACE+PAN+PGS**
Ag	11.8 ± 0.6	10.6 ± 0.5	9.7 ± 0.5	$\mathbf{8.6 \pm 0.4}$
LiF/Li–F	17.4 ± 0.9	15.8 ± 0.8	14.2 ± 0.7	$\mathbf{12.6 \pm 0.6}$

S3.1 Local distortion descriptor

The continuous distortion descriptor used in Fig. 3a of the main text is defined at the atom level. For atom i, the distortion magnitude is

$$D_i = \alpha\, \Delta N_i + \beta\, \sigma_{c,i} + \gamma\, \delta\rho_i, \tag{S1}$$

where ΔN_i is the coordination deficit relative to a bulk reference, $\sigma_{c,i}$ is the cutoff-weighted variance of nearest-neighbor cosine angles $c_{jik} = \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ik}$, and $\delta\rho_i$ is the deviation of the local atomic density from the bulk average. The three components are individually rescaled to $[0, 1]$, combined with weights $(\alpha, \beta, \gamma) = (1/3, 1/3, 1/3)$ to give equal emphasis, and the combined descriptor D_i is then linearly rescaled to $[0, 1]$ across the dataset. This yields a monotonic ranking of atomic environments from bulk-like ($D_i \to 0$) to highly distorted ($D_i \to 1$) regimes: surface, defect-adjacent, and strongly strained.

For the binned analysis in Fig. 3, we partitioned the test atoms into five equal-population quantiles in D_i and computed the force MAE separately within each quantile.

Table S4 Class-resolved energy and force RMSE for Ag. Energy RMSE in meV/atom and force RMSE in meV/Å for bulk fcc, distorted (strain), defect (vacancy and divacancy), and surface ((100), (110), (111)) environments. The same Ag model is evaluated after binning the test environments by structural class. Entries are mean ± standard deviation over five training seeds. The right-most column reports the relative force-RMSE improvement of MACE+PAN over baseline MACE.

Structure class	**MACE** (E / F)	**MACE+PAN** (E / F)	ΔF / F_{MACE}
Bulk fcc	1.20 ± 0.06 / 22.5 ± 0.6	1.10 ± 0.05 / 20.3 ± 0.5	9.8%
Distorted (strain)	2.62 ± 0.13 / 37.3 ± 1.1	2.10 ± 0.10 / 29.7 ± 0.9	20.4%
Defect (vacancy)	5.05 ± 0.22 / 58.7 ± 1.7	3.51 ± 0.16 / 41.5 ± 1.2	29.3%
Surface	7.05 ± 0.30 / 74.5 ± 2.0	4.22 ± 0.18 / 46.9 ± 1.4	37.0%

Table S5 Ag structural-class held-out evaluation. Energy and force MAE for Ag models evaluated on structural classes not included in training. In the defect-held-out split, vacancy and divacancy configurations were held out. In the surface-held-out split, (100), (110), (111) surface configurations were held out. Values are the mean ± standard deviation over five training seeds. Energies are in meV/atom, and forces are in meV/Å. The right-most column reports the relative force-MAE reduction of MACE+PAN+PGS over baseline MACE on the held-out class.

Held-out class	**Quantity**	**MACE**	**MACE+PAN**	**MACE+PGS**	**MACE+PAN+PGS**	**Reduction**
Defect	Energy	5.82 ± 0.25	4.52 ± 0.20	4.78 ± 0.21	$\mathbf{3.91 \pm 0.17}$	32.8%
	Force	66.8 ± 2.0	49.8 ± 1.5	52.6 ± 1.6	$\mathbf{43.9 \pm 1.3}$	34.3%
Surface	Energy	7.64 ± 0.34	5.68 ± 0.25	6.02 ± 0.27	$\mathbf{4.86 \pm 0.22}$	36.4%
	Force	84.6 ± 2.7	61.9 ± 2.0	65.4 ± 2.1	$\mathbf{55.2 \pm 1.7}$	34.8%

The class-resolved breakdown in Table S4 corroborates the continuous trend in the main text. Improvements are smallest in bulk fcc configurations and increase monotonically with structural distortion, reaching their largest values at surface atoms.

S3.2 Ag structural-class held-out evaluation

To test whether the scalar-pathway correction remains effective under a stronger environment shift than a random configuration-level split, we performed two Ag structural-class held-out evaluations. In the surface-held-out split, all surface configurations (100), (110), (111) were removed from the training and validation sets and used only for testing. In the defect-held-out split, vacancy and divacancy configurations were removed from the training and validation sets and used only for testing. Bulk, strained, and high-temperature configurations were retained in the training pool. All models used the same MACE backbone, optimizer, loss weights, cutoff, training budget, and five-seed protocol as in the main Ag experiments; only the scalar-pathway modules differed. This evaluation complements the class-resolved breakdown reported in Table S4, which evaluates class-resolved performance under the main random split, by additionally probing class-level out-of-distribution generalization.

Table S6 Scalar-mixer ablation under matched PAN pooling. Energy and force MAE for different invariant-channel mixers on Ag, Si, and the LiF/Li–F short-range ionic subset. All entries share the same MACE backbone and PAN pooling, with only the scalar mixer varied; parameter and FLOPs budgets are matched within ±5%. Energies are in meV/atom, forces in meV/Å. Values are the mean ± standard deviation over five training seeds. The PGS column reproduces the MACE+PAN+PGS entries of Table 1 of the main text.

System	Quantity	MLP	KAN	KAF	LibraKAN	PGS (ours)
Ag (metallic)	Energy	1.32 ± 0.07	1.30 ± 0.06	1.25 ± 0.06	1.18 ± 0.06	$\mathbf{1.12 \pm 0.05}$
	Forces	32.5 ± 0.8	31.9 ± 0.8	30.7 ± 0.7	29.4 ± 0.7	$\mathbf{27.9 \pm 0.7}$
Si (covalent)	Energy	2.30 ± 0.10	2.25 ± 0.09	2.18 ± 0.09	2.05 ± 0.08	$\mathbf{1.92 \pm 0.08}$
	Forces	41.3 ± 1.1	40.5 ± 1.1	39.0 ± 1.0	37.5 ± 1.0	$\mathbf{35.8 \pm 1.0}$
LiF/Li–F	Energy	2.95 ± 0.12	2.85 ± 0.11	2.72 ± 0.11	2.55 ± 0.10	$\mathbf{2.41 \pm 0.10}$
	Forces	47.5 ± 1.4	46.2 ± 1.3	44.0 ± 1.3	41.8 ± 1.2	$\mathbf{39.6 \pm 1.2}$

Absolute errors increase on both held-out splits relative to the main random-split Ag results (Table 1), as expected for a harder evaluation in which the model has not seen the corresponding structural class during training. The relative MACE+PAN+PGS reduction over baseline MACE, however, is preserved and is in fact larger than on the random split (34–35% versus 22.3%). This is consistent with the PAN mechanism: ρ_i and η_i encode coordination loss and angular distortion, precisely the geometric perturbations that distinguish surface and defect environments from bulk fcc. The result supports the view that the scalar-pathway correction is not a configuration-level interpolation artifact and is most pronounced in the under-coordinated, defect-adjacent regimes that motivate PAN.

S3.3 Scalar mixers under matched PAN pooling

To test whether spectral enrichment via PGS is necessary or whether any sufficiently expressive scalar mixer is enough, we replaced the scalar mixer in the MACE+PAN architecture with a series of alternative mixers under matched backbone, PAN pooling, training data, and parameter budget. The mixers tested were: a standard SiLU-MLP (MLP); a Kolmogorov–Arnold Network (KAN)[13]; a kernel-activation-function network (KAF)[14]; a learnable rational-function variant (LibraKAN); and the proposed PGS edge+readout module. Table S6 reports the resulting energy and force MAE on Ag, Si, and the LiF/Li–F short-range ionic subset. The PGS column matches the MACE+PAN+PGS entries of Table 1 of the main text by construction.

PGS yields the lowest energy and force MAEs across all three systems. The ordering MLP > KAN > KAF > LibraKAN > PGS is consistent across systems and across all five seeds, supporting the mechanistic claim in the main text that the gain is not explained by adding any high-capacity scalar mixer but is specifically associated with spectral enrichment of the invariant channel.

Table S7 Component ablation for the PAN–PGS architecture. Force MAE in meV/Å for the three architectural variants. All variants share the same MACE backbone and training protocol. Values are the mean ± standard deviation over five training seeds.

System	PAN only	PGS only	PAN + PGS
Ag	32.5 ± 0.8	31.5 ± 0.8	$\mathbf{27.9 \pm 0.7}$
Si	41.3 ± 1.1	38.7 ± 1.0	$\mathbf{35.8 \pm 1.0}$
LiF/Li–F	47.5 ± 1.4	44.5 ± 1.3	$\mathbf{39.6 \pm 1.2}$

Table S8 PGS insertion-location ablation. Force MAE in meV/Å for four PGS placements under matched PAN pooling. Values are the mean ± standard deviation over five training seeds.

System	Edge	Node	Readout	Edge + Readout
Ag	30.4 ± 0.8	34.0 ± 0.9	29.5 ± 0.8	$\mathbf{27.9 \pm 0.7}$
Si	38.5 ± 1.1	43.5 ± 1.2	37.4 ± 1.0	$\mathbf{35.8 \pm 1.0}$
LiF/Li–F	43.6 ± 1.3	49.5 ± 1.4	42.2 ± 1.3	$\mathbf{39.6 \pm 1.2}$

S3.4 Component and placement ablations

To separate the contributions of the two PAN–PGS components and the role of PGS placement, we ran two further ablations under the same MACE backbone, training data, and budget.

Component ablation: PAN only, PGS only, PAN+PGS. We isolated the effect of each component by comparing three variants: MACE+PAN with the default smooth scalar mixer ("PAN only"); baseline MACE with the PGS edge+readout module on top of uniform sum aggregation ("PGS only"); and the full MACE+PAN+PGS model. Force MAE on Ag, Si, and the LiF/Li–F subset is reported in Table S7. The "PAN only" column matches the MLP entry of Table S6 by construction (default smooth mixer with PAN). The "PAN+PGS" column matches the MACE+PAN+PGS entry of Table 1 of the main text.

The component comparison shows that PAN and PGS target distinct bottlenecks: neighborhood-adaptive scalar aggregation and scalar spectral resolution, respectively. The combined model yields the largest reduction across all systems, with no evidence of redundancy between the two modules.

Placement ablation: edge, node, readout, edge+readout. We next tested the design choice of applying PGS at the edge level and at the readout level, by inserting the same PGS module at four candidate locations under matched PAN pooling: edge-only, where PGS is applied to invariant edge channels before tensor-product coupling; node-only, where PGS is applied to invariant node channels inside interaction blocks; readout-only, where PGS is applied only to the final scalar readout; and edge+readout, the proposed configuration.

Table S9 Capacity-matched MACE controls vs MACE+PAN+PGS. Force MAE (meV/Å) on Ag, Si, and the LiF/Li–F short-range ionic subset for capacity-matched MACE variants and for MACE+PAN+PGS. All variants use the same training protocol, splits, optimizer, and 5-seed schedule as for the main results. Parameters and inference FLOPs are normalized to baseline MACE. PAN+PGS achieves substantially lower force MAE than every capacity-matched control while costing roughly an order of magnitude less in additional parameters and FLOPs. Values are the mean ± standard deviation over five training seeds.

Model	Params (rel.)	FLOPs (rel.)	F_{Ag}	F_{Si}	$F_{\mathrm{LiF/Li-F}}$	Ag ΔF
MACE (baseline)	1.00×	1.00×	35.9 ± 0.9	47.1 ± 1.4	54.0 ± 1.6	0.0%
MACE-wide (C=192)	1.65×	1.50×	34.1 ± 0.9	44.8 ± 1.3	51.5 ± 1.5	5.0%
MACE-deep (L=3)	1.42×	1.40×	33.5 ± 0.8	44.2 ± 1.3	50.8 ± 1.5	6.7%
MACE-Nrad16	1.08×	1.10×	34.0 ± 0.9	44.5 ± 1.3	51.0 ± 1.5	5.3%
MACE + PAN + PGS (ours)	**1.03**×	**1.05**×	$\mathbf{27.9 \pm 0.7}$	$\mathbf{35.8 \pm 1.0}$	$\mathbf{39.6 \pm 1.2}$	**22.3**%

Three observations follow. First, node-level PGS underperforms even baseline MACE+PAN (the MLP entry of Table S6), indicating that spectral mixing inside the interaction block destabilizes the multiplicative tensor-product structure on which equivariant scaling depends. Node-channel scalars are therefore best left unmodified. Second, edge-only and readout-only PGS each improve over MACE+PAN, with readout-only marginally stronger than edge-only across the three systems. Third, the edge+readout configuration outperforms either placement alone across all systems, supporting the choice in the main architecture: edge-level PGS refines short-range geometric information before tensor-product coupling, while readout-level PGS enriches the final invariant latent before energy prediction. Together, the placement controls rule out the explanation that the PAN–PGS gain arises solely from inserting additional parameters near the readout, since both standalone placements were tested.

Capacity robustness. To confirm that the observed gains are not sensitive to the specific PGS capacity, we varied the number of spectral components $M \in \{2, 4, 8, 12, 16, 24\}$ and the readout projection dimension $d \in \{8, 16, 32\}$. Force MAE varied by less than 4% across this range in every system, and the ordering of mixer performance in Table S6 (MLP > KAN > KAF > LiBraKAN > PGS) was preserved. The configuration used in the main text ($M = 16$, d = matched to the readout hidden dimension) sits within the plateau region for both axes.

S3.5 Capacity-matched MACE controls

A natural alternative explanation of the gain reported in the main text is that the same improvement could be obtained simply by enlarging the MACE backbone. To test this, we trained three capacity-matched MACE controls under the same training protocol, dataset splits, optimizer, and seed schedule used for the main results, varying only one capacity-related hyperparameter at a time: **MACE-wide**, in which the hidden channel multiplicity is increased from 128 to 192 (∼1.65× parameters); **MACE-deep**, in which the number of interaction blocks is increased from 2 to 3 (∼1.42× parameters); and **MACE-Nrad16**, in which the radial Bessel basis size is doubled from $N_{\mathrm{rad}} = 8$ to $N_{\mathrm{rad}} = 16$ (∼1.08× parameters). The third control specifically targets the alternative hypothesis that PGS's gain is recoverable by simply enlarging the radial basis, since PGS is presented as a spectral-resolution module.

Table S10 LiF/Li–F grouped-split sanity check. Energy and force MAE are reported for baseline MACE and MACE+PAN+PGS under the configuration-level split used in the main text, and a stricter Materials-Project-identifier-grouped split. Values are mean ± s.d. over five seeds. Energies are in meV/atom, and forces are in meV/Å.

Split	Qty.	MACE	MACE +PAN+PGS	Reduction (%)	Δ vs config.
Config.-level (main text)	E	3.09 ± 0.12	$\mathbf{2.41 \pm 0.10}$	22.0	—
	F	54.0 ± 1.6	$\mathbf{39.6 \pm 1.2}$	26.7	—
MP-id grouped (this section)	E	3.62 ± 0.16	$\mathbf{2.84 \pm 0.13}$	21.5	−0.5 pp
	F	61.8 ± 2.0	$\mathbf{45.4 \pm 1.5}$	26.5	−0.2 pp

The three capacity-matched controls each reduce the Ag force MAE by 5–7% relative to baseline MACE, with similar relative reductions on Si and the LiF/Li–F subset. By contrast, MACE+PAN+PGS reduces the Ag force MAE by 22.3% at 1.03× the parameters and 1.05× the inference FLOPs of baseline MACE. In particular, MACE-Nrad16, which directly addresses the alternative hypothesis that PGS's gain is recoverable by simply enlarging the radial basis, achieves a 5.3% reduction at 1.10× FLOPs—about one quarter of the PAN+PGS reduction at twice the FLOPs overhead. The capacity-matched comparison, therefore, makes "backbone capacity" a less likely explanation for the gain reported in Table 1 of the main text and supports the mechanistic interpretation discussed in Section 2.3.

S3.6 LiF/Li–F sanity check under a stricter (grouped) split

The LiF/Li–F results in the main text use a configuration-level 80/10/10 split, which can place closely related snapshots from the same MPtrj relaxation trajectory in both the training and the test set. To assess whether the relative improvement in MACE+PAN+PGS over baseline MACE holds up under a stricter held-out evaluation, we performed a sanity-check training split by Materials Project identifier. All configurations sharing a Materials Project identifier (i.e. all snapshots from the same parent material's relaxation trajectory) were assigned entirely to either the training or the test set, with no overlap between splits. The resulting split keeps the same overall 80/10/10 ratio.

Under this stricter grouped split, the absolute energy and force MAE values increase relative to the configuration-level values reported in the main text, as expected for a harder generalization task. However, the relative improvement of MACE+PAN+PGS over baseline MACE is preserved, with reductions within ±1 percentage point of the configuration-level values (Table S10). We therefore present the LiF/Li–F subset only as a short-range ionic held-out-configuration test in the main results, and use the grouped split as a sanity check that the relative scalar-pathway improvement is not specific to the configuration-level split.

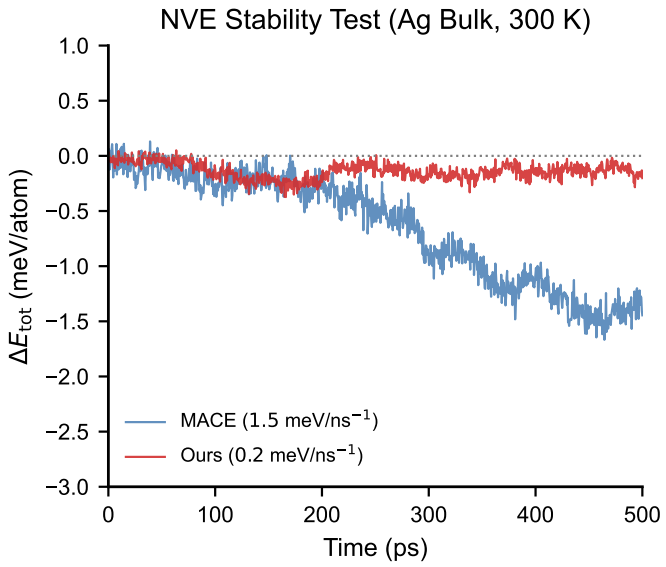

Fig. S1 NVE energy conservation for bulk Ag at 300 K (108-atom supercell). Total-energy deviation $\Delta E_{\mathrm{tot}}(t)$ during a 500 ps NVE trajectory using baseline MACE and MACE+PAN+PGS. Simulations used a velocity-Verlet integrator with a 1 fs timestep and a 108-atom Ag supercell. Curves correspond to a representative seed for visual clarity. The aggregate drift across five seeds is 1.51 ± 0.20 meV ns^{-1} per atom for baseline MACE and 0.23 ± 0.05 meV ns^{-1} per atom for MACE+PAN+PGS.

S4 Additional dynamical and structural validation

This section reports three further validation tests for the dynamical behavior discussed in Section 2.4 of the main text: long-time NVE energy conservation in a smaller Ag system, the liquid Ag radial distribution function under matched conditions, and an additional flexible-molecule torsional distribution (malonaldehyde).

S4.1 NVE energy conservation in a smaller system

Supplementary Fig. S1 shows that the small-system NVE drift follows the same pattern as the large-scale 10,000-atom system reported in Section 2.4. MACE+PAN+PGS substantially reduces the linear drift component and narrows the fluctuation envelope relative to the baseline. The agreement between the small- and large-scale system tests indicates that the drift improvement is a property of the architecture rather than a function of simulation size.

S4.2 Liquid Ag radial distribution function and additional molecular torsions

Supplementary Fig. S2a shows the liquid Ag radial distribution function (RDF) at 1,200 K. Baseline MACE allows atoms to approach unphysically closely, with a spurious sub-shell peak in $g(r)$ at $r \approx 2.3$ Å. MACE+PAN+PGS attenuates

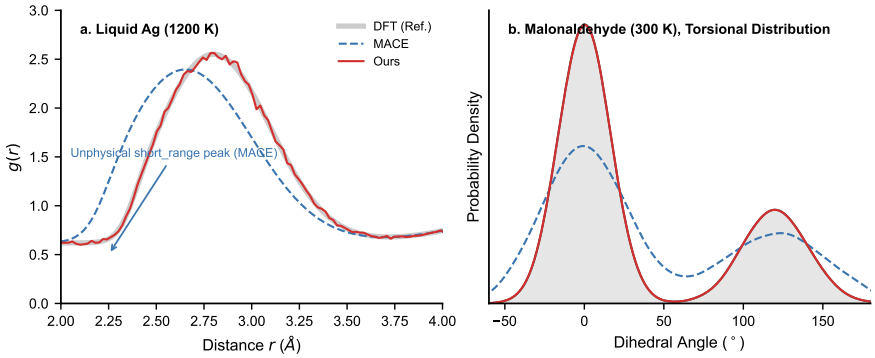

Fig. S2 Additional structural and conformational validation across metallic and molecular systems. (a) Liquid Ag radial distribution function $g(r)$ at 1,200 K for DFT, baseline MACE, and MACE+PAN+PGS. **(b)** Malonaldehyde O–C–C–O torsional-angle distribution at 300 K for DFT, baseline MACE, and MACE+PAN+PGS. Both observables extend the main-text tests to one liquid and one additional flexible molecule under matched protocols.

this artifact and more closely matches the DFT short-range repulsion and medium-range coordination-shell structure. This is consistent with the spectral enrichment mechanism shown for the Ag–Ag and Si–Si compression pathways in the main text.

Supplementary Fig. S2b reports the malonaldehyde torsional distribution at 300 K. As with aspirin in Section 2.4, baseline MACE places excess probability near the torsional barrier, whereas MACE+PAN+PGS more closely matches the sharper bimodal DFT distribution with the correct peak separation and relative occupancy. The agreement between the additional molecule and aspirin indicates that the conformational improvement is not specific to a single molecular system.

S5 Detailed proof of Proposition 1: symmetry and differentiability of PAN and PGS

This section provides a detailed proof of Proposition 1 of the main text. The proposition states that the PAN and PGS modifications defined in Eqs. (10)–(12) of the main text preserve (i) permutation invariance of the energy with respect to neighbor indexing, (ii) $O(3)$-equivariance of the predicted forces and stress tensor, and (iii) the differentiability required for analytic force and stress evaluation under the smooth cutoff and tapered-kernel implementation described in Supplementary Section S6.

For convenience, we recall the notation. The total energy of a configuration $\{(\mathbf{r}_i, Z_i)\}_{i=1}^N$ in a cell with strain $\boldsymbol{\epsilon}$ is

$$E_{\mathrm{tot}} = \sum_{i=1}^{N} \varepsilon_\theta\left(y_i^{(0)}\right), \tag{S2}$$

where $y_i^{(0)} \in \mathbb{R}^{C_n}$ is the readout-level PGS output of Eq. (12), ε_θ is a smooth invariant readout MLP, and the per-atom invariant pre-readout descriptor $x_i^{(0)}$ is itself constructed from the modified scalar edge features $\hat{h}_{ij}^{(0)}$ (Eq. 11) and the PAN-aggregated scalar messages $h_i^{(0)} = \sum_{j\in\mathcal{N}(i)} a_{ij}\hat{h}_{ij}^{(0)}$ (Eq. 10). The PAN gate is $a_{ij} = \sigma(f_\theta(\psi_{ij}))$ with the invariant edge descriptor $\psi_{ij} = [h_{ij}^{(0)}, r_{ij}, Z_i, Z_j, \rho_i, \eta_i]$, where ρ_i and η_i are the cutoff-weighted coordination and angular-variance descriptors defined in the main text. Throughout the proof, we work under the fixed-candidate-neighbor convention used during each force and stress evaluation, so the candidate neighbor set $\mathcal{N}(i)$ is held fixed, and we restrict to noncoincident configurations ($r_{ij} > 0$ for all $i \neq j$).

S5.1 Permutation invariance

Lemma 1 (Permutation invariance of the per-atom scalar pathway) For every atom i, the per-atom invariant descriptor $y_i^{(0)}$ defined through Eqs. (10)–(12) is invariant under any permutation of the neighbor indices in $\mathcal{N}(i)$.

Proof Let $\pi : \mathcal{N}(i) \to \mathcal{N}(i)$ be an arbitrary permutation. The PGS-modified edge feature $\hat{h}_{ij}^{(0)}$ at fixed receiving atom i is a function only of the per-edge data $(h_{ij}^{(0)}, r_{ij}, Z_j)$, so its value attached to neighbor j is unchanged when j is replaced by $\pi(j)$. The local descriptors are

$$\rho_i = \sum_{j\in\mathcal{N}(i)} f_{\mathrm{cut}}(r_{ij}), \qquad \eta_i = \frac{\sum_{j,k\in\mathcal{N}(i),\, j\neq k} f_{\mathrm{cut}}(r_{ij})\, f_{\mathrm{cut}}(r_{ik})\, (c_{jik} - \bar{c}_i)^2}{\epsilon_{\mathrm{reg}} + \sum_{j,k\in\mathcal{N}(i),\, j\neq k} f_{\mathrm{cut}}(r_{ij})\, f_{\mathrm{cut}}(r_{ik})}. \tag{S3}$$

Both expressions are unordered sums over $\mathcal{N}(i)$ (and an unordered double sum over the multiset $\mathcal{N}(i) \times \mathcal{N}(i)$ with the $j \neq k$ constraint), and are therefore invariant under π. The PAN gate $a_{ij} = \sigma(f_\theta(\psi_{ij}))$ is consequently a function only of π-invariant inputs, and the aggregated scalar

$$h_i^{(0)} = \sum_{j\in\mathcal{N}(i)} a_{ij}\, \hat{h}_{ij}^{(0)} = \sum_{j\in\mathcal{N}(i)} a_{i\pi(j)}\, \hat{h}_{i\pi(j)}^{(0)} \tag{S4}$$

is invariant under π because it is a sum over the unordered multiset $\mathcal{N}(i)$. Subsequent equivariant interaction blocks combine π-invariant scalar messages with vector and tensor irreps that are themselves aggregated by unordered sums over $\mathcal{N}(i)$, so the final invariant per-atom descriptor $x_i^{(0)}$ is π-invariant. The readout-level PGS [Eq. 12] is a single-atom operator acting only on $x_i^{(0)}$, so $y_i^{(0)}$ inherits the same invariance. □

Corollary 2 The total energy E_{tot} in Eq. (S2) is invariant under any global relabelling of atomic indices, since it is a sum over atoms of per-atom quantities, each of which is itself invariant under permutations of its neighbor indices.

S5.2 $O(3)$-equivariance of the forces and stress tensor

Lemma 3 ($O(3)$-invariance of the energy) Under a global rotation–inversion $R \in O(3)$ acting on atomic positions by $\mathbf{r}_i \mapsto R\mathbf{r}_i$, $E_{\mathrm{tot}}(R\mathbf{r}_1, \ldots, R\mathbf{r}_N) = E_{\mathrm{tot}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$.

Proof Each entry of the PAN gate input $\psi_{ij} = [h_{ij}^{(0)}, r_{ij}, Z_i, Z_j, \rho_i, \eta_i]$ is built from $O(3)$-invariant primitives:

1. $r_{ij} = \|\mathbf{r}_j - \mathbf{r}_i\|$ is invariant by orthogonality of R.
2. Atomic numbers Z_i, Z_j are scalar labels and unchanged.
3. $\rho_i = \sum_{j \in \mathcal{N}(i)} f_{\text{cut}}(r_{ij})$ depends only on invariant distances.
4. The angular variance η_i uses $c_{jik} = \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ik}$, which is $O(3)$-invariant because $(R\hat{\mathbf{r}}_{ij}) \cdot (R\hat{\mathbf{r}}_{ik}) = \hat{\mathbf{r}}_{ij}^\top R^\top R\, \hat{\mathbf{r}}_{ik} = \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ik}$.
5. $h_{ij}^{(0)}$ is the ℓ=0 component of the MACE pair embedding and transforms as an $O(3)$ scalar by construction of the backbone.

Therefore a_{ij} is $O(3)$-invariant. The edge-level PGS [Eq. 11] applies a real-valued radial basis $B_m^{\text{FB}}(r_{ij})$ to the invariant distance and modifies only the ℓ=0 channel, so $\hat{h}_{ij}^{(0)}$ is also $O(3)$-invariant. The Clebsch–Gordan tensor-product machinery of MACE is unchanged and is $O(3)$-equivariant by construction; in particular, the final invariant per-atom descriptor $x_i^{(0)}$ transforms as an $O(3)$ scalar. The readout-level PGS [Eq. 12] acts on $x_i^{(0)}$ via the learned scalar projection $u_m^\top x_i^{(0)}$, so $y_i^{(0)}$ is also an $O(3)$ scalar. The total energy is a sum of $O(3)$ scalars and is therefore $O(3)$-invariant. □

Corollary 4 (Tensorial transformation of forces and stress) The predicted forces and stress tensor obtained as analytic derivatives of E_{tot} transform as $\mathbf{F}_i \mapsto R\mathbf{F}_i$ and $\boldsymbol{\sigma} \mapsto R\boldsymbol{\sigma}R^\top$ under any $R \in O(3)$.

Proof For the forces, applying Lemma 3 and differentiating with respect to $\mathbf{r}_k$ gives, by the chain rule,

$$\nabla_{\mathbf{r}_k}\left[E_{\text{tot}} \circ R^{\otimes N}\right] = R^\top \nabla_{(R\mathbf{r}_k)} E_{\text{tot}}, \tag{S5}$$

where $R^{\otimes N}$ denotes the simultaneous action of R on every atomic position. Equating both sides with the gradient of the original energy and using the definition $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E_{\text{tot}}$ yields $\mathbf{F}_i(R\mathbf{r}) = R\mathbf{F}_i(\mathbf{r})$. For the stress tensor, the cell strain transforms as $\boldsymbol{\epsilon} \mapsto R\boldsymbol{\epsilon}R^\top$ under $R \in O(3)$, so $\boldsymbol{\sigma} = (1/V)\partial E_{\text{tot}}/\partial \boldsymbol{\epsilon}$ transforms as a rank-2 tensor: $\boldsymbol{\sigma} \mapsto R\boldsymbol{\sigma}R^\top$. □

S5.3 Differentiability

We now establish the C^2 regularity of E_{tot} used in the analytic force and stress derivations. Three lemmas establish the C^2 regularity of the three smooth elementary objects entering the PAN–PGS construction; the main theorem then assembles them via the chain rule.

Lemma 5 (C^2 regularity of cutoff-weighted descriptors) The polynomial cutoff envelope

$$f_{\text{cut}}(r) = \begin{cases} 1 - 6(r/r_{\text{cut}})^5 + 15(r/r_{\text{cut}})^4 - 10(r/r_{\text{cut}})^3, & 0 \le r \le r_{\text{cut}}, \\ 0, & r > r_{\text{cut}}, \end{cases} \tag{S6}$$

is of class C^2 on $[0, \infty)$ and C^∞ on $[0, r_{\text{cut}})$. Consequently, ρ_i and η_i are C^2 functions of atomic positions on the noncoincident manifold.

Proof On $[0, r_{\mathrm{cut}}]$ the cutoff envelope is a polynomial and hence C^∞. At $r = r_{\mathrm{cut}}$, direct differentiation gives $f_{\mathrm{cut}}(r_{\mathrm{cut}}) = 0$, $f'_{\mathrm{cut}}(r_{\mathrm{cut}}) = 0$, $f''_{\mathrm{cut}}(r_{\mathrm{cut}}) = 0$, while $f'''_{\mathrm{cut}}(r_{\mathrm{cut}}) \neq 0$, so f_{cut} is C^2 but not C^3 at the boundary. On $r > r_{\mathrm{cut}}$ it is identically zero. The descriptor ρ_i is a finite sum of $f_{\mathrm{cut}}(r_{ij})$, each of which is C^2 in $(\mathbf{r}_i, \mathbf{r}_j)$ on the noncoincident manifold (since $r_{ij} > 0$ implies r_{ij} is C^∞ in $\mathbf{r}_i, \mathbf{r}_j$); ρ_i is therefore C^2. For η_i, the numerator and denominator are C^2 functions of atomic positions on the noncoincident manifold, because $c_{jik} = \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ik}$ is C^∞ in $(\hat{\mathbf{r}}_{ij}, \hat{\mathbf{r}}_{ik})$, and $\hat{\mathbf{r}}_{ij}$ is C^∞ for $r_{ij} > 0$. The denominator is strictly positive because $\epsilon_{\mathrm{reg}} > 0$, so the quotient is C^2. □

Lemma 6 (C^2 regularity of the Fourier–Bessel basis) The Fourier–Bessel basis $B_m^{\mathrm{FB}}(r) = \sqrt{2/r_{\mathrm{cut}}}\ \sin(m\pi r/r_{\mathrm{cut}})/r \cdot f_{\mathrm{cut}}(r)$ is C^2 on $(0, \infty)$ and admits a C^2 extension to $r = 0$ via the continuous limit $\lim_{r\to 0^+} \sin(m\pi r/r_{\mathrm{cut}})/r = m\pi/r_{\mathrm{cut}}$. On the noncoincident configuration manifold (where all $r_{ij} > 0$), the extension is not invoked.

Proof On $r > 0$, $\sin(m\pi r/r_{\mathrm{cut}})/r$ is a quotient of C^∞ functions with non-vanishing denominator and is therefore C^∞. Its Taylor expansion at $r = 0$ is real-analytic, so multiplication by $f_{\mathrm{cut}} \in C^2([0, \infty))$ yields a C^2 function on $[0, \infty)$. □

Lemma 7 (C^2 regularity of the tapered ES kernel) Let $z = (t - t_m)/w$. The shifted, tapered ES kernel

$$K_m^{\mathrm{ES}}(t) = \begin{cases} \tau_w(z)\left[\exp\big(\beta(\sqrt{1-z^2}-1)\big) - e^{-\beta}\right], & |z| \le 1, \\ 0, & |z| > 1, \end{cases} \tag{S7}$$

where τ_w is the same C^2 polynomial taper as f_{cut} applied symmetrically on $|z| \le 1$, is of class C^2 on $\mathbb{R}$.

Proof On the open interior $|z| < 1$, the kernel is a product of C^∞ factors (taper, exponential, square root with strictly positive argument). At the boundary $|z| = 1$, expanding around $u = \sqrt{1-z^2} \to 0$ gives

$$\exp\big(\beta(\sqrt{1-z^2}-1)\big) - e^{-\beta} = e^{-\beta}\big(e^{\beta u} - 1\big) = e^{-\beta}\beta u + O(u^2), \tag{S8}$$

so the bracketed term behaves as $O((1-|z|)^{1/2})$ as $|z| \to 1^-$. The taper τ_w inherits the C^2 structure of f_{cut} near the boundary, so $\tau_w(z), \tau'_w(z), \tau''_w(z) \to 0$ as $|z| \to 1^-$, with leading-order behavior $\tau_w(z) = O((1-|z|)^3)$. The product therefore behaves as $\tau_w(z) \cdot [\,\cdot\,] = O((1-|z|)^{7/2})$ near $|z| = 1$, whose first three derivatives vanish at $|z| = 1$. Matching to the identically zero exterior region, K_m^{ES} is C^2 at $|z| = 1$ and hence on all of $\mathbb{R}$. □

Theorem 8 (Differentiability of E_{tot}) *Under the noncoincident-configuration and fixed-candidate-neighbor conventions, E_{tot} is a C^2 function of the $3N$ atomic-position coordinates and the 6 independent components of the cell strain $\boldsymbol{\epsilon}$. The forces and stress tensor are defined as analytic autograd derivatives of E_{tot}*

$$\mathbf{F}_i = -\frac{\partial E_{\mathrm{tot}}}{\partial \mathbf{r}_i}, \qquad \boldsymbol{\sigma} = \frac{1}{V}\frac{\partial E_{\mathrm{tot}}}{\partial \boldsymbol{\epsilon}}, \tag{S9}$$

are therefore well defined and continuous, and coincide with their variational definitions.

Proof The energy admits the finite composition $E_{\mathrm{tot}} = \sum_i \varepsilon_\theta \circ \Pi^{(0)} \circ \mathcal{G}$, where $\mathcal{G}$ collects the equivariant tensor-product backbone (polynomial Clebsch–Gordan combinations of equivariant features) and $\Pi^{(0)}$ extracts the invariant ℓ=0 component before the readout-level PGS lift [Eq. 12]. Each constituent operation is at least C^2 on the noncoincident manifold: the radial Bessel embedding through f_{cut} is C^2 (Lemma 5); the edge-level Fourier–Bessel basis B_m^{FB} is C^2 on the relevant domain (Lemma 6); the readout-level ES kernel is C^2 on $\mathbb{R}$ (Lemma 7); the MLPs $f_\theta, g_\theta, h_{\mathrm{low}}$, and ε_θ use smooth (C^∞) activations and are therefore C^∞ in their inputs; tensor products with Clebsch–Gordan coefficients are polynomial and hence C^∞. The composition of C^2 functions is C^2, so E_{tot} is C^2. Forces and stress, defined as first derivatives, are therefore continuous (in fact C^1) on the noncoincident manifold.

The cell-strain dependence enters only through the radial distances $r_{ij}(\boldsymbol{\epsilon})$, which are smooth functions of $\boldsymbol{\epsilon}$ on noncoincident configurations; the same chain of regularity therefore yields C^2 dependence on $\boldsymbol{\epsilon}$. Higher regularity (C^k for $k > 2$) holds on configurations where every pair distance lies strictly below r_{cut}, since on that open subdomain f_{cut} acts as a C^∞ polynomial; we do not claim higher regularity globally because f_{cut} saturates to C^2 regularity at $r = r_{\mathrm{cut}}$. □

S5.4 Remarks on the convention boundaries

Two convention boundaries are explicit in Proposition 1 and are recalled here. First, the proof is restricted to noncoincident configurations. At coincident configurations ($r_{ij} \to 0$), the radial Bessel basis remains continuous via the limit in Lemma 6, but the normalized pair direction $\hat{\mathbf{r}}_{ij}$ used inside the equivariant backbone is undefined. This is a property of the underlying MACE backbone rather than of PAN or PGS, and is outside the scope of the proof. Second, the proof assumes a fixed candidate neighbor graph during each force and stress evaluation. The use of the C^2 cutoff envelope ensures that adding or removing a candidate pair at $r_{ij} = r_{\mathrm{cut}}$ contributes zero to the energy, force, and stress at the crossing; alternative neighbor-list rebuild policies in specific MD engines may introduce additional bookkeeping discontinuities external to the model, which are not properties of the PAN–PGS architecture.

S6 Supplementary methods: algorithmic details

This section provides the explicit forms of the smooth elementary functions used inside PAN and PGS, together with a layer-by-layer summary of how the two modules are inserted into the MACE forward pass.

Cutoff envelope. All cutoff-based smoothing in PAN, PGS, and the MACE radial embedding uses the same polynomial envelope $f_{\mathrm{cut}} \colon [0, r_{\mathrm{cut}}] \to [0, 1]$, defined as

$$f_{\mathrm{cut}}(r) = \begin{cases} 1 - 6\,(r/r_{\mathrm{cut}})^5 + 15\,(r/r_{\mathrm{cut}})^4 - 10\,(r/r_{\mathrm{cut}})^3, & r \le r_{\mathrm{cut}}, \\ 0, & r > r_{\mathrm{cut}}, \end{cases} \tag{S10}$$

which satisfies $f_{\mathrm{cut}}(0) = 1$, $f_{\mathrm{cut}}(r_{\mathrm{cut}}) = 0$, and f_{cut} is C^2 at $r = r_{\mathrm{cut}}$.

Edge-level Fourier–Bessel basis. The edge-level PGS spectral basis $\{B_m^{\mathrm{FB}}(r)\}_{m=1}^{M}$ on $[0, r_{\mathrm{cut}}]$ is the standard real-valued Bessel–root basis

$$B_m^{\mathrm{FB}}(r) = \sqrt{\frac{2}{r_{\mathrm{cut}}}}\,\frac{\sin\!\big(m\pi\, r/r_{\mathrm{cut}}\big)}{r}\, f_{\mathrm{cut}}(r), \qquad m = 1, \ldots, M, \tag{S11}$$

multiplied by the polynomial cutoff envelope f_{cut} to ensure C^2 vanishing at $r = r_{\mathrm{cut}}$. The factor $\sin(m\pi r/r_{\mathrm{cut}})/r$ is taken at its continuous limit $m\pi/r_{\mathrm{cut}}$ when $r \to 0$; in practice all $r = r_{ij}$ are strictly positive in the neighbor graph. The basis is L^2-orthonormal on $[0, r_{\mathrm{cut}}]$ in the absence of the cutoff envelope, and we used $M = 16$ throughout.

Readout-level Exponential-of-Semicircle (ES) kernel. The readout-level PGS uses the ES kernel basis introduced for non-uniform fast Fourier transforms (NUFFTs) [15]. For a projected latent value $t = u_m^{\top} x_i^{(0)} \in [-t_{\max}, t_{\max}]$, the m-th basis function is implemented in the shifted exponential-of-semicircle form, multiplied by a smooth C^2 taper, so that K_m^{ES} and its first two derivatives vanish continuously at the support boundary with no hard indicator function entering the computational graph:

$$K_m^{\mathrm{ES}}(t) = \tau_w(z)\,\Big[\,\exp\!\big(\beta\,(\sqrt{1-z^2}-1)\big) - e^{-\beta}\,\Big], \qquad z = \frac{t - t_m}{w}, \qquad |z| \le 1, \tag{S12}$$

where $\tau_w(z)$ is the same C^2 polynomial taper as the radial cutoff f_{cut}, applied symmetrically on $|z| \le 1$. The shifted exponential-of-semicircle term equals $e^{-\beta}$ at $|z| = 1$, so the bracketed term vanishes exactly at the support boundary; multiplying by τ_w additionally enforces vanishing of the first two derivatives. The centers $\{t_m\}_{m=1}^{M}$ are uniformly spaced on $[-t_{\max}, t_{\max}]$, w is the support half-width, and β is the standard NUFFT shape parameter. Following the recommended NUFFT settings, we used $w = 4\Delta t$ (where Δt is the center spacing) and $\beta = 2.30\, w$. We used $M = 16$ ES components throughout.

Descriptor normalization. The PAN local descriptors ρ_i and η_i are standardized to zero mean and unit variance using statistics computed once on the training set and then frozen. The radial input r_{ij} to ψ_{ij} is encoded through the same Bessel basis used by MACE, so no separate normalization is needed for r_{ij}.

Insertion points in MACE. PAN and PGS are inserted into the standard MACE forward pass with only a few well-defined modifications. Algorithm S1 summarizes the modified forward pass; in MACE notation, edge-level PGS operates on the scalar component $h_{ij}^{(0)}$ of the equivariant edge embedding before tensor-product coupling, PAN modulates the scalar message amplitude before accumulation over the neighbor set within each interaction block, while leaving the learned message functions unchanged, and readout-level PGS is applied after the final interaction block on the invariant per-atom descriptor before the energy MLP.

Algorithm S1 PAN+PGS forward pass on a single configuration.

Require: Atomic positions $\{\mathbf{r}_i\}$, atomic numbers $\{Z_i\}$, cell tensor $\mathbf{h}$.

1: Build neighbor graph using cutoff radius r_{cut}.
2: For each edge (i, j), compute r_{ij}, $\hat{\mathbf{r}}_{ij}$, and the equivariant edge embedding $\phi_{ij,n}^{(\ell,m)} = B_n(r_{ij})\, Y_\ell^m(\hat{\mathbf{r}}_{ij})$.
3: **Edge-level PGS** [Eq. (11)]: replace $h_{ij}^{(0)}$ by $\hat{h}_{ij}^{(0)}$ via the smooth low branch and the Fourier–Bessel spectral mixer.
4: **for** each interaction block $\ell = 1, \ldots, L$ **do**
5: Compute Clebsch–Gordan tensor products using the modified scalar edge channel and the unchanged $\ell{>}0$ channels.
6: **PAN gating** [Eq. (10)]: compute $a_{ij} = \sigma(f_\theta(\psi_{ij}))$ from the smooth invariant descriptor ψ_{ij}, and aggregate the scalar message channel as $h_i^{(0)} = \sum_j a_{ij} \hat{h}_{ij}^{(0)}$. The $\ell{>}0$ channels are aggregated by the unchanged uniform sum.
7: **end for**
8: **Readout-level PGS** [Eq. (12)]: lift $x_i^{(0)}$ to $y_i^{(0)}$ via the ES-kernel expansion of the learned projection $u_m^\top x_i^{(0)}$.
9: Compute atomic energies $E_i = \varepsilon_\theta(y_i^{(0)})$ and total energy $E_{\text{tot}} = \sum_i E_i$.
10: Forces and stress: $\mathbf{F}_i = -\partial E_{\text{tot}}/\partial \mathbf{r}_i$, $\boldsymbol{\sigma} = (1/V)\partial E_{\text{tot}}/\partial \boldsymbol{\epsilon}$ via automatic differentiation.

S7 Supplementary methods: cross-backbone implementation

The cross-backbone entries in Table 1 of the main text use Allegro[2] and NequIP[4] backbones in addition to MACE. For all three backbones, we used the same author-generated Ag dataset, the same public Si and LiF/Li–F datasets, and the same train/validation/test splits as in the MACE experiments, so that any difference between backbone-only and backbone+PAN+PGS columns reflects only the scalar-pathway modification.

Allegro and NequIP backbones. Allegro was instantiated with two interaction layers, $\ell_{\max} = 2$, 128 scalar and tensor channels per irrep, the same $r_{\text{cut}} = 5.0$ Å used for MACE, and a Bessel radial basis of size $N_{\text{rad}} = 8$. NequIP was instantiated with three interaction blocks, $\ell_{\max} = 2$, 128 feature channels, the same cutoff $r_{\text{cut}} = 5.0$ Å, and the same Bessel radial basis size $N_{\text{rad}} = 8$. Both models were trained with AdamW (initial learning rate 5×10^{-3}, weight decay 10^{-8}), the same ReduceLROnPlateau schedule, EMA decay, and SWA settings as MACE, batch size 4, gradient clipping at 100, and the same loss weights $w_E = 1.0$, $w_F = 10.0$, $w_S = 10.0$ when stress labels are available. Each model was trained five times under seeds $\{1, 2, 3, 4, 5\}$.

Insertion of PAN and PGS into Allegro and NequIP. The PAN gate was inserted at the scalar message aggregation stage of each backbone: it modulates the scalar message amplitude before accumulation over the neighbor set, while leaving the learned message functions and all $\ell{>}0$ tensor operations unchanged. The same invariant descriptor $\psi_{ij} = [h_{ij}^{(0)}, r_{ij}, Z_i, Z_j, \rho_i, \eta_i]$ defined in Eq. (10) was used, and

Table S11 Architectural mapping of PAN and PGS across the three tested equivariant backbones. PAN replaces or modulates the scalar aggregation stage at the closest corresponding scalar-processing point in each backbone; edge-level PGS acts on the ℓ=0 component of the pair embedding before tensor-product coupling; readout-level PGS acts on the final invariant descriptor before the energy head. All higher-order (ℓ>0) channels, tensor-product operations, and backbone-specific equivariant blocks remain unchanged.

Backbone	PAN insertion	Edge-level PGS insertion	Readout-level PGS insertion	Unchanged
MACE[1]	Scalar-channel aggregation stage inside each interaction block; applies a coordination- and distortion-conditioned amplitude gate before the scalar messages are summed over the neighbor set	ℓ=0 component of the pair embedding $h_{ij}^{(0)}$, applied before each tensor-product coupling	Final invariant per-atom descriptor $x_i^{(0)}$, applied before the readout MLP ε_θ	All Clebsch–Gordan tensor products, ℓ>0 channels, radial Bessel basis, and equivariant readout
NequIP[4]	Scalar-channel aggregation stage inside each equivariant convolution layer; applies the same PAN amplitude gate before scalar messages are summed over the neighbor set	ℓ=0 component of the pair embedding before each equivariant convolution	Final invariant node descriptor before the per-element energy MLP	All equivariant convolutions, ℓ>0 irreps, radial Bessel basis, and gate nonlinearities
Allegro[2]	Scalar pair-weighting/contraction stage in the local environment block; the PAN gate is applied before the invariant local contribution is accumulated	ℓ=0 component of the pair latent before each tensor-product mixing	Final invariant per-edge contribution before the energy head	All tensor-product mixings, ℓ>0 irreps, and the strictly local Allegro backbone

PAN parameters (f_θ width, descriptor normalization) were matched to the MACE configuration. Edge-level PGS was inserted on the scalar component of the pair embedding before tensor-product coupling in each backbone (i.e. on the ℓ=0 channel of the per-edge invariant latent), with $M = 16$ Fourier–Bessel components on $[0, r_{\text{cut}}]$. Readout-level PGS was inserted between the final invariant per-atom descriptor of each backbone and the energy MLP, with $M = 16$ ES-kernel components and one learned projection direction per component, identical to the MACE configuration. All higher-order (ℓ>0) channels and tensor-product operations of each backbone were left unchanged.

Table S11 summarizes the architectural mapping used for the three backbones, allowing the reader to verify that PAN and PGS act on functionally analogous locations across the three.

Capacity and FLOPs matching. The three backbones are not parameter-matched to each other because their tensor-product structures differ; each backbone was used at the configuration recommended in its original publication and at the radial cutoff matched to the MACE experiments. The within-backbone comparisons (e.g. NequIP vs NequIP+PAN+PGS) are the controlled signal in Table 1: PAN+PGS adds $\approx$ 3–5% parameters and $\approx$ 4–6% inference FLOPs to each backbone, so each row of Table 1 is an approximately parameter-matched comparison. Cross-backbone differences in absolute MAE between rows reflect the relative expressivity of each backbone and are not the object of this study.

Table S12 DFT settings used to generate the author-generated Ag reference dataset.

Parameter	Value
Exchange–correlation functional	PBE
PAW pseudopotential	`PAW_PBE Ag 06Sep2000`
Plane-wave cutoff (`ENCUT`)	500 eV
Electronic convergence (`EDIFF`)	10^{-5} eV/atom
Force convergence (`EDIFFG`)	-10^{-2} eV/Å
Smearing (`ISMEAR`, `SIGMA`)	1 (Methfessel–Paxton), 0.1 eV
k-mesh, primitive/strained cells	Γ-centered, density ≥ 0.20 Å^{-1}
k-mesh, supercells (defects/surfaces/MD)	Single Γ-point, supercell sized to match density
Strain range	±5% (isotropic and uniaxial)
Defects	Mono- and divacancies, self-interstitials
Surfaces	(100), (110), (111) slabs with vacuum ≥ 12 Å
High-T AIMD sampling	300–1200 K, 1 fs timestep, Nosé–Hoover thermostat

S8 Supplementary methods: DFT settings and molecular-dynamics protocols

Ag DFT settings. Table S12 summarizes the VASP parameters used to generate the author-generated Ag reference dataset.

Aspirin torsion distribution. For Fig. 4(a), aspirin torsional distributions were computed from 1 ns NVT trajectories at 300 K with a Nosé–Hoover thermostat (chain length 3, coupling time 100 fs) and a 0.5 fs timestep. The first 50 ps were discarded as equilibration; torsional values were sampled every 0.5 fs over the final 900 ps, giving 1.8×10^6 torsional values per trajectory. The DFT reference distribution was obtained from *ab initio* molecular dynamics under identical thermostat and timestep settings, with torsional values sampled every 1 fs over the final 90 ps, giving 9×10^4 torsional values.

Si vibrational density of states. For Fig. 4(b), the Si VDOS was computed from the mass-weighted velocity autocorrelation function over 200 ps NVE trajectories at 300 K (after 50 ps NVT equilibration), with velocities recorded every 0.5 fs (4×10^5 velocity snapshots per trajectory). The DFT reference VDOS was generated from *ab initio* molecular dynamics over 20 ps under matched conditions, with velocities recorded every 0.5 fs (4×10^4 snapshots), Fourier-transformed with the same window function as the MLIP VDOS to ensure comparable spectral resolution.

Large-scale Ag NVE drift. For Fig. 4(c), the Ag NVE test used a 10,000-atom fcc Ag bulk system at 300 K. Each trajectory was integrated for 500 ps with a 1 fs timestep after 20 ps NVT equilibration; total energy was recorded every 0.1 ps. Energy drift was estimated by linear regression of accumulated energy deviation over the production segment and reported in meV ns^{-1} per atom. No DFT reference is needed for this metric because the comparison is between energy-conservation properties of the MLIP trajectories themselves.

Liquid Ag radial distribution function. For Fig. 4(d), liquid Ag RDFs were computed from 200 ps NVT trajectories at 1,200 K (after 50 ps NVT equilibration

with the same thermostat settings as the MLIP trajectories), with frames saved every 0.5 ps to give 400 saved frames per trajectory; the RDF for each trajectory was accumulated by summing pair distances over all 400 frames on a uniform grid with bin width 0.02 Å up to $r_{\mathrm{cut}}^{\mathrm{RDF}} = 6.0$ Å. The DFT reference RDF was generated from *ab initio* molecular dynamics over 30 ps at the same temperature, with frames saved every 0.5 ps (60 saved frames), accumulated using identical bin width and identical $r_{\mathrm{cut}}^{\mathrm{RDF}}$ for all three curves.

Finite-sampling uncertainty of DFT references. The DFT-MD reference trajectories used for the VDOS and RDF analyses are necessarily shorter than the corresponding MLIP trajectories because of the cost of *ab initio* molecular dynamics. The reported distances to the DFT reference should therefore be interpreted relative to finite-sampling reference curves rather than as distances to an infinite-sampling limit. To keep the comparison controlled, DFT and MLIP observables were processed with identical bin widths, window functions, normalization conventions, and integration ranges. The main conclusions rely on relative changes between baseline MACE and MACE+PAN+PGS under the same finite DFT reference.

S9 Supplementary methods: computational benchmarks

Table S13 reports the relative computational cost of PAN+PGS and the scalar-mixer controls used in the main text. Costs were measured under matched hardware, batch size, cutoff radius, and force/stress evaluation protocol. Inference FLOPs include the automatic differentiation backward pass through the scalar energy with respect to atomic positions and cell strain.

For practical reference, we measured wall-clock molecular-dynamics throughput on the 10,000-atom Ag NVE benchmark used in Section 2.4 of the main text using a single NVIDIA A100 (40 GB) in float64. At a 1 fs timestep, baseline MACE achieves 4.8 ns/day (≈ 56 MD force evaluations per second; equivalently $\approx 5.6 \times 10^5$ atom-force evaluations per second for the 10,000-atom system). MACE+PAN+PGS achieves 4.6 ns/day under the same protocol (≈ 53 MD force evaluations per second), corresponding to a $\approx 4\%$ throughput cost. The corresponding throughput for the capacity-matched MACE-wide control is 3.2 ns/day (≈ 37 MD force evaluations per second). The PAN+PGS gain reported in Table 1, therefore, comes at a small fraction of the throughput cost of a comparable backbone-capacity expansion.

A capacity-matched MACE control, in which the baseline MACE backbone is widened, deepened, or given a larger radial basis under a matched training budget, is reported in Supplementary Table S9. The capacity-matched variants achieve only 4–7% force-MAE reductions relative to baseline MACE—substantially smaller than the 22–27% reductions reported in the main text for MACE+PAN+PGS—and require 1.4–1.7× more parameters and 1.1–1.5× more inference FLOPs. This supports the interpretation that the gain in Table 1 is not adequately explained by backbone capacity.

Table S13 Relative computational cost across backbones and scalar-mixer alternatives. Parameter count, training time per epoch, and inference FLOPs are normalized to baseline MACE under identical batch size, identical hardware (NVIDIA A100, float64), and matched supercell size. The top block compares the three equivariant backbones reported in the main-text cross-regime accuracy table; the bottom block compares alternative scalar mixers under the MACE backbone.

Model	Params	Train time	FLOPs	Additional cost source
Backbones with and without PAN+PGS				
Allegro	1.10×	1.25×	1.40×	Local tensor products without iterative message passing
Allegro + PAN + PGS	1.13×	1.30×	1.46×	Lightweight scalar gating and spectral projections
NequIP	1.05×	1.10×	1.30×	Low-body-order equivariant convolutions
NequIP + PAN + PGS	1.08×	1.14×	1.36×	Lightweight scalar gating and spectral projections
MACE	1.00×	1.00×	1.00×	Standard SiLU–MLP scalar mixer
MACE + PAN + PGS	1.03×	1.06×	1.05×	Lightweight scalar gating and spectral projections
Alternative scalar mixers under the MACE backbone				
MACE + KAN	2.20×	1.95×	1.50–2.10×	Dense B-spline activation grids on all scalar channels
MACE + KAF	1.55×	1.40×	1.30–1.45×	Kernel activation functions on all scalar channels
MACE + LibraKAN	1.25×	1.20×	1.15–1.25×	Learned rational activations on all scalar channels

References

[1] Batatia, I., *et al.*: Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. Advances in Neural Information Processing Systems (NeurIPS) **35**, 11423–11436 (2022)

[2] Musaelian, A., *et al.*: Learning local equivariant representations for large-scale atomistic dynamics. Nat. Commun. **14**, 579 (2023)

[3] Batatia, I., Batzner, S., Kovács, D.P., Musaelian, A., Simm, G.N.C., Drautz, R., Ortner, C., Kozinsky, B., Csányi, G.: The design space of E(3)-equivariant atom-centred interatomic potentials. Nature Machine Intelligence **7**, 56–67 (2025) https://doi.org/10.1038/s42256-024-00956-x

[4] Batzner, S., *et al.*: E(3)-equivariant graph neural networks for data-efficient interatomic potentials. Nat. Commun. **13**, 2453 (2022)

[5] Gasteiger, J., Giri, S., Margraf, J.T., Günnemann, S.: Gemnet: Universal directional graph neural networks for molecules. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 6790–6802 (2021)

[6] Drautz, R.: Atomic cluster expansion for accurate and transferable interatomic potentials. Phys. Rev. B **99**, 014104 (2019)

[7] Christensen, A.S., Lilienfeld, O.A.: On the role of gradients for machine learning of molecular energies and forces. Mach. Learn.: Sci. Technol. **1**(4), 045018 (2020)

[8] Bartók, A.P., Payne, M.C., Kondor, R., Csányi, G.: Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. Phys. Rev. Lett. **104**(13), 136403 (2010)

[9] Smith, J.S., Isayev, O., Roitberg, A.E.: Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. Chem. Sci. **8**(4), 3192–3203 (2017)

[10] Schütt, K., Unke, O., Gastegger, M.: Equivariant message passing for the prediction of tensorial properties and molecular spectra. In: International Conference on Machine Learning (ICML), pp. 9377–9388 (2021). PMLR

[11] Klicpera, J., Groß, J., Günnemann, S.: Directional message passing on molecular graphs via synthetic coordinates. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 13313–13323 (2020)

[12] Haghighatlari, M., Li, J., Guan, X., Zhang, O., Das, A., Stein, C.J., Heidar-Zadeh, F., Liu, M., Head-Gordon, M., Bertels, L., Hao, H., Leven, I., Head-Gordon, T.: NewtonNet: A newtonian message passing network for deep learning of interatomic potentials and forces. Digital Discovery **1**, 333–343 (2022) https://doi.org/10.1039/D2DD00008C

[13] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: KAN: Kolmogorov–arnold networks. arXiv preprint arXiv:2404.19756 (2024) arXiv:2404.19756 [cs.LG]

[14] Scardapane, S., Van Vaerenbergh, S., Totaro, S., Uncini, A.: Kafnets: Kernel-based non-parametric activation functions for neural networks. Neural Networks **110**, 19–32 (2019) https://doi.org/10.1016/j.neunet.2018.11.002

[15] Barnett, A.H., Magland, J., Klinteberg, L.: A parallel nonuniform fast fourier transform library based on an "Exponential of Semicircle" kernel. SIAM Journal on Scientific Computing **41**(5), 479–504 (2019) https://doi.org/10.1137/18M120885X