STRAINPAIR: LEARNING ELASTIC RESPONSE IN PRETRAINED INTERATOMIC POTENTIALS

Anonymous authors
Paper under double-blind review

ABSTRACT

Elastic moduli measure how stress changes under deformation. Yet interatomic potentials are typically fine-tuned to match stresses one configuration at a time. We find that these two objectives can diverge: an adapted potential can predict elastic moduli more accurately even as its common stress-offset error increases. Motivated by this distinction, we introduce StrainPair, which augments conventional fine-tuning with a simple paired loss that matches stress differences between deformations of the same material. The loss uses existing labels and leaves the potential architecture and inference procedure unchanged. In our evaluation, StrainPair reduces shear- and bulk-modulus errors by 20.00% and 17.52% over matched pointwise fine-tuning across 10,987 materials. Three-seed experiments on 224 external materials confirm the gains. Stronger pointwise supervision, cross-material pairing, and affine output calibration do not recover the same improvements. Instead, the gains accompany a reduction in deformation-response error, despite larger common stress offsets. Our analysis explains this separation: stress differences cancel shared offsets and constrain variation along deformation paths. The results show that learning stress changes provides a useful target for adapting pretrained potentials to elasticity.

1 INTRODUCTION

Predicting how a crystal resists deformation requires more than an accurate stress at a single structure. Elastic moduli describe how stress changes as the crystal is strained; a standard computational procedure obtains them by applying small deformations, calculating stresses, and fitting an elastic tensor (de Jong et al., 2015). Pretrained interatomic potentials make this workflow accessible beyond direct first-principles calculations. Conservative potentials obtain forces and stresses from a learned energy, allowing the same model to support relaxation and response calculations (Rhodes et al., 2025). Strained-structure fine-tuning can further improve elastic predictions (Gao & Wang, 2026). But adding deformation data leaves an important choice open: should adaptation only match each configuration's stress, or should it also match the changes between configurations?

The distinction is easy to see in a stress–strain curve. A vertical shift changes its absolute values while leaving its slope unchanged. Conversely, small errors that vary with strain can distort the slope. At fixed configurations and within a common stress convention, these correspond to a shared offset and a response error, respectively. A pointwise loss penalizes both, but does not explicitly distinguish their roles in the response. Related concerns arise elsewhere in atomistic modeling: lower held-out prediction errors need not translate into better physical properties (Fu et al., 2025), and standard energy, force, and stress training can leave curvature errors that affect phonons (Koker et al., 2026). Systematic softening in several universal potentials further illustrates how biased potential-energy curvature affects physical predictions (Deng et al., 2025). For elasticity, this suggests examining how stress errors vary along a deformation path, alongside their absolute magnitude.

Our experiments expose this distinction in pretrained-potential adaptation. In our controlled study, physically paired supervision improves shear and bulk predictions while reducing deformation-response error, even though the common stress-offset error increases. Across three seeds, the mean response error decreases by approximately 22.9% relative to matched pointwise fine-tuning (Section 4.4). Here, the offset is the fitted intercept of the stress residual along a deformation path. Improved elasticity follows more accurate stress changes even when this intercept becomes less

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

accurate. Measuring response error therefore reveals a change that aggregate pointwise errors can miss.

We introduce *StrainPair*, a physically paired stress-supervision strategy for elastic adaptation. Two configurations of the same material along the same strain axis pass through a shared potential, and their predicted stress difference is matched to the difference of their existing reference labels. Native energy, force, and stress losses continue to anchor absolute predictions; the paired term constrains variation along deformation paths. Figure 1 summarizes the joint training procedure. No new labels, prediction head, or inference procedure is required. Learning derivatives and output differences has established precedents in Sobolev training and GAR (Czarnecki et al., 2017; Zhu et al., 2025). Here, the physical relationship between configurations gives the difference target a deformation-response interpretation. Our analysis makes that interpretation precise: common residual offsets cancel, and on an affine residual path the paired loss penalizes response mismatch in proportion to the strain span. This gives the paired term a direct role: it teaches the potential how stress should change as a material deforms.

We evaluate StrainPair using one fixed checkpoint per training condition on all 10,987 materials in the Matbench elasticity pool (Dunn et al., 2020), together with a three-seed comparison on 224 external materials. StrainPair reduces shear- and bulk-modulus MAE by 20.00% and 17.52%, respectively, relative to matched pointwise adaptation in the full-set evaluation. The external controls address three alternative explanations: stronger pointwise stress supervision, cross-material pair rewiring, and affine readout calibration. None reproduces the same gains. Combined with the response diagnostics, these comparisons support the value of supervising changes between physically related configurations. Section 4.1 describes the data separation and training protocols.

Our main finding is that elasticity can improve through more accurate stress changes, independently of the common stress offset. StrainPair turns this observation into a simple training objective using existing labels. Matched controls test the role of physical pairing, response diagnostics connect the objective to the measured gains, and application evaluations examine how the adapted potential performs in material screening. Together, these experiments make the case for supervising deformation response explicitly during elastic adaptation.

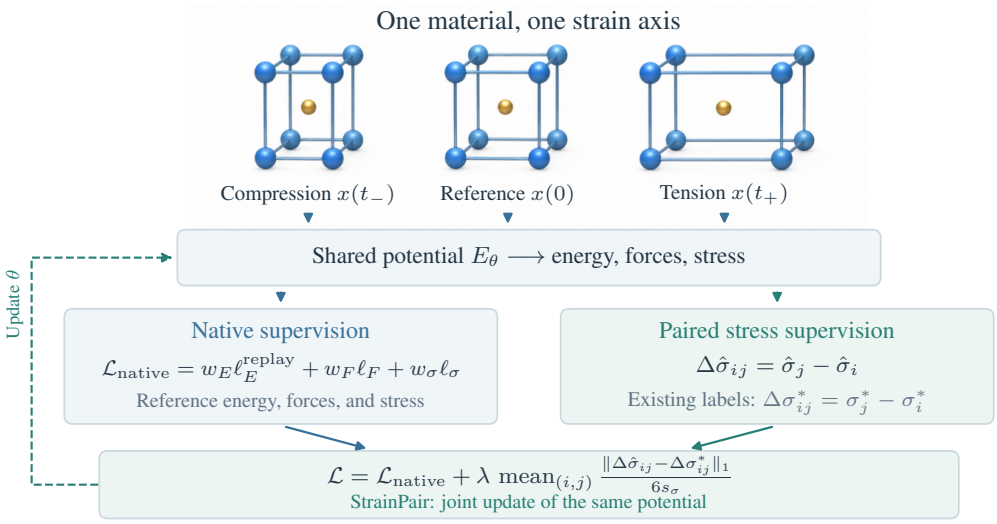

Figure 1: StrainPair: physically paired stress supervision for adapting a pretrained interatomic potential. Configurations of the same material along one strain axis share the potential and its energy-derived outputs. Native energy, force, and stress supervision is combined with an L_1 loss on predicted versus reference stress differences, averaged over eligible within-group pairs and normalized by the frozen pretrained stress scale s_σ. The relaxed reference is included when available in the batch. The dashed arrow denotes joint parameter updating. Crystal cells schematically represent deformations of the same material.

2 RELATED WORK

Relational supervision and physical pairing. Learning relationships between outputs can complement pointwise regression. GAR combines absolute prediction with pairwise label differences and normalized differences, relating these objectives to residual dispersion, correlation, and function derivatives (Zhu et al., 2025). GARM extends relational supervision to genetic perturbations through multiple decoders and configurable pair types (Zhu & Jerby, 2025). These works establish difference supervision as a general learning principle. StrainPair belongs to that family; its application-specific structure pairs deformations of the same material along the same strain axis within a shared conservative potential. We study whether these physical relationships improve elastic adaptation and distinguish common-offset from response error. Our controls examine the contribution of physical pairing relative to pointwise weighting and readout calibration.

Response supervision for atomistic potentials. Force matching and Sobolev training establish the value of supervising derivatives alongside function values (Ercolessi & Adams, 1994; Czarnecki et al., 2017), including stress and tangent constraints in constitutive modeling (Vlassis et al., 2022). Recent methods target potential-energy curvature: Hessian distillation transfers teacher curvature (Amin et al., 2025), PFT supervises crystal force constants using sampled Hessian columns (Koker et al., 2026), and PHL and HINT use stochastic Hessian projections (Rodriguez et al., 2026; Yin et al., 2026). Taylor augmentation uses derivative information to construct additional training labels (Cooper et al., 2020); UniAug and ModeAug extend this strategy using Hessians (Kwak & Jo, 2026). StrainPair uses differences of existing finite-strain stress labels, without a teacher or Hessian labels. These targets supervise stress variation over finite strain intervals. We retain potential-based elastic readout, whereas MatTen directly predicts elasticity tensors (Wen et al., 2024). CHIPS-FF evaluates downstream material properties of universal potentials (Wines & Choudhary, 2025), and targeted strained-data adaptation is studied by Gao & Wang (2026). The question here is what explicitly supervising within-material stress changes adds to matched adaptation on the same data.

Adaptation and capability retention. Pretrained potentials such as MACE-MP, MatterSim, and Orb-v3 can be specialized to downstream tasks (Batatia et al., 2025; Yang et al., 2024; Rhodes et al., 2025). Frozen transfer learning and LoRA-based Equitrain restrict parameter updates (Radova et al., 2025; Grandel et al., 2026); reEWC combines replay and parameter regularization (Kim et al., 2026), while PFT co-trains on upstream data during response adaptation (Koker et al., 2026). StrainPair retains native supervision and replay while adding a paired stress term. We evaluate held-out energy, force, and stress errors alongside elasticity to track how adaptation changes the potential's predictions on other configurations.

3 STRAINPAIR: PHYSICALLY PAIRED STRESS SUPERVISION

3.1 FROM CONFIGURATION ACCURACY TO DEFORMATION RESPONSE

StrainPair adapts a pretrained conservative potential E_{θ_0} by updating its parameters, retaining the architecture and its energy-derived force and Cauchy-stress outputs. Elastic properties are computed from the adapted potential by deforming and relaxing a material. Accurate stresses at individual configurations and accurate changes along deformation paths are related but distinct objectives.

A physical group $g = (m, k)$ consists of configurations $x_g(t_i)$ of the same material m along strain axis k, with amplitudes t_i and existing reference stresses $\boldsymbol{\sigma}_i^*$. We collect final deformation frames and, when present in the batch, the relaxed reference frame. All members share a rotation and a six-component Voigt convention. Predictions and labels refer to exactly the same configurations. Define the stress residual

$$\mathbf{r}_i = \boldsymbol{\sigma}_\theta(x_g(t_i)) - \boldsymbol{\sigma}_i^*. \tag{1}$$

Physical pairing supplies a direction of change with a common material identity. Arbitrary cross-material differences generally mix changes in composition, structure, and deformation; they do not isolate the same response.

3.2 JOINT POINTWISE AND PAIRED STRESS SUPERVISION

Let $\mathcal{E}_B$ be the multiset of all unordered within-group pairs in a batch. The implemented regularizer is

$$\mathcal{R}_{\text{pair}} = \frac{1}{|\mathcal{E}_B|} \sum_{(i,j)\in\mathcal{E}_B} \frac{\|[\boldsymbol{\sigma}_\theta(x_j) - \boldsymbol{\sigma}_\theta(x_i)] - [\boldsymbol{\sigma}_j^* - \boldsymbol{\sigma}_i^*]\|_1}{6s_\sigma} = \frac{1}{|\mathcal{E}_B|} \sum_{(i,j)\in\mathcal{E}_B} \frac{\|\mathbf{r}_j - \mathbf{r}_i\|_1}{6s_\sigma}. \quad (2)$$

Here $s_\sigma > 0$ is the frozen pretrained stress standard deviation. We average over all eligible pairs in the batch, so groups with more pairs contribute more terms. An empty pair set contributes zero. The loss matches stress differences directly, with no division by strain separation.

The adaptation objective is

$$\mathcal{L}_{\text{single}} = w_E \ell_E^{\text{replay}} + w_F \ell_F + w_\sigma \ell_\sigma, \quad (3)$$

$$\mathcal{L}_{\text{StrainPair}} = \mathcal{L}_{\text{single}} + \lambda \mathcal{R}_{\text{pair}}. \quad (4)$$

The native losses anchor absolute predictions; the added term penalizes mismatched stress changes. Energy anchoring uses replay frames, while force and stress supervision use applicable deformation and replay labels. The matched Single control uses the same pretrained initialization, data, rotations, update schedule, and native losses, with $\lambda = 0$. Our experiments use $(w_E, w_F, w_\sigma) = (10, 2, 5)$ and $\lambda = 5$.

3.3 THEORETICAL PROPERTIES AND PROOFS

Offset invariance and the nullspace. We analyze the implemented L_1 objective at fixed configurations. Write $\mathcal{E}_g$ for a group's edge set and $\mathcal{R}_g$ for its mean normalized pair penalty. For any common vector $\mathbf{b}$, $(\mathbf{r}_j + \mathbf{b}) - (\mathbf{r}_i + \mathbf{b}) = \mathbf{r}_j - \mathbf{r}_i$. Hence the pair loss cannot identify a common offset. If the graph with edges $\mathcal{E}_g$ is connected, zero loss implies $\mathbf{r}_i = \mathbf{r}_j$ on every edge and therefore on every path; all residuals are equal. The converse is immediate. For a disconnected graph there is one unconstrained constant per connected component. These statements also explain why native pointwise supervision is necessary.

Proposition 1 (Response interpretation on an affine residual path) *Suppose that, on the sampled configurations of a group, $\mathbf{r}_i = \mathbf{b}_g + t_i \mathbf{a}_g$. Then its mean paired penalty equals*

$$\mathcal{R}_g = \frac{\overline{\Delta t}_g}{6s_\sigma} \|\mathbf{a}_g\|_1, \qquad \overline{\Delta t}_g = \frac{1}{|\mathcal{E}_g|} \sum_{(i,j)\in\mathcal{E}_g} |t_j - t_i|. \quad (5)$$

Proof. For an affine residual $\mathbf{r}_i = \mathbf{b} + t_i \mathbf{a}$, every edge has $\|\mathbf{r}_j - \mathbf{r}_i\|_1 = |t_j - t_i| \|\mathbf{a}\|_1$, proving Proposition 1. At a fixed response mismatch, pairs with a larger strain span contribute more to the loss. □

A conditional bound on fitted response Assume that a group contains n frames with $Q = \sum_i (t_i - \bar{t})^2 > 0$ and all unordered pairs are included, as in the implementation. The error between least-squares stress slopes with free intercepts is

$$\widehat{\mathbf{a}} = \frac{\sum_i (t_i - \bar{t})\mathbf{r}_i}{Q} = \frac{\sum_{i<j} (t_j - t_i)(\mathbf{r}_j - \mathbf{r}_i)}{\sum_{i<j} (t_j - t_i)^2}. \quad (6)$$

To verify the second equality, expand the pair sums: the numerator is $n \sum_i (t_i - \bar{t})\mathbf{r}_i$ and the denominator is nQ. Let $M = \max_{i<j} |t_j - t_i|$ and $D = \sum_{i<j} (t_j - t_i)^2$. The triangle inequality gives

$$\|\widehat{\mathbf{a}}\|_1 \le \frac{M}{D} \sum_{i<j} \|\mathbf{r}_j - \mathbf{r}_i\|_1 = \frac{6s_\sigma |\mathcal{E}_g| M}{D} \mathcal{R}_g. \quad (7)$$

For a fixed group with nonzero strain span, this bounds the fitted response error by its paired loss. The coefficient depends on the sampled amplitudes, so the bound applies group by group.

Relation to known squared-pair identities For comparison only, the squared Euclidean surrogate satisfies

$$\sum_{i<j} \|\mathbf{r}_j - \mathbf{r}_i\|_2^2 = n \sum_i \|\mathbf{r}_i - \bar{\mathbf{r}}\|_2^2. \tag{8}$$

Expanding both sides proves the identity. Residual-dispersion interpretations of difference supervision are established in GAR (Zhu et al., 2025). It describes a squared-loss surrogate; StrainPair uses the componentwise L_1 objective in Equation 2.

Finite intervals and downstream scope On a differentiable shared path, $\mathbf{r}(t_j) - \mathbf{r}(t_i) = \int_{t_i}^{t_j} \mathbf{r}'(t)\, dt$. Pair supervision therefore constrains the accumulated response between sampled configurations. Interpreting it as an infinitesimal tangent requires smoothness and sufficiently dense sampling. Training uses Cauchy stresses on the labeled configurations; elastic evaluation converts them to second-Piola stresses, $S = JF^{-1}\sigma F^{-\mathsf{T}}$, after model-dependent relaxation. Since both the stress transformation and the relaxed geometries enter the readout, we evaluate downstream moduli directly alongside the fixed-geometry response diagnostic.

3.4 Implementation and elastic readout

Stress-difference targets are computed from existing stress labels. The additional loss requires no new labels, teacher, prediction head, or coordinate-derivative order beyond native stress training. Small within-batch groups are explicitly enumerated. Pretrained normalization remains frozen and confidence loss is disabled; paired structures share a rotation so their stress components are comparable. The shared rotation keeps paired stress components in the same coordinate frame; the componentwise L_1 metric itself depends on that frame.

At inference, one adapted potential processes each structure independently. Relaxation, finite deformation, conversion to second-Piola stress, and elastic-tensor fitting yield the shear and bulk moduli. Physical pair groups are needed only during training. We assess the adapted potential through elastic accuracy, fixed-geometry response errors, held-out force and stress accuracy, and candidate-screening diagnostics.

4 Experiments

4.1 Experimental setup

Evaluation design. We ask whether supervising stress changes improves elastic prediction beyond matched pointwise adaptation, and whether the improvement accompanies better deformation response. One fixed potential per arm is evaluated on 10,987 Matbench materials; a separate 224-material set supports three-seed replication and controlled comparisons. Fixed-geometry stress diagnostics examine the response interpretation, while retention and screening evaluations test the extent of the benefit. Throughout, *Single* denotes native pointwise fine-tuning and *StrainPair* adds the absolute stress-difference loss.

Data and separation. The adaptation pool is constructed from successful Materials Project elasticity records (Jain et al., 2013; de Jong et al., 2015) after excluding all matched Matbench material identifiers and all reduced compositions represented in the 10,987-material Matbench elasticity pool. Deterministic composition-grouped partitioning divides the remaining 1,624 materials into 1,000 fit, 200 tune, 200 calibration, and 224 external-test materials. All frames of a material remain in its partition. The fit set contains 105,493 strain-source frames and is supplemented with 40,000 audited replay frames from the Materials Project Trajectory (MPtrj) dataset (Deng et al., 2023). An additional 1,500-frame retention set is excluded from training. Replay selection excludes mapped reserved materials and compositions; unknown mappings are excluded.

The tune split supplies the predefined execution gate, and the calibration split is used only for affine readout fitting. The complete Matbench pool is evaluated with one fixed checkpoint per arm. This fixed-checkpoint protocol evaluates the same adapted weights across materials, rather than training a model for each fold. The exclusion audit applies to adaptation data; foundation-model pretraining exposure is unknown. Historical development included Matbench-derived materials and inspection

of the external test, so we use these evaluations for large-scale comparison, matched replication, and mechanism analysis.

Matched adaptation. The primary experiments start from the same Orb-v3 conservative MPA checkpoint (Rhodes et al., 2025). Single and StrainPair use the same training data, batch schedule, initialization seed, and 5,000 optimizer updates. Native energy, force, and stress weights are 10, 2, and 5; the energy term anchors replay frames. StrainPair adds the absolute error between predicted and reference stress differences with weight 5, using the pretrained stress normalization. Paired configurations belong to the same material and strain axis. Group batches occur with probability 0.5 and share a rotation; ordinary frames receive independent rotations. Batches contain at most 192 atoms, with ordinary batches capped at eight structures. Target normalizers are frozen and the confidence loss is disabled. Native Adam/OneCycle training uses an initial learning rate of 1.5×10^{-6}, a peak of 1.5×10^{-5}, and gradient clipping at 0.5. We evaluate final saved weights.

Physical readout and scoring. Each potential is evaluated by cell and ionic relaxation, positive and negative Green–Lagrange strains of 0.5% and 1%, and internally relaxed stress calculations. Fitting second-Piola stress yields the elastic tensor (de Jong et al., 2015), from which Voigt–Reuss–Hill shear and bulk moduli are computed (Hill, 1952). The force tolerance is 10^{-3} eV/Å, with at most 200 base-relaxation and 100 per-strain ionic steps. The primary endpoint is shear-modulus MAE in $\log_{10}(\mathrm{GPa})$; bulk MAE is also reported. Predictions failing the prescribed physical validity checks remain in the denominator and receive a target-median fallback computed from the fit split. Validity is reported separately.

The full-set and application comparisons use the fixed seed-0 models. External controls use three training seeds and report mean and sample standard deviation. Paired bootstrap intervals resample materials jointly across those fixed seeds; they do not estimate population uncertainty over training seeds. The 50 orbital-field matrix (OFM) groups are post-hoc strata of the same full-set predictions; OFM represents local environments using valence-orbital information (Pham et al., 2017).

4.2 Physical pairs improve elasticity across the evaluated materials

StrainPair improves both moduli, whereas pointwise adaptation alone produces mixed gains (Table 1). Single lowers shear error relative to the pretrained potential but increases bulk error from 0.05509 to 0.06033. StrainPair reaches 0.06356 for shear and 0.04977 for bulk, reducing errors by 20.00% and 17.52% relative to Single, and by 27.03% and 9.67% relative to the pretrained model. Thus the paired term improves an already-improved shear prediction and reverses the bulk degradation. Because Single and StrainPair see the same deformation data, this comparison supports the value of organizing existing labels into response targets.

Table 1: Physical pairing improves both elastic endpoints in the fixed-checkpoint evaluation on 10,987 Matbench materials. MAE is in $\log_{10}(\mathrm{GPa})$, including the predefined fallback for invalid predictions. Adapted rows each use one seed-0 checkpoint, not five-fold training averages.

Model	G MAE ↓	K MAE ↓	Valid ↑
Pretrained	0.08709	0.05509	10,650
Single	0.07944	0.06033	10,699
StrainPair	**0.06356**	**0.04977**	**10,783**

The aggregate improvement is accompanied by broader validity and lower errors across material groups. StrainPair returns valid estimates for 84 more materials than Single, so the all-case MAE reflects both accuracy and validity. Stratifying the same predictions into 50 OFM groups, StrainPair improves shear in 49 groups and bulk in all 50 (Figure 2). The unweighted mean of group MAEs falls from 0.08454 to 0.06615 for shear and from 0.05543 to 0.04252 for bulk; the maximum group errors also decrease. The gains therefore occur across most evaluated groups rather than only in the aggregate. The OFM analysis groups the fixed full-set predictions after evaluation; each group uses the same checkpoint.

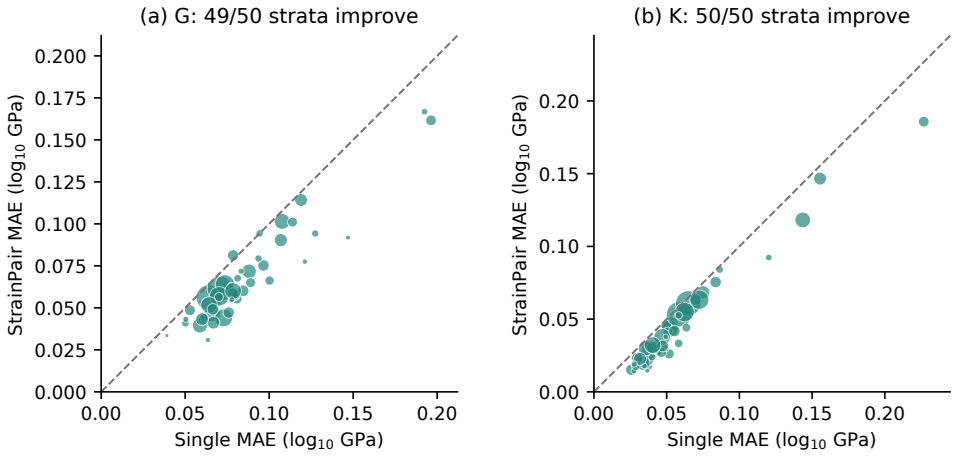

Figure 2: The benefit extends across most material strata. Each point compares Single and StrainPair on one of 50 OFM groups; points below the diagonal favor StrainPair and marker area reflects group size. Both panels reuse the fixed full-set checkpoints, without group-specific retraining.

4.3 Stronger pointwise supervision and calibration do not recover the gains

The gains repeat across seeds. The elastic advantage repeats across all three seeds on the 224 external materials (Table 2). Mean shear and bulk errors decrease by 13.55% and 13.97% relative to Single. StrainPair-minus-Single material-bootstrap intervals are $[-0.04150, -0.01614]$ for shear and $[-0.02804, -0.00747]$ for bulk. This replication motivates a more specific question: does StrainPair help because it strengthens stress supervision, changes the final readout, or uses physically related configurations?

Table 2: The tested alternative explanations do not match physical pairing. External MAE on 224 materials is reported as mean $\pm$ sample standard deviation over three seeds, in $\log_{10}(\mathrm{GPa})$. All methods retain invalid cases. Affine calibration uses only the separate 200-material calibration split.

Method	G MAE $\downarrow$	K MAE $\downarrow$
Single	0.20826 ± 0.00891	0.11789 ± 0.01005
Stronger single	0.20824 ± 0.00190	0.12302 ± 0.00473
Rewired pairs	0.21347 ± 0.00173	0.12198 ± 0.00256
Single + affine	0.19400 ± 0.00926	0.12772 ± 0.00960
StrainPair	0.18005 ± 0.00407	0.10142 ± 0.00430

Increasing pointwise stress weight is insufficient. Stronger single scales the stress term using a fixed training-only calibration of initial gradients on paired batches. Its mean shear error is almost unchanged from Single (0.20824 versus 0.20826), and its bulk error is higher. StrainPair instead improves both endpoints. The tested increase in pointwise stress weight therefore does not explain StrainPair's gains.

The relationship between paired configurations matters. Rewired pairs use cross-material edges within the same batches whenever available, retaining physical edges otherwise. Their errors exceed Single's on both endpoints. Merely adding a difference target therefore does not reproduce the gain in this control: the physical organization of the pairs matters. The control changes which structures are compared, including their locality and label-difference distribution.

A global readout correction is insufficient. Single + affine fits a slope and intercept for each log modulus on separate calibration data. It improves shear MAE to 0.19400 but remains above StrainPair's 0.18005, while bulk error rises to 0.12772. The advantage is therefore not recovered

by the tested global output correction. Together, these controls motivate examining the learned deformation response directly.

4.4 Better response accompanies larger common stress offsets

StrainPair improves elasticity even though its shared stress-offset error increases. To expose this separation, we evaluate each model on the same final DFT geometries and fit stress residuals against strain within each material–axis group. The fitted intercept measures common-offset error, and the fitted slope measures response error. Across three seeds, StrainPair reduces the mean recorded material-level response error from 0.05015 to 0.03866 GPa per 1% strain, a 22.9% reduction, while the intercept magnitude rises from 0.12962 to 0.15310 GPa (Figure 3a–b).

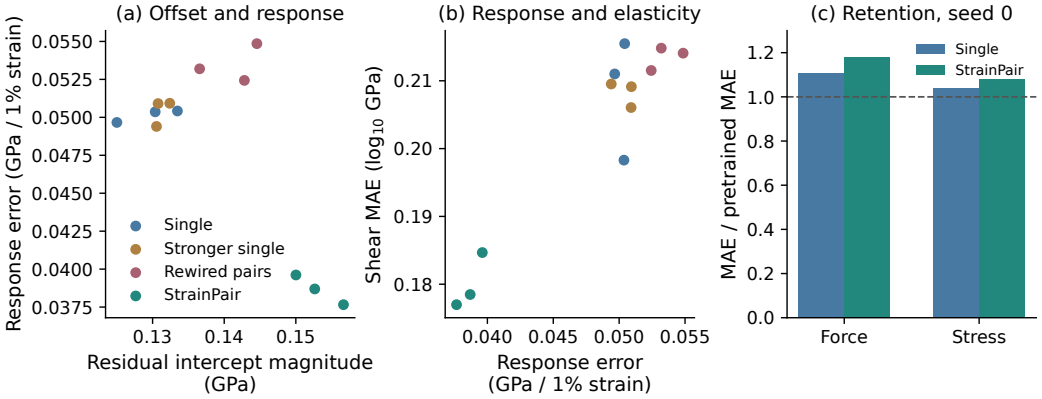

Figure 3: Elastic response, shared offsets, and retained accuracy change differently. (a–b) Each marker represents one training seed evaluated on the same 224 materials. Fixed-geometry Cauchy-stress diagnostics separate common-offset and response errors; the scatter shows association, not a causal fit. (c) On 1,500 held-out frames, seed-0 force and stress errors increase relative to the pretrained model, whose MAE ratio is one.

This is the empirical distinction motivating paired adaptation. A shared offset cancels in a within-path stress difference, allowing better changes to coexist with worse absolute alignment. The joint decrease in response and modulus errors is consistent with that interpretation and explains why offset accuracy alone is an incomplete diagnostic for elastic adaptation. The intercept describes the shared residual along each path, rather than its full pointwise stress error. The response diagnostic uses Cauchy stresses at fixed DFT geometries, while modulus evaluation uses second-Piola stresses after model-dependent relaxation. Agreement between these two evaluations supports the response interpretation: StrainPair improves the stress changes relevant to elasticity, and the tested pointwise weighting and output calibration do not recover the same gains.

4.5 Held-out accuracy and material screening

Held-out force and stress accuracy. The reduction in elastic error does not extend uniformly to other configurations. On the reserved 1,500-frame set, seed-0 StrainPair raises force and stress MAE by 17.76% and 7.84% relative to the pretrained model; Single raises them by 10.73% and 3.97% (Figure 3c). Both arms remain within the predefined 1.2 ratio gate. These measurements track force and stress accuracy separately from the elastic endpoints during adaptation.

Retrieval of high-modulus materials. We test the ranking consequence by selecting materials with the largest predicted moduli and counting known-label threshold exceedances (Table 3a). At 1,099 queries, StrainPair finds 927 materials with $G \geq 100$ GPa, compared with 919 for Single and 894 for the pretrained potential. Relative to Single, this is eight additional hits and a precision increase from 83.62% to 84.35%. The smallest shear budget is already saturated. Bulk retrieval does not improve: StrainPair ties Single at 110 and 1,099 queries and has one fewer hit at 549. The measured retrieval gain is concentrated in shear selection at the larger query budget. The distinction between regression error and selection performance is also central to Matbench Discovery, which

evaluates thermodynamic-stability screening rather than the high-modulus ranking task considered here (Riebesell et al., 2025).

Table 3: High-modulus screening and generated-candidate assessment. (a) Known-label hits at all recorded query budgets for $G \geq 100$ GPa and $K \geq 150$ GPa. (b) Predictions and diagnostics on the same 400 Chemeleon2 candidates. Born-stable counts are model predictions, not DFT-confirmed successes. Continuous diagnostics are medians, with smaller deviations preferred.

(a) Known-label screening

Target	Queries	Pretrained	Single	StrainPair
G	110	110	110	110
G	549	542	544	545
G	1,099	894	919	927
K	110	109	110	110
K	549	546	546	545
K	1,099	1088	1089	1089

(b) Generated-candidate assessment

Metric	Pretrained	Single	StrainPair
Valid elastic estimates	386	388	389
Predicted Born-stable	393	396	395
Major-symmetry error (GPa)	0.17654	0.16835	0.16788
Rotation-force deviation (eV/Å)	0.00394	0.00587	0.00912
Rotation-stress deviation (GPa)	0.00963	0.02420	0.04378

Assessment of generated candidates. On a frozen pool of 400 Chemeleon2 candidates (Park & Walsh, 2026), StrainPair supplies 389 valid estimates, versus 388 for Single and 386 for the pretrained model. Single and StrainPair share 19 of their top 20 candidates for both shear and bulk, indicating that adaptation changes these selections only slightly. StrainPair marginally reduces tensor major-symmetry error, but increases rotational force and stress deviations and predicts one fewer Born-stable candidate than Single (Table 3b). These diagnostics describe different aspects of the generated structures: elastic validity, tensor symmetry, rotation consistency, and predicted local mechanical stability (Mouhat & Coudert, 2014). They are model-based assessments without DFT confirmation of the selected candidates. The Single/StrainPair comparison uses matching recorded protocol hashes; archived pretrained scores serve as a reference and lack those hashes.

5 Conclusion

We introduced StrainPair, a simple fine-tuning strategy that learns stress changes between physically paired deformations using existing labels. StrainPair improves elastic predictions over matched pointwise adaptation, with gains supported by three-seed replication and controlled comparisons. The central finding is that response error can decrease even as common stress-offset error increases: better elasticity does not require every component of stress error to improve. This distinction explains the role of paired supervision, which constrains changes along deformation paths while native losses anchor absolute predictions. These results support learning deformation response explicitly when adapting pretrained interatomic potentials for elasticity.

AI USE STATEMENT

Generative AI tools were used to assist with drafting parts of the manuscript, polishing and improving the clarity of the writing, retrieving and discovering relevant literature, and generating or refining code. All AI-generated or AI-assisted text, references, and code were carefully reviewed, verified, and, when necessary, revised by the authors before inclusion in the paper or implementation. The authors take full responsibility for the final manuscript, references, experimental code, and reported results.

REFERENCES

Ishan Amin, Sanjeev Raja, and Aditi S. Krishnapriyan. Towards Fast, Specialized Machine Learning Force Fields: Distilling Foundation Models via Energy Hessians. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2501.09009.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, et al. A foundation model for atomistic materials chemistry. *The Journal of Chemical Physics*, 163: 184110, 2025. doi: 10.1063/5.0297006. URL https://doi.org/10.1063/5.0297006.

April M. Cooper, Johannes Kästner, Alexander Urban, and Nongnuch Artrith. Efficient training of ANN potentials by including atomic forces via Taylor expansion and application to water and a transition-metal oxide. *npj Computational Materials*, 6:54, 2020. doi: 10.1038/s41524-020-0323-8. URL https://doi.org/10.1038/s41524-020-0323-8.

Wojciech Marian Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Świrszcz, and Razvan Pascanu. Sobolev Training for Neural Networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.04859.

Maarten de Jong, Wei Chen, Thomas Angsten, Anubhav Jain, Randy Notestine, Anthony Gamst, et al. Charting the complete elastic properties of inorganic crystalline compounds. *Scientific Data*, 2:150009, 2015. doi: 10.1038/sdata.2015.9. URL https://doi.org/10.1038/sdata.2015.9.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5:1031–1041, 2023. doi: 10.1038/s42256-023-00716-3.

Bowen Deng, Yunyeong Choi, Peichen Zhong, Janosh Riebesell, Shashwat Anand, Zhuohan Li, KyuJung Jun, Kristin A. Persson, and Gerbrand Ceder. Systematic softening in universal machine learning interatomic potentials. *npj Computational Materials*, 11:9, 2025. doi: 10.1038/s41524-024-01500-6. URL https://doi.org/10.1038/s41524-024-01500-6.

Alexander Dunn, Qi Wang, Alex Ganose, Daniel Dopp, and Anubhav Jain. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. *npj Computational Materials*, 6:138, 2020. doi: 10.1038/s41524-020-00406-3.

F Ercolessi and J. B Adams. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhysics Letters (EPL)*, 26:583–588, 1994. doi: 10.1209/0295-5075/26/8/005. URL https://doi.org/10.1209/0295-5075/26/8/005.

Xiang Fu, Brandon M. Wood, Luis Barroso-Luque, Daniel S. Levine, Meng Gao, Misko Dzamba, and C. Lawrence Zitnick. Learning Smooth and Expressive Interatomic Potentials for Physical Property Prediction. arXiv preprint arXiv:2502.12147v2, 2025. URL https://arxiv.org/abs/2502.12147v2.

Pengfei Gao and Haidi Wang. Benchmarking universal machine learning interatomic potentials for elastic property prediction. arXiv preprint arXiv:2510.22999v3, 2026. URL https://arxiv.org/abs/2510.22999v3. Version 3, revised March 2026.

Jonas Grandel, Philipp Benner, and Janine George. Parameter-Efficient Fine-Tuning of Machine-Learning Interatomic Potentials for Phonon and Thermal Properties. arXiv preprint arXiv:2604.01017v2, 2026. URL https://arxiv.org/abs/2604.01017v2.

R. Hill. The elastic behaviour of a crystalline aggregate. *Proceedings of the Physical Society. Section A*, 65(5):349–354, 1952. doi: 10.1088/0370-1298/65/5/307. URL https://doi.org/10.1088/0370-1298/65/5/307.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1:011002, 2013. doi: 10.1063/1.4812323.

Jisu Kim, Jiho Lee, Sangmin Oh, Yutack Park, Seungwoo Hwang, Seungwu Han, Sungwoo Kang, and Youngho Kang. An efficient forgetting-aware fine-tuning framework for pretrained universal machine-learning interatomic potentials. *npj Computational Materials*, 12:26, 2026. doi: 10.1038/s41524-025-01895-w. URL https://doi.org/10.1038/s41524-025-01895-w.

Teddy Koker, Abhijeet Gangan, Mit Kotak, Jaime Marian, and Tess Smidt. PFT: Phonon Fine-tuning for Machine Learned Interatomic Potentials. arXiv preprint arXiv:2601.07742v4, 2026. URL https://arxiv.org/abs/2601.07742v4.

Bumju Kwak and Jeonghee Jo. Hessian-based molecular conformation augmentation for a scalable and efficient strategy of machine learning interatomic potentials. arXiv preprint arXiv:2609.05233v2, 2026. URL https://arxiv.org/abs/2609.05233v2.

Félix Mouhat and François-Xavier Coudert. Necessary and sufficient elastic stability conditions in various crystal systems. *Physical Review B*, 90(22):224104, 2014. doi: 10.1103/PhysRevB.90.224104. URL https://doi.org/10.1103/PhysRevB.90.224104.

Hyunsoo Park and Aron Walsh. Guiding generative models to uncover diverse and novel crystals via reinforcement learning. *Nature Machine Intelligence*, 8:1087–1099, 2026. doi: 10.1038/s42256-026-01262-4.

Tien Lam Pham, Hiori Kino, Kiyoyuki Terakura, Takashi Miyake, Koji Tsuda, Ichigaku Takigawa, and Hieu Chi Dam. Machine learning reveals orbital interaction in materials. *Science and Technology of Advanced Materials*, 18(1):756–765, 2017. doi: 10.1080/14686996.2017.1378060. URL https://doi.org/10.1080/14686996.2017.1378060.

Mariia Radova, Wojciech G. Stark, Connor S. Allen, Reinhard J. Maurer, and Albert P. Bartók. Fine-tuning foundation models of materials interatomic potentials with frozen transfer learning. *npj Computational Materials*, 11:237, 2025. doi: 10.1038/s41524-025-01727-x. URL https://doi.org/10.1038/s41524-025-01727-x.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: Atomistic simulation at scale. arXiv preprint arXiv:2504.06231v2, 2025. URL https://arxiv.org/abs/2504.06231v2.

Janosh Riebesell, Rhys E. A. Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Gerbrand Ceder, Mark Asta, Alpha A. Lee, Anubhav Jain, and Kristin A. Persson. A framework to evaluate machine learning crystal stability predictions. *Nature Machine Intelligence*, 7:836–847, 2025. doi: 10.1038/s42256-025-01055-1. URL https://doi.org/10.1038/s42256-025-01055-1.

Austin Rodriguez, Justin S. Smith, Sakib Matin, Nicholas Lubbers, Kipton Barros, and Jose L. Mendoza-Cortes. Projected Hessian Learning: Fast Curvature Supervision for Accurate Machine-Learning Interatomic Potentials. arXiv preprint arXiv:2603.04523v1, 2026. URL https://arxiv.org/abs/2603.04523v1.

Nikolaos N. Vlassis, Puhan Zhao, Ran Ma, Tommy Sewell, and WaiChing Sun. Molecular dynamics inferred transfer learning models for finite-strain hyperelasticity of monoclinic crystals: Sobolev training and validations against physical constraints. *International Journal for Numerical Methods in Engineering*, 123:3922–3949, 2022. doi: 10.1002/nme.6992. URL https://doi.org/10.1002/nme.6992.

Mingjian Wen, Matthew K. Horton, Jason M. Munro, Patrick Huck, and Kristin A. Persson. An equivariant graph neural network for the elasticity tensors of all seven crystal systems. *Digital Discovery*, 3:869–882, 2024. doi: 10.1039/D3DD00233K. URL https://doi.org/10.1039/D3DD00233K.

Daniel Wines and Kamal Choudhary. CHIPS-FF: Evaluating Universal Machine Learning Force Fields for Material Properties. *ACS Materials Letters*, 7:2105–2114, 2025. doi: 10.1021/acsmaterialslett.5c00093. URL https://doi.org/10.1021/acsmaterialslett.5c00093.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, et al. MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures. arXiv preprint arXiv:2405.04967v2, 2024. URL https://arxiv.org/abs/2405.04967v2.

Bangchen Yin, Jian Ouyang, Zhen Fan, Kailai Lin, Hanshi Hu, Dingshun Lv, Weiluo Ren, Hai Xiao, Ji Chen, and Changsu Cao. Hessian-informed machine learning interatomic potential towards bridging theory and experiments. arXiv preprint arXiv:2603.25373v1, 2026. URL https://arxiv.org/abs/2603.25373v1.

Dixian Zhu and Livnat Jerby. Gradient-aware modeling advances AI-driven prediction of genetic perturbation effects. bioRxiv preprint, 2025. URL https://doi.org/10.1101/2025.10.03.680360.

Dixian Zhu, Tianbao Yang, and Livnat Jerby. Gradient Aligned Regression via Pairwise Losses. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 80259–80281. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhu25x.html.