FAME: FACTORIZED ASTRONOMICAL MULTIMODAL EMBEDDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Stellar astrophysics observes the same stars through different instruments, across surveys with vastly different noise, resolution, and coverage, and some key properties, like stellar age, are not measured by any single one of them. Building a foundation model that combines these modalities therefore remains an open challenge. We present FAME, a foundation model for stellar astrophysics that combines seven modalities from six large-scale surveys via pretrained modality-specific encoders and a small alignment module. FAME departs from classical multimodal alignment in two ways. Its latent space is factorized into a shared and a residual subspace, so that aligning the modalities preserves what each one holds alone. And the shared subspace is trained with a joint-embedding objective extended by a graph-Laplacian penalty over catalogued physical coordinates. Under frozen linear probing, FAME outperforms unaligned features, a single-latent alignment baseline, and three existing astrophysics foundation models. Against fully supervised models it is more label-efficient on the labels that enter no pretraining objective and is beaten only on the one label read directly from a single observational capability; when the surveys carrying such a label are withheld, FAME still recovers it ($R^2 = 0.46$ and 0.59), where a supervised model trained under the same schedule retains little (0.07 and 0.19). Rare stellar populations are linearly decodable from the residual subspace, and the representation supports cross-modal spectral generation. Neither ingredient is specific to astronomy: the factorization applies wherever modalities carry private information, and the smoothness penalty wherever samples come with catalogued continuous coordinates. To our knowledge, FAME is the first foundation model for stellar astrophysics to combine time series, spectra and astrometry at scale.

1 INTRODUCTION

Foundation models have transformed many domains like natural language and computer vision (Devlin et al., 2018; Brown et al., 2020; Dosovitskiy et al., 2021; Radford et al., 2021), and recently, they have also started to appear in various scientific domains such as biology, chemistry, earth science, and particle physics (Lin et al., 2023; Candido et al., 2026; Batatia et al., 2025; Price et al., 2025; Bhimji et al., 2026). Within the scientific landscape, astrophysics is a promising domain, with large volumes of data spanning diverse and complex modalities. Indeed, there is a growing line of work in the field of multimodal learning for astrophysics (Zhang et al., 2024; Rizhko and Bloom, 2025; Parker et al., 2024; Kamai et al., 2025; Parker et al., 2025). However, many of these works rely primarily on contrastive objectives, autoregressive models, or limited modalities. Specifically, combining all modality types, including time-domain surveys, and across various noise levels, in a unified and optimal model, is still an open problem in astrophysics.

Here we present FAME (Factorized Astronomical Multimodal Embeddings), a foundation model for stellar astrophysics, combining seven modalities from six large-scale surveys, including time series, spectra, and astrometry. As such, FAME is the first foundation model in astrophysics to combine time-domain information from multiple sources on a large scale.

FAME is a two-step model, with modality-specific encoders, followed by a small alignment module. This has two advantages: first, it allows modality-specific inductive biases to be imposed in the first step, without interfering with other modalities. Second, the small alignment module allows easy

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.

fine-tuning for specific tasks. Motivated by failure modes of classical multimodal objectives, such as contrastive and reconstruction-based methods, FAME's latent space is factorized into shared and modality-specific subspaces, allowing for both types of information flow through the model, without losing any of them.

We further augment the alignment objective with a physics-informed smoothness prior. Alignment objectives already contain a Laplacian: VICReg's invariance term (Bardes et al., 2021) is a Dirichlet energy over the graph of positive pairs, whose minimizers under the variance and covariance constraints are Laplacian eigenmaps (Balestriero and LeCun, 2022). In multimodal alignment that graph is a disjoint union of per-star cliques: it constrains agreement between modalities of the same star and says nothing about how distinct stars relate. We therefore keep the penalty and add a second graph. $\mathcal{L}_{\mathrm{smooth}}$ draws its edges not from pairing but from distance in a catalogued physical coordinate space, $(T_{\mathrm{eff}}, \log g)$, so the geometry it induces follows from the loss rather than being discovered after the fact (Section 5.1). It improves downstream probing on every label we test. Pretraining is hybrid: survey-derived labels are injected alongside the self-supervised objectives (Section 4), so well-measured quantities are reused rather than rediscovered. In summary, our concrete contributions are:

- A general factorization of the latent space into shared and residual streams, which aligns the modalities while preserving what each one holds alone. Importantly, this is not unique to astrophysics.
- A new alignment objective combining VICReg with a graph-Laplacian smoothness term defined in physical-parameter space. Because its continuum limit selects Laplacian eigenfunctions, the geometry it induces is predictable from the loss rather than empirical. We show that the resulting shared space carries the Toeplitz and sinusoidal-mode structure of a representation that is approximately translation-equivariant in the stellar parameters, and improves downstream performance. This is a general extension of the VICReg objective to domains with known continuous parameters.
- Evidence that FAME outperforms three existing astrophysics foundation models under frozen linear probing, with controlled ablations tying the gains to our novel design: naive alignment can be *destructive*, falling below the unaligned pretrained features on most labels, and the factorization recovers what it discards. Against a fully supervised model, FAME is more label-efficient on the labels absent from pretraining. Where a single observational capability defines the label, it recovers that label once the surveys carrying the capability are withheld, which a supervised model trained under the same schedule fails to do (Section 5.3).

2 Related work

Multimodal alignment in astronomy. Two general families dominate how heterogeneous modalities get mapped into a shared representation. The first is joint-embedding alignment: paired observations of the same underlying object are pulled together in embedding space, either with explicit negatives (as in CLIP (Radford et al., 2021)) or, more recently, without them, using variance and covariance regularizers to prevent the representation from collapsing (as in VICReg (Bardes et al., 2021)). The second is generative, sequence-modeling alignment: each modality is discretized into tokens from a (shared or per-modality) vocabulary, and a single transformer is trained with a masked- or next-token objective over the concatenated multimodal sequence, treating alignment as a byproduct of predicting one part of the sequence from another. Existing multimodal foundation models in astronomy fall cleanly into these two categories. AstroCLIP (Parker et al., 2024), AstroM3 (Rizhko and Bloom, 2025), and MAVEN (Zhang et al., 2024) are explicitly contrastive, modeled directly on CLIP. DESA (Kamai et al., 2025) aligns *Kepler* and LAMOST with a non-contrastive joint-embedding objective, without explicit negatives. AION-1 (Parker et al., 2025), the most direct comparison point for this work, instead tokenizes its modalities across five surveys and trains a transformer with a masked-modeling objective over the resulting token sequences. Similarly, Leung and Bovy (2023) train a transformer on *Gaia* XP with APOGEE labels as a foundation model for stars.

Spectral structure in joint-embedding objectives. Joint-embedding methods are spectral methods in disguise: VICReg (Bardes et al., 2021), SimCLR (Chen et al., 2020) and Barlow Twins (Zbontar et al., 2021) recover Laplacian eigenmaps (Belkin and Niyogi, 2003) and multidimensional scaling respectively, once the positive-pair relation is read as a graph (Balestriero and LeCun, 2022), and the same correspondence underlies deep spectral clustering (Shaham et al., 2018). The geometry such an objective produces is a property of that graph. In image self-supervision the augmentation distributions of distinct samples overlap, so the graph is connected across samples and the embedding inherits data geometry; improving the edge set directly is an active lever (Simai et al., 2025). Alignment by exact cross-match has no such connectivity, since positive pairs join modalities of the same object and nothing else. To our knowledge the consequences of this for the resulting geometry have not been examined.

Geometry-imposing regularization. Graph-Laplacian penalties are the standard tool for imposing continuity on a representation — manifold regularization over a neighbour graph (Belkin et al., 2006), whose continuum limit is the density-weighted Laplacian of diffusion maps (Coifman and Lafon, 2006), and its use in data fusion and manifold alignment (Lafon et al., 2006; Ham et al., 2005; Wang and Mahadevan, 2009) — but there the graph is estimated from the data rather than given by catalogued coordinates. Geometry has also been treated as something to measure rather than impose: translation-symmetric structure emerges spontaneously in language models from translation-invariant co-occurrence statistics (Karkada et al., 2026).

Shared and private representations. That only part of a multimodal representation admits a common basis is long established (Tucker, 1958; Flury, 1984; Eynard et al., 2015; Klami et al., 2013; Tsai et al., 2018), as is posing the recovery of a withheld modality as graph-regularized matrix completion (Kalofolias et al., 2014; Boyarski et al., 2019). Recent identifiability theory formalizes the split (Yao et al., 2023), but treats whatever appears in every view as shared content, without distinguishing physics from a shared systematic.

3 DATA

3.1 OVERVIEW

Stellar parameters such as temperature, surface gravity, chemical composition, rotation, and age describe a star's structure and its stage of evolution, and underpin much of stellar and Galactic astrophysics. They are inferred from three kinds of observations: *spectroscopy*, flux against wavelength; *light curves*, a stochastic time series of brightness against time; and *astrometry*, position and parallax, which give distance and kinematics. Some parameters are constrained by several of these channels, some by a single instrument capability, and some, like age, by none directly. We assemble a stellar dataset spanning **six surveys** and **seven observational modalities**, drawn from all measurement channels. **Spectroscopy**: LAMOST low-resolution optical spectra (Zhao et al., 2012), APOGEE near-infrared spectra (Majewski et al., 2017), GALAH high-resolution optical spectra (De Silva et al., 2015), and *Gaia* BP/RP (XP) spectrophotometry (De Angeli et al., 2023); **Photometric time series (light curve)**: *Kepler* long-cadence light curves (Borucki et al., 2010) and TESS light curves (Ricker et al., 2015); **Astrometry**: *Gaia* DR3 parallaxes, proper motions, and derived space velocities (Gaia Collaboration et al., 2023). More information about the surveys and the preprocessing steps is given in Section 7.1.

4 TWO-STEP MULTIMODAL MODEL

FAME is a 1.2B-parameter, two-step model. The first step, which accounts for most of the parameters, is pretraining of *specific-modality encoders*. The second step is a small (4.5M parameters), physics-motivated *alignment of modality features*. This separation allows optimizing the architecture, maximizing training samples for each modality, and easily fine-tuning the model and adapting it to new tasks. Pretraining combines self-supervised and supervised objectives, following established practice in astronomy (Walmsley et al., 2022; Kamai et al., 2025). For each survey we add a supervised loss over the subset of its catalogue labels that the modality can predict, when the label exists ($T_{\rm eff}$, $\log g$, [Fe/H] for the spectroscopic encoders; $P_{\rm rot}$, $\log g$ and absolute magnitude

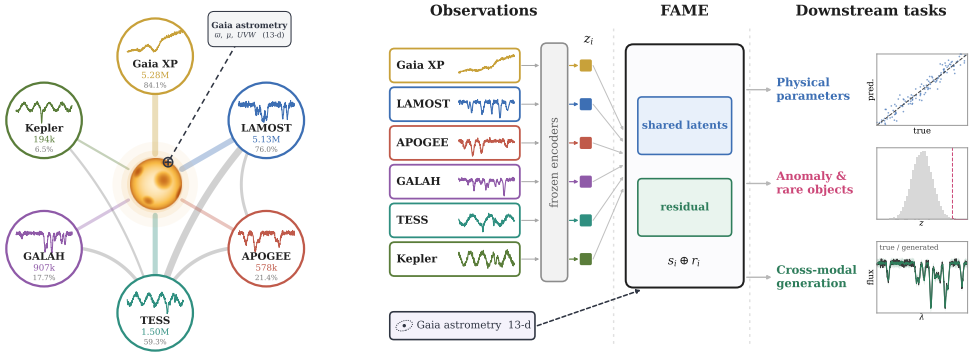

Figure 1: FAME observes one star through six encoded modalities, cross-matched on *Gaia* DR3 `source_id`; *Gaia* astrometry enters as a 13-d auxiliary channel. Spoke widths scale with coverage, chords with pairwise overlap. Stars encoded: *Gaia* XP 5,282,319; LAMOST 5,131,599; TESS 1,503,260; GALAH 906,695; APOGEE 578,163; *Kepler* 193,951; astrometry 1,301,941. The factorized latent supports parameter prediction, rare-population analysis, and cross-modal generation. Full census in Table 4.

for the light-curve encoders; abundances and $v \sin i$ for GALAH; see Table 5). This imposes a modality-specific inductive bias and, more importantly, reflects *concept reuse*: the model need not bootstrap all physical knowledge from scratch, but can build on properties already measured well. Two labels, age and chromospheric activity, enter no pretraining objective and serve as the held-out tests throughout. The choice of self-supervised objective in the modality-specific encoders is meant to be informed and motivated by the differences between spectroscopic and photometric modalities: spectra are less noisy and stochastic and therefore fit reconstruction-based learning. For light curves, strong systematics and the stochastic nature of the signal make joint embedding objectives preferable. This aligns with the findings of Van Assel et al. (2025) which showed that joint embedding methods are preferable when nuisance features are strong. Next, we specify the details of each part in the model.

4.1 Modality-specific encoders

All four spectroscopic modalities (LAMOST, APOGEE, GALAH, Gaia XP) share one architecture: a 1-D patch transformer trained with masked reconstruction (15%), plus a CQR pinball loss (Romano et al., 2019) over five quantiles on the survey's own catalogue parameters ($T_{\rm eff}$ for all; $\log g$ and [Fe/H] where the survey provides them; $v \sin i$ and 11 abundances for GALAH), weighted per star by spectral SNR.

The light-curve encoders use joint-embedding objectives instead: both transform the light curve into time-domain and frequency-domain channels (ACF and Lomb–Scargle), processed by an AstroConformer (Pan et al., 2024) and a CNN respectively, followed by a DualFormer module (Kamai et al., 2025; 2026a). The supervised loss is again CQR pinball, on rotation periods (Kamai et al., 2026a; Boyle et al., 2026; Van-Lane et al., 2025), $\log g$, and absolute magnitude. The two architectures differ by instrument: TESS light curves carry strong sector-dependent systematics, so the TESS encoder is a causal model (Audenaert et al., 2025) trained on triplets (same star across sectors, different stars within a sector) that splits its capacity between a physics and an instrument latent space; Kepler, with a single stable instrument, does not need this and uses a dual-view encoder.

4.2 Factorized Alignment

We note that classical alignment is not optimal in our case. The framework of Kamai et al. (2026b) predicts, for a modality pair and label, whether signal is recoverable by cross-alignment, by cross-prediction, by both, or by neither. Applied to our modality pairs, no single approach suffices, and many pairs, particularly for stellar age, fall in the *Neither* regime (Figure 6). Moreover, a single

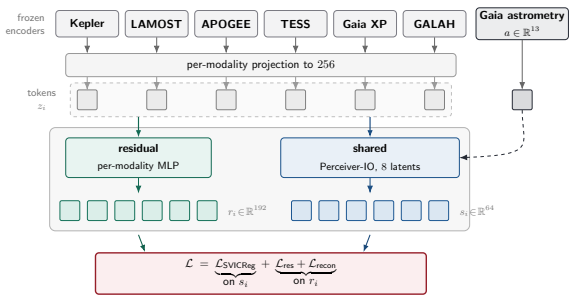

Figure 2: **The factorized alignment module.** Two branches read the same tokens, with 60% of z_i dimensions masked in a single pass. The *residual* stream is a per-modality MLP with no cross-token mixing, so a star's r_i depends only on its own z_i; two heads regress the modality's catalogue targets ($\mathcal{L}_{\text{res}}$) and reconstruct the unmasked features ($\mathcal{L}_{\text{recon}}$). The *shared* stream is a Perceiver-IO module, with *Gaia* astrometry entering as an auxiliary token in the cross-attention K/V. Its objective $\mathcal{L}_{\text{SVICReg}}$ is VICReg with smoothness Laplacian regularization. See Section 4.2 for details.

shared latent becomes a sharper bottleneck as modalities are added, and joint-embedding methods can collapse onto correlated nuisance features (Kamai et al., 2026b), which in astrophysics arise from shared background objects and systematics.

We therefore split FAME's latent space into shared and residual streams. The shared stream is a Perceiver-IO module (Jaegle et al., 2021) that compresses masked modality tokens into a learnable latent code and decodes them, with one learned query per modality, into *shared tokens*. It is optimized with *Smoothed VICReg (SVICReg)*, a variant of VICReg (Bardes et al., 2021). As noted in Section 1, VICReg's invariance term builds a graph from the cross-match alone, with no edges between distinct stars. We therefore add $\mathcal{L}_{\text{smooth}}$, which supplies those edges from catalogued coordinates, penalizing the squared difference of shared tokens that are close in the HR diagram:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{d_s} \frac{\sum_{(a,b)\in\mathcal{E}} \tilde{w}_{ab} \, \|\bar{s}_a - \bar{s}_b\|^2}{\sum_{(a,b)\in\mathcal{E}} \tilde{w}_{ab}}, \quad w_{ab} = \exp\!\Big(-\frac{\|c_a - c_b\|^2}{2\varepsilon^2}\Big), \quad \tilde{w}_{ab} = \frac{w_{ab}}{q_a \, q_b}, \quad q_a = \hat{p}(c_a), \tag{1}$$

where $a \neq b$ index stars within a minibatch, $\bar{s}_a \in \mathbb{R}^{d_s}$ ($d_s = 64$) is the presence-masked mean $(\sum_i m_{a,i} \, s_{a,i}) / \sum_i m_{a,i}$ of the per-modality shared tokens ($m_{a,i} = 1$ if star a has modality i), c_a is the min–max-normalized catalogue $(T_{\text{eff}}, \log g)$ of star a, $\mathcal{E}$ retains pairs with $\|c_a - c_b\| < 3\varepsilon$, and $\hat{p}$ is a Gaussian kernel density estimate of the sampling density, computed once over the training set. The $1/d_s$ and the $\sum \tilde{w}_{ab}$ denominator place the term on the scale of the VICReg invariance term.

Two choices are not cosmetic. The bandwidth ε and the density $\hat{p}$ are *fixed* rather than estimated per batch ($\varepsilon = 0.014$, about 60 K and 0.07 dex), so the kernel measures the same physical scale everywhere; and the per-node division by $q_a q_b$ is the $\alpha = 1$ member of the Coifman–Lafon family (Coifman and Lafon, 2006), which makes the continuum limit the *unweighted* Dirichlet energy $\int \|\nabla_c s\|^2 \, dc$ rather than one weighted by the sampling density. The scale of the minimiser is set by the VICReg variance and covariance terms, and as they are averaged over the batch, we weight them by $1/q_a$ as well. The invariance term is left unweighted: it compares pairs of modalities of the same star, with no density to cancel. Importantly, continuity in a catalogued coordinate space is not specific to astrophysics, so we see SVICReg as a general extension of VICReg.

In the residual stream, the masked z_i are projected to a lower dimension and sent to two predictors: a supervision MLP for labels that a modality was pretrained on ($\mathcal{L}_{\text{res}}$), and a decoder that reconstructs the masked features ($\mathcal{L}_{\text{recon}}$). We impose no orthogonality constraint between the streams, because some stellar parameters are inherently redundant: T_{eff} can be read from a single instrument yet correlates with everything the others see, and it serves as a label in the residual stream and as a coordinate in the shared one. Decorrelating the streams would oppose both roles. Figure 2 shows the module; training details are in Section 7.4.

5 RESULTS

5.1 THE ROLE OF $\mathcal{L}_{\mathrm{smooth}}$

We start by investigating the effect of $\mathcal{L}_{\mathrm{smooth}}$ on the latent space and on performance, with an ablation that sets $\mathcal{L}_{\mathrm{smooth}}$ to zero while keeping everything else fixed, including the random seed, data split, and architecture.

The geometry is approximately translation-equivariant. Beyond making the shared space smoother (Figure 7), $\mathcal{L}_{\mathrm{smooth}}$ organises it as Fourier modes in the parameters it penalises. The eigenmodes of the bin-by-bin Gram matrix of the shared token are low-order sinusoids in T_{eff} ($R^2_{\mathrm{Toep}} = 0.92$, the fraction of the Gram matrix explained by its best Toeplitz approximation; Figure 3, upper left), the signature of a stationary kernel, and successive principal-component pairs trace smooth Lissajous curves along which T_{eff} runs monotonically (upper middle and right), the geometry Karkada et al. (2026) derive for a translation-symmetric representation; the same holds for $\log g$ (Figure 8). This follows from the loss: in its continuum limit $\mathcal{L}_{\mathrm{smooth}}$ is a Dirichlet energy whose operator is the Laplacian of parameter space, and its low eigenfunctions are Fourier modes (Appendix 7.6). Removing $\mathcal{L}_{\mathrm{smooth}}$ breaks the geometry: the mode ladder disorders ($R^2_{\mathrm{Toep}} = 0.63$) and the Lissajous curves fragment (Figure 3, lower row).

The prior improves performance. While the latent space geometry is illuminating, the ultimate test for $\mathcal{L}_{\mathrm{smooth}}$ is on downstream tasks. We use a linear probe on the model's features with and without the smoothness term, on six labels, scored on test-set stars. On the shared stream alone, $\mathcal{L}_{\mathrm{smooth}}$ improves every label at every label budget (Figure 9). Figure 4 probes the full latent (shared and residual) and adds three reference arms: the unaligned pre-trained features z_i, a vanilla VICReg model (a single latent, width-matched to FAME, trained with the VICReg shared-stream objective and same data), and a supervised model trained end-to-end from scratch on the raw inputs. Read together, the arms form an ablation ladder. Vanilla VICReg falls *below* the pre-alignment features on most labels ($v \sin i$ throughout; T_{eff}, [Fe/H] and activity at all but the largest budgets; age at low budgets): compressing all modalities into a single shared code discards modality-specific information, so alignment alone can be destructive. Adding the residual stream recovers it: FAME, with or without $\mathcal{L}_{\mathrm{smooth}}$, matches or beats both baselines on every label at every budget shown, and $\mathcal{L}_{\mathrm{smooth}}$ improves performance further at nearly every budget, with smaller margins than in the shared stream alone, as expected: the prior acts only on the shared tokens and leaves the residual untouched.

The comparison with the supervised model splits by how a label relates to the modalities. The two labels absent from pretraining, age and activity, are the label-efficiency tests proper. Age is *multimodal* (best unimodal $R^2 = 0.72$, against 0.84 multimodal; Table 6), and FAME wins at every budget. Activity is read from the Ca II H & K cores that only LAMOST covers; FAME leads clearly up to 5k labels and the two are level from 10k, where the supervised model has enough labels to read the capability directly. FAME's margin on the *redundant* labels (T_{eff}, $\log g$, [Fe/H]) is expected: these were injected during pretraining (Section 4), and their curves test how well that knowledge survives alignment. Supervision wins outright only on $v \sin i$, read from line broadening that only the high-resolution spectrographs (APOGEE and GALAH) resolve: when the label-bearing capability is present and the label is direct, alignment has nothing to add. Section 5.3 turns this around: with the surveys carrying that capability withheld, FAME recovers these same *modality-specific* labels better than any surviving modality, and far better than a supervised model trained under the same modality-dropout schedule.

5.2 COMPARISON WITH OTHER FOUNDATION MODELS

We compare FAME against three existing models: AION-1 (Parker et al., 2025), a general-purpose omnimodal foundation model for astronomy; the stellar transformer of Leung and Bovy (2023), trained on *Gaia* XP with APOGEE labels; and DESA (Kamai et al., 2025), a Kepler–LAMOST multimodal model. All comparisons use a frozen ridge probe on identical stars and identical target definitions, scored with R^2 on test stars unseen by FAME's alignment module. Encoders are pretrained per survey with independent splits, so for labels that were pretraining targets (T_{eff}, $\log g$, [Fe/H], $v \sin i$, P_{rot}) FAME's encoders have seen most test stars' labels; these rows test whether pretraining knowledge survives alignment. The like-for-like rows are age and activity (bold), which enter no objective of any model. Table 1 shows the comparison with AION-1, whose tokens are mean-pooled within each survey block and concatenated. FAME outperforms AION-1 on every target. We also test how adding modalities affects each model, comparing the full arm against a

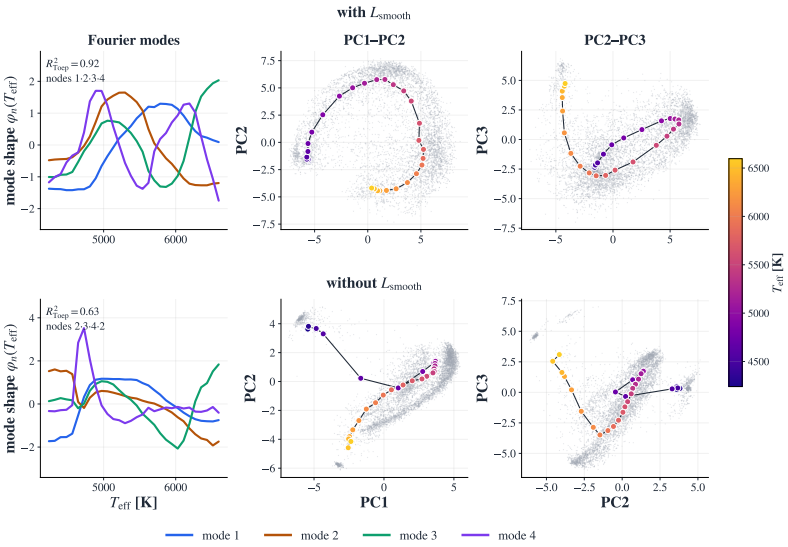

Figure 3: Fourier geometry induced in the shared latent space. Upper row: the full model, FAME with $\mathcal{L}_{\text{smooth}}$; the Gram-matrix eigenmodes are sinusoids in T_{eff} whose node count matches the mode index, and the principal-component pairs trace the curves Karkada et al. (2026) predict for a translation-symmetric representation. Lower row: the same model without $\mathcal{L}_{\text{smooth}}$ shows neither.

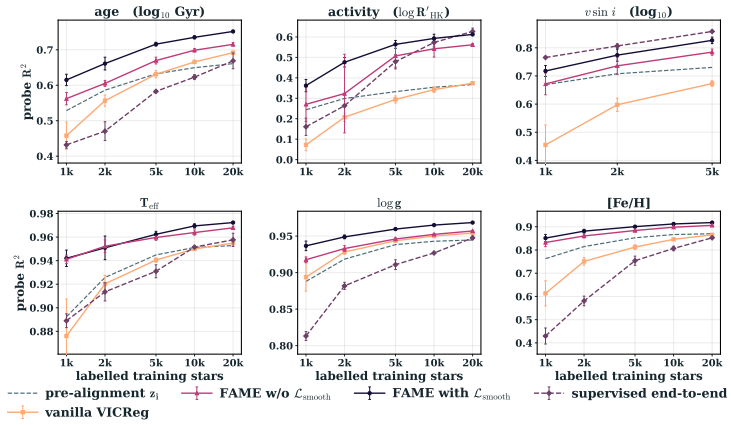

Figure 4: Linear probing of the full latent space (shared and residual) against label budget, on test stars unseen by the alignment module. Points are medians over 30 draws (3 probe seeds $\times$ 10 training subsamples) and bars a robust s.d. ($1.4826\times$MAD); FAME with $\mathcal{L}_{\text{smooth}}$ outperforms vanilla VICReg and the pre-alignment features on every label. Against the supervised model it leads on all labels but $v \sin i$, and is level with it on activity from 10k labels.

Gaia-only restriction. FAME benefits from additional modalities, while AION-1 does not (AION is worse than AION-gaia on every target, given *Gaia*, Legacy Survey photometry and imaging). The difference is striking since AION-gaia is better than FAME-gaia on six of seven targets (its *Gaia* block also ingests parallax, position and fluxes, so the two are not input-matched); only once all modalities are used does FAME come out ahead. We conclude that under linear probing of a frozen embedding, multimodality helps FAME and hurts AION-1; this result is scoped to this setting, since AION-1's own reported numbers use a learned, arm-specific cross-attention, under which multimodality helps. The comparison with Leung and Bovy (2023) shows the same pattern (Table 10). Restricted to *Gaia* inputs the two are close on the held-out labels (age 0.76 vs. 0.78, activity 0.23 vs. 0.25), and with all modalities FAME reaches 0.83 and 0.59. FAME's advantage is multimodality, not a better single-survey encoder. Table 2 compares FAME with DESA, the closest model in domain, which uses a similar hybrid pretraining. $v \sin i$, activity and $\log P_{\text{rot}}$ have fewer than 300 stars unseen by both models and are omitted. FAME wins in all other labels and more than doubles DESA's R^2 on age (0.49 vs. 0.22).

Table 1: R^2 of FAME vs. AION-1 under a frozen ridge probe, on test stars unseen by FAME's alignment module. AION columns take the best of its three sizes per task, which favours AION (all sizes in Table 9). *Gaia-only* restricts both models to one single shared modality. Bold values: best in row; bold targets (activity, age): labels that enter no pretraining objective of either model, the like-for-like comparison. The other targets are pretraining labels of FAME's encoders, which saw most of these stars' labels. FAME leads on all seven targets; AION-gaia is better than AION on all seven and than FAME-gaia on six (tied on $T_{\rm eff}$). Mean over 5 training subsamples (fixed test fold); s.d. ≤ 0.004 in every cell.

target	FAME	AION	FAME-gaia	AION-gaia
activity	**0.653**	0.310	0.286	0.415
$\log P_{\rm rot}$	**0.641**	0.403	0.302	0.475
$v \sin i$	**0.837**	0.596	0.597	0.680
[Fe/H]	**0.929**	0.823	0.784	0.864
age	**0.756**	0.659	0.648	0.711
$T_{\rm eff}$	**0.988**	0.970	0.975	0.975
$\log g$	**0.980**	0.961	0.944	0.969

Table 2: R^2 of FAME vs. DESA on Kepler×LAMOST stars unseen by DESA and by FAME's alignment module (median and 16–84% range of 50 repeats). *K+L* restricts FAME to DESA's two surveys. Age (bold) is the like-for-like label.

target	DESA	FAME-K+L	FAME-full
$T_{\rm eff}$	$0.935^{+0.015}_{-0.027}$	$0.979^{+0.007}_{-0.013}$	$\mathbf{0.986}^{+0.004}_{-0.005}$
$\log g$	$0.962^{+0.009}_{-0.016}$	$0.966^{+0.007}_{-0.023}$	$\mathbf{0.973}^{+0.007}_{-0.006}$
[Fe/H]	$0.869^{+0.038}_{-0.078}$	$0.947^{+0.005}_{-0.016}$	$\mathbf{0.951}^{+0.010}_{-0.012}$
age	$0.219^{+0.059}_{-0.098}$	$\mathbf{0.486}^{+0.031}_{-0.086}$	$0.468^{+0.050}_{-0.076}$

Table 3: R^2 of supervised baseline vs. FAME under modality withholding (activity $n = 4{,}633$, $v \sin i$ $n = 1{,}901$). Both arms are trained with identical modality-dropout of the label-bearing surveys; *withheld* scores the same weights with those surveys fully removed (LAMOST for activity; APOGEE and GALAH for $v \sin i$).

	Activity		$v \sin i$	
	full	withheld	full	withheld
Supervised	0.744	0.073	0.883	0.186
FAME	0.727	**0.457**	0.890	**0.592**

5.3 Cross-modal prediction

Chromospheric activity ($\log R'_{\rm HK}$) needs the Ca II H & K cores, which only LAMOST covers; $v \sin i$ needs resolution sufficient for line broadening, which only APOGEE and GALAH reach (99.8% of labelled stars carry one of the two). With all modalities present these are exactly the labels where supervision is competitive (Section 5.1). We ask what happens when the capability is absent: we fine-tune FAME with modality dropout (label-bearing surveys masked in 50% of training samples) and evaluate with *every* survey carrying the capability withheld from every test star.

Recovery. Fine-tuned FAME reaches $R^2 = 0.736$ (activity) and 0.921 ($v \sin i$) with all modalities, against 0.658 and 0.713 for the best single survey (Table 7). Withheld, it retains 0.469 and 0.640: ahead of every surviving unimodal arm (best *Gaia* XP, 0.352 and 0.564), though below the defining survey. The fusion has to adapt rather than merely be probed: with the spectra withheld a probe on the frozen backbone reaches only $R^2 \approx 0.2$ on $v \sin i$.

Against a supervised model under the same schedule. Table 3 repeats the experiment with a supervised model (a 21M-parameter multi-branch transformer on the raw inputs, trained from scratch, as in Section 5.1) under the same dropout schedule, training stars and labels; scores are on the matched test stars (smaller subsets, so cells are not comparable with Table 7). At full input it is competitive, slightly ahead on activity and level on $v \sin i$; withheld, it retains $R^2 = 0.07$ and 0.19 against FAME's 0.46 and 0.59. The supervised baseline is not worse at the task, it is worse at missing data: the capability transfers through the representation, not through the schedule.

5.4 Rare populations in the residual stream

If the factorization works as designed, peculiarity should survive in the residual stream, which keeps what a single modality sees that the others do not. We test this on six known-peculiar populations

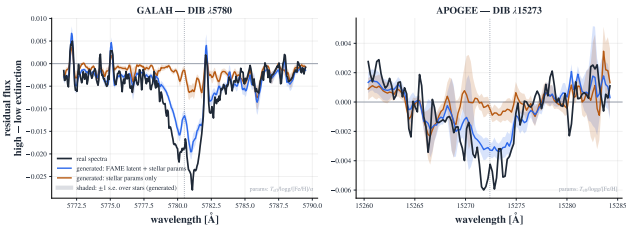

Figure 5: Diffuse interstellar band in generated spectra. Blue: flow-matching model conditioned on stellar parameters and FAME's latent (label survey excluded); brown: the same model conditioned on parameters alone; black: real spectrum. Only the latent-conditioned model recovers the band.

(Norfolk et al., 2019; Gao et al., 2025; Buder et al., 2025), each scored against controls matched on $(T_{\rm eff}, \log g, [\mathrm{Fe/H}], M_G, \varpi)$ and on modality signature, so neither bulk parameters nor survey selection can drive the score. A cross-validated linear probe on the residual features reads every population (Table 8, AUC 0.70–0.95) from as few as 46 labelled members, and is ahead of the unaligned encoder features z_i on four of the six populations despite being $30\times$ more compact (192 vs. 5,888 dimensions). On the shared stream the probe is weaker on every population (mean -0.17 AUC) and near chance on the s-process and δ Scuti stars: the peculiarity the encoders capture is preserved in the residual and largely lost in the modality-agnostic stream, as the design intends. This is population analysis, not label-free anomaly detection, which we do not claim.

5.5 Cross-modal generation

Finally, we probe what the representation encodes by generating through it: a flow-matching model (Lipman et al., 2022) conditioned on FAME's latent tokens and three bulk parameters ($T_{\rm eff}$, $\log g$, [Fe/H]) generates high-resolution APOGEE or GALAH spectra for held-out stars with that survey withheld (Figures 10 and 11). Over a capacity-matched baseline conditioned on the bulk parameters alone the latent adds a small, consistent increment ($\sim$ 1–2% RMSE); the informative question is where. The clearest case is interstellar dust (Fig. 5): the latent reproduces the diffuse interstellar band, which the parameters-only baseline leaves absent. The band is interstellar, independent of the stellar atmosphere, so its recovery shows that FAME's latents carry the *Gaia* extinction signal into spectral generation.

6 Conclusions

We presented FAME, a foundation model for stellar astrophysics that fuses seven modalities from six surveys through pretrained encoders and a small, factorized alignment module. Our central methodological finding is that multimodal fusion is not automatically beneficial: under joint-embedding alignment, forcing all modalities into a single shared latent can perform worse than not aligning at all, and under frozen probing adding modalities degrades AION-1's representations while it strengthens FAME's. Controlled ablations tie the gains to our design decisions: the shared–residual factorization recovers the modality-specific information a single latent discards, and the physical smoothness prior improves probing on every label while organizing the shared space along stellar parameters.

Under frozen probing the representation outperforms three existing astrophysics foundation models and is more label-efficient than a supervised baseline on the labels absent from pretraining; supervision wins outright only on $v \sin i$, read directly from high-resolution spectra. Where the surveys carrying a label's capability are withheld, FAME recovers the label while a supervised model trained under the same schedule retains little: the representation, not the training curriculum, carries each instrument's knowledge beyond the instrument itself.

FAME's advantages are in frozen probing and missing-modality regimes, not in end-to-end fine-tuning with every modality present (Table 6). Recovery between surveys with no stars in common, which the shared space admits in principle, is untested, as is the generality of latent factorization and SVICReg outside astronomy; these are the direct next steps.

AI USE STATEMENT

In this work, we used generative AI tools for code generation, creation of scientific figures, suggesting experimental parameters, and drafting parts of the paper. We have not used generative AI tools for data generation and interpretation of results. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by the main author. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All data are public survey products; Appendix 7.1 lists each survey's release, the selection cuts, and the preprocessing applied to every modality, and Appendix 7.2 describes the cross-match on *Gaia* DR3 `source_id` with the resulting census (Table 4). The age label construction, including the catalogue priority order and quality cuts, is given in Appendix 7.3. Encoder architectures, parameter counts, and pretraining heads are in Table 5; the alignment objective is specified by Equation 1 and Section 4.2, with the training subset, loss weights, bandwidth, optimizer settings and compute in Appendix 7.4. Appendix 7.5 gives the evaluation protocol: which stars each test set excludes, the probe, and what the reported spreads measure. Encoders and the alignment module are single training runs at a fixed seed; where results carry repeats, their number and kind are stated in the caption or in Appendix 7.5. Code for the alignment module, the probing and withholding experiments, and the frozen encoder weights will be released upon publication.

ACKNOWLEDGMENTS

REFERENCES

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, art. arXiv:1810.04805, October 2018. doi: 10.48550/arXiv.1810.04805.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *arXiv e-prints*, art. arXiv:2005.14165, May 2020. doi: 10.48550/arXiv.2005.14165.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2021.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL `https://www.science.org/doi/abs/10.1126/science.ade2574`.

Salvatore Candido, Thomas Hayes, Alexander Derry, Roshan Rao, Zeming Lin, Robert Verkuil, Bryan Z. Wu, Jin Sub Lee, Elise S. Bruguera, Jehan A. Keval, Mykhailo Kopylov, John E. Pak, Wesley Wu, Neil Thomas, Samson Mataraso, Alvin Hsu, Ashton C. Trotman-Grant, Kilian Fatras, Allan dos Santos Costa, Rohil Badkundri, Halil Akın, Deniz Oktay, Jonathan Deaton, Elizabeth Montabana, Hrishita Sitwala, Yue Yu, Marius Wiggert, Dylan Alexander Carlin, Anthony W.

Goering, Tomasz Blazejewski, McCullen Sandora, Michael Hla, Tina Z. Jia, Leon H. Kloker, Nicholas J. Sofroniew, Masatoshi Uehara, Jassi Pannu, Sharrol Bachas, Daniel S. Liu, Tom Sercu, and Alexander Rives. Language modeling materializes a world model of protein biology. *bioRxiv*, 2026. doi: 10.64898/2026.06.03.729735. URL https://www.biorxiv.org/content/early/2026/06/04/2026.06.03.729735.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhowmik, Filippo Bigi, Samuel M. Blau, Vlad Cărare, Michele Ceriotti, Sanggyu Chong, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Edvin Fako, Fabio Falcioni, Andrea C. Ferrari, John L. A. Gardner, Mikołaj J. Gawkowski, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Jonas Grandel, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Cheuk Hin Ho, Stephan Hofmann, Christian Holm, Jad Jaafar, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Panagiotis Kourtis, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Chen Lin, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Louise A. M. Rosset, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Christopher Sutton, Viktor Svahn, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Santiago Vargas, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Thomas Wolf, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. J. Chem. Phys., 163(18):184110, November 2025. doi: 10.1063/5.0297006.

Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R. Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, Remi Lam, and Matthew Willson. Probabilistic weather forecasting with machine learning. Nature, 637(8044): 84–90, January 2025. doi: 10.1038/s41586-024-08252-9.

Wahid Bhimji, Chris Harris, Vinicius Mikuni, and Benjamin Nachman. Foundation model framework for all tasks involving jet physics. Phys. Rev. D, 113(3):032020, February 2026. doi: 10.1103/knmd-f5jm.

Gemma Zhang, Thomas Helfer, Alexander T. Gagliano, Siddharth Mishra-Sharma, and V. Ashley Villar. Maven: a multimodal foundation model for supernova science. *Machine Learning: Science and Technology*, 5(4):045069, December 2024. doi: 10.1088/2632-2153/ad990d.

M. Rizhko and J. S. Bloom. AstroM^3: A Self-supervised Multimodal Model for Astronomy. AJ, 170(1):28, July 2025. doi: 10.3847/1538-3881/adcbad.

Liam Parker, Francois Lanusse, Siavash Golkar, Leopoldo Sarra, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Geraud Krawezik, Michael McCabe, Rudy Morel, Ruben Ohana, Mariel Pettee, Bruno Régaldo-Saint Blancard, Kyunghyun Cho, Shirley Ho, and Polymathic AI Collaboration. AstroCLIP: a cross-modal foundation model for galaxies. MNRAS, 531(4):4990–5011, July 2024. doi: 10.1093/mnras/stae1450.

Ilay Kamai, Alex M. Bronstein, and Hagai B. Perets. Machine Learning Inference of Stellar Properties Using Integrated Photometric and Spectroscopic Data. ApJ, 994(1):110, November 2025. doi: 10.3847/1538-4357/ae0cbc.

Liam Parker, Francois Lanusse, Jeff Shen, Ollie Liu, Tom Hehir, Leopoldo Sarra, Lucas Meyer, Micah Bowles, Sebastian Wagner-Carena, Helen Qu, Siavash Golkar, Alberto Bietti, Hatim Bourfoune, Nathan Casserau, Pierre Cornette, Keiya Hirashima, Geraud Krawezik, Ruben Ohana, Nicholas Lourie, Michael McCabe, Rudy Morel, Payel Mukhopadhyay, Mariel Pettee, Bruno Regaldo-Saint Blancard, Kyunghyun Cho, Miles Cranmer, and Shirley Ho. AION-1: Omnimodal Foundation Model for Astronomical Sciences. *arXiv e-prints*, art. arXiv:2510.17960, October 2025. doi: 10.48550/arXiv.2510.17960.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *arXiv e-prints*, art. arXiv:2105.04906, May 2021. doi: 10.48550/arXiv.2105.04906.

Randall Balestriero and Yann LeCun. Contrastive and non-contrastive self-supervised learning recover global and local spectral embedding methods. *ArXiv*, abs/2205.11508, 2022. URL https://api.semanticscholar.org/CorpusID:248986152.

Henry W. Leung and Jo Bovy. Towards an astronomical foundation model for stars with a transformer-based model. 2023. URL https://api.semanticscholar.org/CorpusID:261064643.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. *ArXiv*, abs/2002.05709, 2020. URL https://api.semanticscholar.org/CorpusID:211096730.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, 2021. URL https://api.semanticscholar.org/CorpusID:232110471.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

Uri Shaham, Kelly P. Stanton, Henry Li, Boaz Nadler, Ronen Basri, and Yuval Kluger. Spectralnet: Spectral clustering using deep neural networks. *ArXiv*, abs/1801.01587, 2018. URL https://api.semanticscholar.org/CorpusID:3278749.

Idan Simai, Ronen Talmon, and Uri Shaham. Enhancing vicreg: Random-walk pairing for improved generalization and better global semantics capturing. *ArXiv*, abs/2506.18104, 2025. URL https://api.semanticscholar.org/CorpusID:279999190.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7(85):2399–2434, 2006.

Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2006.04.006. URL https://www.sciencedirect.com/science/article/pii/S1063520306000546. Special Issue: Diffusion Maps and Wavelets.

S. Lafon, Y. Keller, and R.R. Coifman. Data fusion and multicue data matching by diffusion maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1784–1797, 2006. doi: 10.1109/TPAMI.2006.223.

Jihun Ham, Daniel D. Lee, and Lawrence K. Saul. Semisupervised alignment of manifolds. In *International Conference on Artificial Intelligence and Statistics*, 2005. URL https://api.semanticscholar.org/CorpusID:2913820.

Chang Wang and Sridhar Mahadevan. Manifold alignment without correspondence. IJCAI'09, page 1273–1278, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

Dhruva Karkada, Daniel J. Korchinski, Andres Nava, Matthieu Wyart, and Yasaman Bahri. Symmetry in language statistics shapes the geometry of model representations. *arXiv e-prints*, art. arXiv:2602.15029, February 2026. doi: 10.48550/arXiv.2602.15029.

Ledyard R. Tucker. An inter-battery method of factor analysis. *Psychometrika*, 23:111–136, 1958. URL https://api.semanticscholar.org/CorpusID:123711893.

Bernhard N. Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984. doi: 10.1080/01621459.1984.10477108. URL https://doi.org/10.1080/01621459.1984.10477108.

Davide Eynard, Artiom Kovnatsky, Michael M. Bronstein, Klaus Glashoff, and Alexander M. Bronstein. Multimodal manifold analysis by simultaneous diagonalization of laplacians. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(12):2505–2517, 2015. doi: 10.1109/TPAMI.2015.2408348.

Arto Klami, Seppo Virtanen, and Samuel Kaski. Bayesian canonical correlation analysis. *J. Mach. Learn. Res.*, 14(1):965–1003, April 2013. ISSN 1532-4435.

Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis philippe Morency, and Ruslan Salakhutdinov. Learning factorized multimodal representations. *ArXiv*, abs/1806.06176, 2018. URL https://api.semanticscholar.org/CorpusID:49303347.

Vassilis Kalofolias, Xavier Bresson, Michael M. Bronstein, and Pierre Vandergheynst. Matrix completion on graphs. *ArXiv*, abs/1408.1717, 2014. URL https://api.semanticscholar.org/CorpusID:14746688.

Amit Boyarski, Sanketh Vedula, and Alexander M. Bronstein. Spectral geometric matrix completion. In *Mathematical and Scientific Machine Learning*, 2019. URL https://api.semanticscholar.org/CorpusID:235363958.

Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. *ArXiv*, abs/2311.04056, 2023. URL https://api.semanticscholar.org/CorpusID:265043539.

Gang Zhao, Yongheng Zhao, Y. Chu, Y. Jing, and L. Deng. Lamost spectral survey — an overview. *Research in Astronomy and Astrophysics*, 2012. doi: 10.1088/1674-4527/12/7/002. URL https://www.semanticscholar.org/paper/6dd33c0f7d83c6cbc8e2ee544cee259518d6f474.

Steven R. Majewski, Ricardo P. Schiavon, Peter M. Frinchaboy, Carlos Allende Prieto, Robert Barkhouser, Dmitry Bizyaev, Basil Blank, Sophia Brunner, Adam Burton, Ricardo Carrera, S. Drew Chojnowski, Kátia Cunha, Courtney Epstein, Greg Fitzgerald, Ana E. García Pérez, Fred R. Hearty, Chuck Henderson, Jon A. Holtzman, Jennifer A. Johnson, Charles R. Lam, James E. Lawler, Paul Maseman, Szabolcs Mészáros, Matthew Nelson, Duy Coung Nguyen, David L. Nidever, Marc Pinsonneault, Matthew Shetrone, Stephen Smee, Verne V. Smith, Todd Stolberg, Michael F. Skrutskie, Eric Walker, John C. Wilson, Gail Zasowski, Friedrich Anders, Sarbani Basu, Stephane Beland, Michael R. Blanton, Jo Bovy, Joel R. Brownstein, Joleen Carlberg, William Chaplin, Cristina Chiappini, Daniel J. Eisenstein, Yvonne Elsworth, Diane Feuillet, Scott W. Fleming, Jessica Galbraith-Frew, Rafael A. García, D. Aníbal García-Hernández, Bruce A. Gillespie, Léo Girardi, James E. Gunn, Sten Hasselquist, Michael R. Hayden, Saskia Hekker, Inese Ivans, Karen Kinemuchi, Mark Klaene, Suvrath Mahadevan, Savita Mathur, Benoît Mosser, Demitri Muna, Jeffrey A. Munn, Robert C. Nichol, Robert W. O'Connell, John K. Parejko, A. C. Robin, Helio Rocha-Pinto, Matthias Schultheis, Aldo M. Serenelli, Neville Shane, Victor Silva Aguirre, Jennifer S. Sobeck, Benjamin Thompson, Nicholas W. Troup, David H. Weinberg, and Olga Zamora. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). AJ, 154(3):94, September 2017. doi: 10.3847/1538-3881/aa784d.

G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. Wylie de Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D'Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O'Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D. Simpson, D. Stello, Y.-S. Ting, G. Traven, F. Watson, R. Wittenmyer, D. Yong, and M. Žerjal. The GALAH survey: scientific motivation. MNRAS, 449(3): 2604–2617, May 2015. doi: 10.1093/mnras/stv327.

F. De Angeli, M. Weiler, P. Montegriffo, D. W. Evans, M. Riello, R. Andrae, J. M. Carrasco, G. Busso, P. W. Burgess, C. Cacciari, M. Davidson, D. L. Harrison, S. T. Hodgkin, C. Jordi, P. J. Osborne, E. Pancino, G. Altavilla, M. A. Barstow, C. A. L. Bailer-Jones, M. Bellazzini, A. G. A. Brown, M. Castellani, S. Cowell, L. Delchambre, F. De Luise, C. Diener, C. Fabricius, M. Fouesneau, Y. Frémat, G. Gilmore, G. Giuffrida, N. C. Hambly, S. Hidalgo, G. Holland, Z. Kostrzewa-Rutkowska, F. van Leeuwen, A. Lobel, S. Marinoni, N. Miller, C. Pagani, L. Palaversa, A. M. Piersimoni, L. Pulone, S. Ragaini, M. Rainer, P. J. Richards, G. T. Rixon, D. Ruz-Mieres, N. Sanna, L. M. Sarro, N. Rowell, R. Sordo, N. A. Walton, and A. Yoldas. Gaia Data Release 3. Processing and validation of BP/RP low-resolution spectral data. A&A, 674:A2, June 2023. doi: 10.1051/0004-6361/202243680.

William J. Borucki, David Koch, Gibor Basri, Natalie Batalha, Timothy Brown, Douglas Caldwell, John Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, Andrea K. Dupree, Thomas N. Gautier, John C. Geary, Ronald Gilliland, Alan Gould, Steve B. Howell, Jon M. Jenkins, Yoji Kondo, David W. Latham, Geoffrey W. Marcy, Søren Meibom, Hans Kjeldsen, Jack J. Lissauer, David G. Monet, David Morrison, Dimitar Sasselov, Jill Tarter, Alan Boss, Don Brownlee, Toby Owen, Derek Buzasi, David Charbonneau, Laurance Doyle, Jonathan Fortney, Eric B. Ford, Matthew J. Holman, Sara Seager, Jason H. Steffen, William F. Welsh, Jason Rowe, Howard Anderson, Lars Buchhave, David Ciardi, Lucianne Walkowicz, William Sherry, Elliott Horch, Howard Isaacson, Mark E. Everett, Debra Fischer, Guillermo Torres, John Asher Johnson, Michael Endl, Phillip MacQueen, Stephen T. Bryson, Jessie Dotson, Michael Haas, Jeffrey Kolodziejczak, Jeffrey Van Cleve, Hema Chandrasekaran, Joseph D. Twicken, Elisa V. Quintana, Bruce D. Clarke, Christopher Allen, Jie Li, Haley Wu, Peter Tenenbaum, Ekaterina Verner, Frederick Bruhweiler, Jason Barnes, and Andrej Prsa. Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327(5968):977, February 2010. doi: 10.1126/science.1185402.

George R. Ricker, Joshua N. Winn, Roland Vanderspek, David W. Latham, Gáspár Á. Bakos, Jacob L. Bean, Zachory K. Berta-Thompson, Timothy M. Brown, Lars Buchhave, Nathaniel R. Butler, R. Paul Butler, William J. Chaplin, David Charbonneau, Jørgen Christensen-Dalsgaard, Mark Clampin, Drake Deming, John Doty, Nathan De Lee, Courtney Dressing, Edward W. Dunham, Michael Endl, Francois Fressin, Jian Ge, Thomas Henning, Matthew J. Holman, Andrew W. Howard, Shigeru Ida, Jon M. Jenkins, Garrett Jernigan, John Asher Johnson, Lisa Kaltenegger, Nobuyuki Kawai, Hans Kjeldsen, Gregory Laughlin, Alan M. Levine, Douglas Lin, Jack J. Lissauer, Phillip MacQueen, Geoffrey Marcy, Peter R. McCullough, Timothy D. Morton, Norio Narita, Martin Paegert, Enric Palle, Francesco Pepe, Joshua Pepper, Andreas Quirrenbach, Stephen A. Rinehart, Dimitar Sasselov, Bun'ei Sato, Sara Seager, Alessandro Sozzetti, Keivan G. Stassun, Peter Sullivan, Andrew Szentgyorgyi, Guillermo Torres, Stephane Udry, and Joel Villasenor. Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1:014003, January 2015. doi: 10.1117/1.JATIS.1.1.014003.

Gaia Collaboration, A. Vallenari, A. G. A. Brown, T. Prusti, J. H. J. de Bruijne, F. Arenou, C. Babusiaux, M. Biermann, O. L. Creevey, C. Ducourant, D. W. Evans, L. Eyer, R. Guerra, A. Hutton, C. Jordi, S. A. Klioner, U. L. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, P. Tanga, N. A. Walton, C. A. L. Bailer-Jones, U. Bastian, R. Drimmel, F. Jansen, D. Katz, M. G. Lattanzi, F. van Leeuwen, J. Bakker, C. Cacciari, J. Castañeda, F. De Angeli, C. Fabricius, M. Fouesneau, Y. Frémat, L. Galluccio, A. Guerrier, U. Heiter, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzzo, F. Riclet, W. Roux, G. M. Seabroke, R. Sordo, F. Thévenin, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. W. Burgess, D. Busonero, G. Busso, H. Cánovas, B. Carry, A. Cellino, N. Cheek, G. Clementini, Y. Damerdji, M. Davidson, P. de Teodoro, M. Nuñez Campos, L. Delchambre, A. Dell'Oro, P. Esquej, J. Fernández-Hernández, E. Fraile, D. Garabato, P. García-Lario, E. Gosset, R. Haigron, J.-L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, B. Holl, K. Janßen, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Löffler, O. Marchal, P. M. Marrese, A. Moitinho, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, A. Recio-Blanco, C. Reylé, M. Riello, L. Rimoldini, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, E. Utrilla, M. van Leeuwen, U. Abbas, P. Ábrahám, A. Abreu Aramburu, C. Aerts, J. J. Aguado, M. Ajaj, F. Aldea-Montero, G. Altavilla, M. A. Álvarez, J. Alves, F. Anders, R. I. Anderson, E. Anglada Varela, T. Antoja, D. Baines, S. G. Baker, L. Balaguer-Núñez, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S. Bartolomé, J.-L. Bassilana, N. Bauchet, U. Becciani, M. Bellazzini, A. Berihuete, M. Bernet, S. Bertone, L. Bianchi, A. Binnenfeld, S. Blanco-Cuaresma, A. Blazere, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, E. Brugaletta, B. Bucciarelli, A. Burlacu, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, T. Cantat-Gaudin, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, L. Chaoul, P. Charlot, L. Chemin, V. Chiaramida, A. Chiavassa, N. Chornay, G. Comoretto, G. Contursi, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Cropper, M. Crosta, C. Crowley, C. Dafonte, A. Dapergolas, M. David, P. David, P. de Laverny, F. De

Luise, and R. De March. Gaia Data Release 3. Summary of the content and survey properties. A&A, 674:A1, June 2023. doi: 10.1051/0004-6361/202243940.

Mike Walmsley, Inigo Slijepcevic, Micah R. Bowles, and Anna Scaife. Toward Galaxy Foundation Models with Hybrid Contrastive Learning. In *Machine Learning for Astrophysics*, page 29, July 2022. doi: 10.48550/arXiv.2206.11927.

Hugues Van Assel, Mark Ibrahim, Tommaso Biancalani, Aviv Regev, and Randall Balestriero. Joint Embedding vs Reconstruction: Provable Benefits of Latent Space Prediction for Self Supervised Learning. *arXiv e-prints*, art. arXiv:2505.12477, May 2025. doi: 10.48550/arXiv.2505.12477.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized Quantile Regression. *arXiv e-prints*, art. arXiv:1905.03222, May 2019. doi: 10.48550/arXiv.1905.03222.

Jia-Shu Pan, Yuan-Sen Ting, and Jie Yu. Astroconformer: The prospects of analysing stellar light curves with transformer-based deep learning models. MNRAS, 528(4):5890–5903, March 2024. doi: 10.1093/mnras/stae068.

Ilay Kamai, Lavi Somers, and Hagai B. Perets. The Maunder Model and Catalog: Stellar Rotation, Bimodal Activity, and Magnetic Braking in Kepler Main-Sequence Stars. *arXiv e-prints*, art. arXiv:2608.06604, August 2026a.

Andrew W. Boyle, Luke G. Bouma, and Andrew W. Mann. The TESS All-Sky Rotation Survey: Periods for 1,046,317 Stars within 500 pc. ApJS, 284(2):75, June 2026. doi: 10.3847/1538-4365/ae6657.

Phil R. Van-Lane, Joshua S. Speagle, Gwendolyn M. Eadie, Stephanie T. Douglas, Phillip A. Cargile, Catherine Zucker, Yuxi (Lucy) Lu, and Ruth Angus. ChronoFlow: A Data-driven Model for Gyrochronology. ApJ, 986(1):59, June 2025. doi: 10.3847/1538-4357/adcd73.

Jeroen Audenaert, Daniel Muthukrishna, Paul F. Gregory, David W. Hogg, and V. Ashley Villar. Causal Foundation Models: Disentangling Physics from Instrument Properties. *arXiv e-prints*, art. arXiv:2507.05333, July 2025. doi: 10.48550/arXiv.2507.05333.

Ilay Kamai, Hugues Van Assel, Aviv Regev, Hagai B. Perets, and Randall Balestriero. When to Align, When to Predict: A Phase Diagram for Multimodal Learning. *arXiv e-prints*, art. arXiv:2606.11190, June 2026b. doi: 10.48550/arXiv.2606.11190.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver IO: A General Architecture for Structured Inputs & Outputs. *arXiv e-prints*, art. arXiv:2107.14795, July 2021. doi: 10.48550/arXiv.2107.14795.

Brodie J. Norfolk, Andrew R. Casey, Amanda I. Karakas, Matthew T. Miles, Alex J. Kemp, Kevin C. Schlaufman, Melissa Ness, Anna Y. Q. Ho, John C. Lattanzio, and Alexander P. Ji. Discovery of s-process enhanced stars in the LAMOST survey. MNRAS, 490(2):2219–2227, December 2019. doi: 10.1093/mnras/stz2630.

Xinyi Gao, Xiaodian Chen, Shu Wang, and Jifeng Liu. Classification of Periodic Variable Stars from TESS. ApJS, 276(2):57, February 2025. doi: 10.3847/1538-4365/ad9dd6.

Sven Buder, Janez Kos, Xi Ella Wang, Madeleine McKenzie, Madeleine Howell, Sarah Martell, Michael R. Hayden, Daniel B. Zucker, Thomas Nordlander, Benjamin Montet, Gregor Traven, Joss Bland-Hawthorn, Gayandhi M. de Silva, Kenneth Freeman, Geraint Lewis, Karin Lind, Sanjib Sharma, Jeffrey D. Simpson, Dennis Stello, Tomaz Zwitter, Anish M. Amarsi, Joseph J. Armstrong, Kirsten Banks, Mark Beavis, Kevin-Luke Beeson, Boquan Chen, Ioana Ciucă, Gary S. da Costa, Richard de Grijs, Bailey Martin, David Moise Nataf, Melissa Ness, Adam D. Rains, Tim Scarr, Rok Vogrinčič, Zixian Purmortal Wang, Rob A. Wittenmyer, Yi Anne Xie, and The Galah Collaboration. The GALAH survey: Data release 4. PASA, 42:e051, May 2025. doi: 10.1017/pasa.2025.26.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. *arXiv e-prints*, art. arXiv:2210.02747, October 2022. doi: 10.48550/arXiv.2210.02747.

STScI. Kepler lc, q0-q17, 2016. URL `http://archive.stsci.edu/doi/resolve/resolve.html?doi=10.17909/T9488N`.

A. McQuillan, S. Aigrain, and T. Mazeh. Measuring the rotation period distribution of field M dwarfs with Kepler. MNRAS, 432(2):1203–1216, June 2013. doi: 10.1093/mnras/stt536.

N. R. Lomb. Least-Squares Frequency Analysis of Unequally Spaced Data. Ap&SS, 39(2):447–462, February 1976. doi: 10.1007/BF00648343.

J. D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. ApJ, 263:835–853, December 1982. doi: 10.1086/160554.

Chelsea X. Huang, Andrew Vanderburg, Andras Pál, Lizhou Sha, Liang Yu, Willie Fong, Michael Fausnaugh, Avi Shporer, Natalia Guerrero, Roland Vanderspek, and George Ricker. Photometry of 10 Million Stars from the First Two Years of TESS Full Frame Images: Part II. *Research Notes of the American Astronomical Society*, 4(11):206, November 2020. doi: 10.3847/2515-5172/abca2d.

Abdurro'uf, Katherine Accetta, Conny Aerts, Víctor Silva Aguirre, Romina Ahumada, Nikhil Ajgaonkar, N. Filiz Ak, Shadab Alam, Carlos Allende Prieto, Andrés Almeida, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Erik Aquino-Ortíz, Alfonso Aragón-Salamanca, Maria Argudo-Fernández, Metin Ata, Marie Aubert, Vladimir Avila-Reese, Carles Badenes, Rodolfo H. Barbá, Kat Barger, Jorge K. Barrera-Ballesteros, Rachael L. Beaton, Timothy C. Beers, Francesco Belfiore, Chad F. Bender, Mariangela Bernardi, Matthew A. Bershady, Florian Beutler, Christian Moni Bidin, Jonathan C. Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Nicholas Fraser Boardman, Adam S. Bolton, Médéric Boquien, Jura Borissova, Jo Bovy, W. N. Brandt, Jordan Brown, Joel R. Brownstein, Marcella Brusa, Johannes Buchner, Kevin Bundy, Joseph N. Burchett, Martin Bureau, Adam Burgasser, Tuesday K. Cabang, Stephanie Campbell, Michele Cappellari, Joleen K. Carlberg, Fábio Carneiro Wanderley, Ricardo Carrera, Jennifer Cash, Yan-Ping Chen, Wei-Huai Chen, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, S. Drew Chojnowski, Haeun Chung, Nicolas Clerc, Roger E. Cohen, Julia M. Comerford, Johan Comparat, Luiz da Costa, Kevin Covey, Jeffrey D. Crane, Irene Cruz-Gonzalez, Connor Culhane, Katia Cunha, Y. Sophia Dai, Guillermo Damke, Jeremy Darling, James W. Davidson, Jr., Roger Davies, Kyle Dawson, Nathan De Lee, Aleksandar M. Diamond-Stanic, Mariana Cano-Díaz, Helena Domínguez Sánchez, John Donor, Chris Duckworth, Tom Dwelly, Daniel J. Eisenstein, Yvonne P. Elsworth, Eric Emsellem, Mike Eracleous, Stephanie Escoffier, Xiaohui Fan, Emily Farr, Shuai Feng, José G. Fernández-Trincado, Diane Feuillet, Andreas Filipp, Sean P. Fillingham, Peter M. Frinchaboy, Sebastien Fromenteau, Lluís Galbany, Rafael A. García, D. A. García-Hernández, Junqiang Ge, Doug Geisler, Joseph Gelfand, Tobias Géron, Benjamin J. Gibson, Julian Goddy, Diego Godoy-Rivera, Kathleen Grabowski, Paul J. Green, Michael Greener, Catherine J. Grier, Emily Griffith, Hong Guo, Julien Guy, Massinissa Hadjara, Paul Harding, Sten Hasselquist, Christian R. Hayes, Fred Hearty, Jesús Hernández, Lewis Hill, David W. Hogg, Jon A. Holtzman, Danny Horta, Bau-Ching Hsieh, Chin-Hao Hsu, Yun-Hsin Hsu, Daniel Huber, Marc Huertas-Company, Brian Hutchinson, Ho Seong Hwang, Héctor J. Ibarra-Medel, Jacob Ider Chitham, Gabriele S. Ilha, Julie Imig, Will Jaekle, Tharindu Jayasinghe, Xihan Ji, Jennifer A. Johnson, Amy Jones, Henrik Jönsson, Ivan Katkov, Arman Khalatyan, Dr., Karen Kinemuchi, Shobhit Kisku, Johan H. Knapen, Jean-Paul Kneib, Juna A. Kollmeier, Miranda Kong, Marina Kounkel, Kathryn Kreckel, Dhanesh Krishnarao, Ivan Lacerna, Richard R. Lane, Rachel Langgin, Ramon Lavender, David R. Law, Daniel Lazarz, Henry W. Leung, Ho-Hin Leung, Hannah M. Lewis, Cheng Li, Ran Li, Jianhui Lian, Fu-Heng Liang, Lihwai Lin, Yen-Ting Lin, Sicheng Lin, Chris Lintott, Dan Long, Penélope Longa-Peña, Carlos López-Cobá, Shengdong Lu, Britt F. Lundgren, Yuanze Luo, J. Ted Mackereth, Axel de la Macorra, Suvrath Mahadevan, Steven R. Majewski, Arturo Manchado, Travis Mandeville, Claudia Maraston, Berta Margalef-Bentabol, Thomas Masseron, Karen L. Masters, Savita Mathur, Richard M. McDermid, Myles Mckay, Andrea Merloni, Michael Merrifield, Szabolcs Meszaros, Andrea Miglio,

Francesco Di Mille, Dante Minniti, Rebecca Minsley, and Antonela Monachesi. The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. ApJS, 259(2):35, April 2022. doi: 10.3847/1538-4365/ac4414.

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, D. J. Milligan, C. Panem, V. Poinsignon, D. Pourbaix, S. Randich, G. Sarri, P. Sartoretti, H. I. Siddiqui, C. Soubiran, V. Valette, F. van Leeuwen, N. A. Walton, C. Aerts, F. Arenou, M. Cropper, R. Drimmel, E. Høg, D. Katz, M. G. Lattanzi, W. O'Mullane, E. K. Grebel, A. D. Holland, C. Huc, X. Passot, L. Bramante, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, J. Hernández, A. Jean-Antoine-Piccolo, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, D. Ordóñez-Blanco, P. Panuzzo, J. Portell, P. J. Richards, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, J. Torra, S. G. Els, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, T. Lock, E. Mercier, M. Altmann, R. Andrae, T. L. Astraatmadja, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, G. Busso, B. Carry, A. Cellino, G. Clementini, S. Cowell, O. Creevey, J. Cuypers, M. Davidson, J. De Ridder, A. de Torres, L. Delchambre, A. Dell'Oro, C. Ducourant, Y. Frémat, M. García-Torres, E. Gosset, J.-L. Halbwachs, N. C. Hambly, D. L. Harrison, M. Hauser, D. Hestroffer, S. T. Hodgkin, H. E. Huckle, A. Hutton, G. Jasniewicz, S. Jordan, M. Kontizas, A. J. Korn, A. C. Lanzafame, M. Manteiga, A. Moitinho, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J.-M. Petit, A. Recio-Blanco, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, K. W. Smith, A. Sozzetti, W. Thuillot, W. van Reeven, Y. Viala, U. Abbas, A. Abreu Aramburu, S. Accart, J. J. Aguado, P. M. Allan, W. Allasia, G. Altavilla, M. A. Álvarez, J. Alves, R. I. Anderson, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, S. Antón, B. Arcay, A. Atzei, L. Ayache, N. Bach, S. G. Baker, L. Balaguer-Núñez, C. Barache, C. Barata, A. Barbier, F. Barblan, M. Baroni, D. Barrado y Navascués, M. Barros, M. A. Barstow, U. Becciani, M. Bellazzini, G. Bellei, A. Bello García, V. Belokurov, P. Bendjoya, A. Berihuete, L. Bianchi, O. Bienaymé, F. Billebaud, N. Blagorodnova, S. Blanco-Cuaresma, T. Boch, A. Bombrun, R. Borrachero, S. Bouquillon, G. Bourda, H. Bouy, A. Bragaglia, M. A. Breddels, N. Brouillet, T. Brüsemeister, B. Bucciarelli, F. Budnik, P. Burgess, R. Burgon, A. Burlacu, D. Busonero, R. Buzzi, E. Caffau, J. Cambras, H. Campbell, R. Cancelliere, T. Cantat-Gaudin, T. Carlucci, J. M. Carrasco, M. Castellani, P. Charlot, J. Charnas, P. Charvet, F. Chassat, A. Chiavassa, M. Clotet, G. Cocozza, R. S. Collins, P. Collins, and G. Costigan. The Gaia mission. A&A, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.

Dean R. H. Johnson and David R. Soderblom. Calculating Galactic Space Velocities and Their Uncertainties, with an Application to the Ursa Major Group. AJ, 93:864, April 1987. doi: 10.1086/114370.

Marc H. Pinsonneault, Joel C. Zinn, Jamie Tayar, Aldo Serenelli, Rafael A. García, Savita Mathur, Mathieu Vrard, Yvonne P. Elsworth, Benoit Mosser, Dennis Stello, Keaton J. Bell, Lisa Bugnet, Enrico Corsaro, Patrick Gaulme, Saskia Hekker, Marc Hon, Daniel Huber, Thomas Kallinger, Kaili Cao, Jennifer A. Johnson, Bastien Liagre, Rachel A. Patton, Ângela R. G. Santos, Sarbani Basu, Paul G. Beck, Timothy C. Beers, William J. Chaplin, Katia Cunha, Peter M. Frinchaboy, Léo Girardi, Diego Godoy-Rivera, Jon A. Holtzman, Henrik Jönsson, Szabolcs Mészáros, Claudia Reyes, Hans-Walter Rix, Matthew Shetrone, Verne V. Smith, Taylor Spoo, Keivan G. Stassun, and Ji Wang. APOKASC-3: The Third Joint Spectroscopic and Asteroseismic Catalog for Evolved Stars in the Kepler Fields. ApJS, 276(2):69, February 2025. doi: 10.3847/1538-4365/ad9fef.

Emily L. Hunt and Sabine Reffert. Improving the open cluster census. III. Using cluster masses, radii, and dynamics to create a cleaned open cluster catalogue. A&A, 686:A42, June 2024. doi: 10.1051/0004-6361/202348662.

Luke G. Bouma, Lynne A. Hillenbrand, Andrew W. Howard, Howard Isaacson, Kento Masuda, and Elsa K. Palumbo. Ages of Stars and Planets in the Kepler Field Younger than Four Billion Years. ApJ, 976(2):234, December 2024. doi: 10.3847/1538-4357/ad855f.

Yuxi(Lucy) Lu, Ruth Angus, Daniel Foreman-Mackey, and Soichiro Hattori. In This Day and Age: An Empirical Gyrochronology Relation for Partially and Fully Convective Single Field Stars. AJ, 167(4):159, April 2024. doi: 10.3847/1538-3881/ad28b9.

F. Anders, A. Khalatyan, A. B. A. Queiroz, C. Chiappini, J. Ardèvol, L. Casamiquela, F. Figueras, Ó. Jiménez-Arranz, C. Jordi, M. Monguió, M. Romero-Gómez, D. Altamirano, T. Antoja, R. Assaad, T. Cantat-Gaudin, A. Castro-Ginard, H. Enke, L. Girardi, G. Guiglion, S. Khan, X. Luri, A. Miglio, I. Minchev, P. Ramos, B. X. Santiago, and M. Steinmetz. Photo-astrometric distances, extinctions, and astrophysical parameters for Gaia EDR3 stars brighter than G = 18.5. A&A, 658: A91, February 2022. doi: 10.1051/0004-6361/202142369.

Maosheng Xiang and Hans-Walter Rix. A time-resolved picture of our Milky Way's early formation history. Nature, 603(7902):599–603, March 2022. doi: 10.1038/s41586-022-04496-5.

Luke G. Bouma, Elsa K. Palumbo, and Lynne A. Hillenbrand. The Empirical Limits of Gyrochronology. ApJ, 947(1):L3, April 2023. doi: 10.3847/2041-8213/acc589.

7 APPENDIX

7.1 DATA AND PREPROCESSING

***Kepler*.** The *Kepler* mission (Borucki et al., 2010) monitored a single $\sim$115 deg^2 field in Cygnus–Lyra for four years at 30-minute (long) cadence, delivering the highest-precision, longest-baseline optical photometry available for main-sequence stars. We use the long-cadence PDCSAP light curves (STScI, 2016) on their native, uniform $48\,\mathrm{samples\,day^{-1}}$ grid, with gaps zero-filled and no rebinning, cut into 450-day windows ($21{,}600$ points). Each window is lightly smoothed (Savitzky–Golay, $\sim$0.5 d) and expanded into six channels: standardized flux, median-normalized flux, two activity-amplitude envelopes (a slow envelope on a $6\,P_{\rm ref}$ window and a high-passed one on a $1\,P_{\rm ref}$ window), the autocorrelation function (ACF, McQuillan et al., 2013), and the Lomb–Scargle periodogram (LS, Lomb, 1976; Scargle, 1982). The reference period $P_{\rm ref}$ only sets the envelope windows; it is a flux-derived estimate from an earlier four-channel stage of the model, available for every star and never taken from a catalogue. The resulting sample is $193{,}951$ stars.

TESS. TESS (Ricker et al., 2015) trades *Kepler*'s depth for sky coverage, observing nearly the full sky in $\sim$27-day sectors at multiple cadences. We use light curves from the Quick-Look Pipeline (Huang et al., 2020), restricted to $T_{\rm mag} < 16$ and to targets observed in more than one sector. Each light curve is resampled to a 30-minute grid and truncated to 1000 points in four channels: standardized flux, median-filtered flux, ACF, and LS. These are the same channels as in *Kepler*, but without the activity envelopes, as the sectors are too short for them. Compared to *Kepler*, TESS has a wider coverage ($1{,}503{,}260$ encoded stars) but a shorter per-sector baseline, larger pixels ($21''$ vs. $4''$), and is generally noisier.

LAMOST. LAMOST (Zhao et al., 2012) is a 4-m telescope with 4000 fibers, conducting the largest optical spectroscopic survey of the northern sky. We use low-resolution ($R \sim 1800$) spectra from Data Release 8, continuum-normalized and resampled onto a common 4096-point wavelength grid. Its combination of scale (5.1M spectra) and northern footprint makes LAMOST the dominant spectroscopic partner for both *Kepler* and TESS.

APOGEE. APOGEE (Majewski et al., 2017) is a high-resolution ($R \sim 22{,}500$) H-band spectrograph. Working in the near-infrared lets it observe through dust into the Galactic plane and bulge, and it targets predominantly giants, a population complementary to the dwarf-dominated *Kepler* and GALAH samples. We use the continuum-normalized combined spectra from Abdurro'uf et al. (2022), resampled to 8576 points.

GALAH. GALAH (De Silva et al., 2015) is a high-resolution ($R \sim 28{,}000$) southern optical survey using the HERMES spectrograph on the Anglo-Australian Telescope, designed for Galactic archaeology and delivering up to $\sim$30 elemental abundances per star. We use the continuum-normalized spectra of Buder et al. (2025) over 4700–7900 Å, concatenated across the four HERMES arms onto a $14{,}420$-point grid, and learn, in addition to basic stellar parameters and $v \sin i$, 11 elemental abundances ([Mg/Fe], [Si/Fe], [Ca/Fe], [Ti/Fe], [Na/Fe], [Al/Fe], [Mn/Fe], [Ni/Fe], [Cr/Fe], [Ba/Fe], [Y/Fe]). GALAH's southern footprint is disjoint from *Kepler*'s.

Table 4: Survey census and pairwise overlaps in the cross-match master (unique stars; *Gaia* is the anchor and covers essentially all 1,386,210 stars). The diagonal gives each survey's total; off-diagonal entries give the number of stars observed by both surveys.

	Kepler	TESS	LAMOST	APOGEE	GALAH
Kepler	91,213	69,701	66,615	20,712	0
TESS		1,247,706	935,262	226,347	188,513
LAMOST			1,053,269	135,429	60,388
APOGEE				310,347	55,623
GALAH					247,444

Modality	Architecture	Params	Input Length	d_i	Pre-training labels
Gaia XP	Spectral ViT	38.5 M	340	512	$T_{\rm eff}$ [a]
LAMOST	Spectral ViT	615.3 M	4096	2048	$T_{\rm eff}$, $\log g$, [Fe/H]
APOGEE	Spectral ViT	154.2 M	8576	1024	$T_{\rm eff}$, $\log g$, [Fe/H]
GALAH	Spectral ViT	154.4 M	14,420	1024	$T_{\rm eff}$, $\log g$, [Fe/H] + 11 elements, $v \sin i$
Kepler	Dual-view LC	129.6 M	$6 \times 21{,}600$	768	$P_{\rm rot}$, $\log g$, M_K
TESS	Dual-branch LC[c]	139.5 M	4×1000	512	$P_{\rm rot}$, $\log g$[b], M_T

Table 5: The six frozen unimodal encoders. Input Length is the number of flux bins (spectra) or channels and time samples (light curves). d_i is the feature dimensionality. All encoders combine a self-supervised objective with the listed supervised heads; missing labels contribute zero gradient. Each encoder is trained on its own survey with its own split, independent of the alignment split (Section 7.5). [a] GSP-Phot, back-filled through the cross-match from LAMOST/APOGEE/GALAH where GSP-Phot is absent. [b] *Gaia* `logg_gspphot`, not spectroscopic. [c] A 100-way sector classification additionally supervises the instrument branch, which is discarded after pre-training.

***Gaia*.** *Gaia* (Gaia Collaboration et al., 2016) provides all-sky astrometry and low-resolution spectrophotometry, and serves here as the universal hub. We use two products from DR3 (Gaia Collaboration et al., 2023): (i) The BP/RP (XP) *sampled* spectra (De Angeli et al., 2023): 340 flux points spanning 330–1050 nm, scaled per-spectrum with a robust (median/IQR) scaler and deliberately *not* continuum-normalized, so that the absolute spectral-energy-distribution shape, which carries temperature and extinction information, is preserved. (ii) A 13-dimensional auxiliary vector: parallax with its error and signal-to-noise, the two proper motions, the G magnitude, RUWE and astrometric excess noise (log-scaled), the UVW space velocities derived with the radial velocity via the transformation of Johnson and Soderblom (1987), their total σ, and a flag marking stars with valid kinematics. Missing entries are imputed with the training median and all components are standardized with training statistics; a star without an auxiliary row receives the training mean.

7.2 Cross-match

All surveys are cross-matched onto a common star identity keyed by the *Gaia* DR3 `source_id`. The master cross-match retains every star observed by at least two of the five non-*Gaia* surveys: 1,386,210 unique stars in 1,613,142 rows, since repeat observations of a star by the same survey are kept as separate rows. Table 4 lists the per-survey census and the pairwise overlaps. The pair frequencies span roughly two orders of magnitude, from TESS×LAMOST (935,262 stars) down to *Kepler*×APOGEE (20,712), which motivates the inverse-square-root pair weighting used in the alignment objective (Section 7.4). Notably, *Kepler* and GALAH have *zero* common stars (northern vs. southern footprints): these two modalities are never observed together and can only be related transitively, through the shared latent space. Counting the number of non-*Gaia* surveys per star, 1,225,913 stars have exactly two, 143,035 have three, and 17,262 have four; no star is covered by all five. For 42,854 of these stars only one modality is available in our encoded data; they enter alignment training as single-modality stars (Section 7.4).

7.3 LABELS

Age labels are drawn from a tiered merge of twelve catalogues, combined into a single consensus age by a fixed priority order that reflects each method's reliability: asteroseismology (APOKASC-3, Pinsonneault et al., 2025) first, then cluster ages (Hunt and Reffert, 2024; Van-Lane et al., 2025), then field gyrochronology (Bouma et al., 2024; Lu et al., 2024), then isochrone and spectroscopic ages, drawn from StarHorse (Anders et al., 2022), GALAH DR4 (Buder et al., 2025) and the LAMOST subgiant sample of Xiang and Rix (2022), and the young moving-group sample of **?**. Three further catalogues apply the *gyrointerp* framework (Bouma et al., 2023) to rotation periods from *Kepler* (Kamai et al., 2026a) and TESS (Boyle et al., 2026). Where a star is dated by several catalogues the highest-priority method wins, and its origin is recorded per star. Of $683{,}951$ dated stars we keep $518{,}638$: asteroseismic and cluster ages regardless of quoted error; ages with relative error below 0.4; and ages from catalogues without errors only when at least two other catalogues agree within 30% of their median.

Rotation-based tiers (about 13% of age-labelled probe stars) derive from rotation periods that FAME's light-curve encoders were trained on. Restricting the age label to the other tiers preserves the margins over DESA (0.58 vs. 0.28) and Leung and Bovy (2023) (0.86 vs. 0.82).

7.4 TRAINING

Pre-training was done on a mixture of H100 and B200 GPUs. Each encoder was trained until convergence with early stopping; training times vary between several hundred and several thousand GPU-hours. The alignment module trains in about 1.3 hours.

Data. The alignment module is trained on a stratified subset of $300{,}000$ stars of the cross-match master ($9{,}000$ held for validation), drawn without regard to the evaluation split; all reported test sets therefore exclude these stars (Section 7.5). Half of every epoch is drawn, with replacement and balanced across modalities, from the $42{,}854$ single-modality stars (Section 7.2), so that each residual branch and the smoothness term also see stars without a cross-modal partner; the other half are multimodal stars, weighted by their rarest modality pair.

Objective. The invariance term sums the mean squared difference of shared tokens over every pair of modalities present for a star, each pair weighted by the inverse square root of its frequency; *Gaia* astrometry is never a view. Single-modality stars contribute to $\mathcal{L}_{\mathrm{res}}$, $\mathcal{L}_{\mathrm{recon}}$ and $\mathcal{L}_{\mathrm{smooth}}$ only. The variance and covariance moments are weighted by $1/q_a$, clipped at 20 and renormalized. $\mathcal{L}_{\mathrm{recon}}$ is the squared error on the masked dimensions of the clean z_i (60% masked, single pass). The invariance and variance terms have weight 25, all others 1. For $\mathcal{L}_{\mathrm{smooth}}$, $\varepsilon = 0.0139$ in coordinates normalized over $T_{\mathrm{eff}} \in [3000, 7500]$ K and $\log g \in [0, 5]$, giving $\sim$10 neighbours per batch.

Architecture and optimization. The Perceiver-IO uses 8 latents of width 32 and one learned output query per modality, projected to $d_s = 64$. We use AdamW (learning rate 10^{-4}, weight decay 10^{-4}, $1{,}000$ warm-up steps, cosine decay), batch size 256, and keep the checkpoint with the lowest validation loss.

7.5 EVALUATION PROTOCOL

Exposure. All probe test stars, and the test stars of Tables 3 and 7, are unseen by the alignment module: the $300{,}000$ stars of its training subset (Section 7.4) are removed from these test sets, which leaves about 80% of the original test split. Encoders are pretrained per survey with independent splits, so test stars may appear in encoder pretraining, including, for the pretraining labels (T_{eff}, $\log g$, [Fe/H], $v \sin i$, P_{rot}, abundances), their labels. Rows for these labels measure how pretraining knowledge survives alignment, not generalization. Age and activity enter no encoder or alignment objective.

Probes. All probes are ridge regressions on frozen, standardized features, with the regularization chosen by leave-one-out error on the training stars. Figure 4 pools 30 draws per point (3 probe seeds $\times$ 10 training subsamples; median and $1.4826\times$MAD), with n_{test} between $2{,}113$ ($v \sin i$) and $29{,}091$ (T_{eff}, $\log g$); the supervised curve uses 3 training runs. Figure 9 uses 10 subsamples (mean and s.d.). Tables 1 and 10 use 5 training subsamples on a fixed test fold at the largest budget available per label (2k–20k); Table 2 uses 50 subsamples with $n_{\mathrm{train}} = 500$ and $n_{\mathrm{test}} = 307$ (195 for age), and reports medians because a few DESA draws fail catastrophically. Error bars and spreads measure readout noise, the variation from resampling which stars the probe is fit on; they do not

include encoder or alignment training noise, for which we have single runs. Absent modalities enter the probe as zero blocks; replacing them by the training-set mean with a presence indicator per modality moves no arm of Figure 4 by more than 0.02, except at the smallest activity budgets, and changes no ordering relative to the pre-alignment features.

Fine-tuned models. Tables 3 and 7 are scored on the unseen test stars only. Table 6 is scored on the full test split (see its caption), and the generation examples of Section 5.5 are not restricted.

External models. AION-1 embeddings are mean-pooled within each survey block and concatenated; its *Gaia* block also ingests parallax, position and fluxes, whereas FAME-gaia reads only the sampled XP spectrum and the astrometry channel. The model of Leung and Bovy (2023) is given *Gaia* XP coefficients only, never label tokens, and is probed on its concatenated output tokens; Table 10 uses stars outside its training set. For DESA, rows require at least 300 stars outside its training set.

Rare populations. Each member is matched to its 5 nearest neighbours in standardized $(T_{\rm eff}, \log g, [\mathrm{Fe/H}], M_G, \varpi)$ among stars with the same modality signature and at least two modalities. Per-modality tokens are mean-pooled over present modalities. Each feature block is reduced to 64 principal components and standardized within the training folds, and classified with class-balanced logistic regression under 5-fold stratified cross-validation.

7.6 Why a continuity penalty produces a Fourier geometry

We sketch why minimising the smoothness penalty organises the shared coordinates into sinusoidal modes. The argument motivates the observed geometry rather than proving a theorem about the full objective.

The unnormalised penalty carries the density squared. We let $s(c)$ be the shared embedding at parameter c. An unnormalised graph penalty $\frac{1}{n^2}\sum_{a\neq b} w_{ab}\|s_a - s_b\|^2$ is a sum over pairs of stars drawn i.i.d. from the sampling density p, so its expectation integrates against $p(x)p(y)$. Expanding at small bandwidth,

$$\frac{1}{n^2}\sum_{a\neq b} w_{ab}\|s_a - s_b\|^2 \;\longrightarrow\; \varepsilon^2 \int \|s'(c)\|^2\, p(c)^2\, dc, \tag{2}$$

one factor of p per summation index. More generally, dividing the kernel by $q_a^\alpha q_b^\alpha$ (with $q_a \propto p(c_a)$ a density estimate; the kernel degree in Coifman and Lafon (2006), a precomputed kernel density estimate here) gives a continuum limit weighted by $p^{\,2-2\alpha}$. Taking $\alpha = 1$, i.e. $\tilde{w}_{ab} = w_{ab}/(q_a q_b)$, cancels the density to leading order and leaves the unweighted Dirichlet energy $E[s] = \int \|s'(c)\|^2\, dc$. Its Euler–Lagrange operator, under Neumann boundary conditions, is the ordinary Laplacian $-d^2/dc^2$.

The scale constraint carries the density too. The energy alone is minimised by a constant; VICReg's variance and covariance terms fix the scale, but as batch averages they impose unit variance under the sampling density p. Minimising the energy under that constraint is a Rayleigh–Ritz problem whose eigenfunctions satisfy $-s'' = \lambda\, p\, s$ and are warped by p. Weighting each variance and covariance moment by $1/q_a \propto 1/p(c_a)$, the $\alpha = 1$ correction a second time, replaces the constraint by unit variance under Lebesgue measure; treating the hinge and covariance terms as equality constraints, the problem becomes

$$-s_n'' = \lambda_n\, s_n, \tag{3}$$

with no density anywhere. The invariance term is left unweighted: it compares two modalities of the same star, a per-pair quantity with no density to cancel.

Fourier modes and their robustness. On $[0, L]$ with Neumann boundaries the eigenfunctions are $s_n(c) = \cos(n\pi c/L)$, $\lambda_n = (n\pi/L)^2$: the observed ladder of increasingly oscillatory coordinates. Even where the correction is imperfect, the operator remains of Sturm–Liouville type, whose n-th eigenfunction has exactly n nodes, so the node ladder is preserved regardless of residual density weighting. This robustness is why the ladder survives approximate stationarity.

Why two coordinates degrade. The one-dimensional argument does not extend directly to two dimensions. The distribution in $(T_{\rm eff}, \log g)$ is neither translation invariant nor defined on a rectangular domain, so higher eigenfunctions distort and mix through near-degenerate eigenvalues.

Measuring stationarity. We bin stars in the penalised coordinate (32 equal-width $T_{\rm eff}$ bins over 4200–6800 K with at least 50 stars each, on the slice $\log g > 4.2$, $|[{\rm Fe/H}]| < 0.2$), average the shared embedding per bin, centre, and form the Gram matrix G. Its best Toeplitz approximation T replaces each diagonal of G by its mean, and $R^2_{\rm Toep} = 1 - \|G - T\|_F^2 / \|G\|_F^2$; a translation-invariant kernel gives 1. Equal-count bins give 0.88 in place of 0.92.

Bandwidth. ε is fixed (chosen for $\sim$10 graph neighbours per batch); edges are retained within a 3ε ball. Optimising the bandwidth, including per-coordinate, is left to future work.

7.7 Additional results

Table 6: Stellar age ($\log_{10}$ Gyr) from single surveys and multimodal arms on the TARS test split ($n_{\rm test} = 98{,}774$; a different split from Table 7). Unimodal arms are XGBoost on one survey's frozen z_i without the astrometry channel; all multimodal arms receive it. The two last rows are end-to-end heads of comparable size (3.1M and 4.4M trainable parameters), mean $\pm$ half-range over 3 seeds. MAE in dex. About 19% of these test stars were seen in alignment training; restricting to the rest changes no re-scorable row by more than 0.002.

	R^2	MAE
Unimodal		
LAMOST	0.719	0.100
Gaia XP	0.685	0.116
TESS LC	0.595	0.158
Kepler LC	0.591	0.153
APOGEE	0.560	0.159
Astrometry	0.553	0.162
GALAH	0.543	0.163
Multimodal		
XGBoost on z_i (pre-alignment)	0.814	0.079
Transformer on z_i (pre-alignment)	0.851 ± 0.002	0.067
FAME (fine-tuned)	0.844 ± 0.001	0.069

Table 7: Cross-modal recovery of two survey-specific quantities ($n_{\rm test} = 48{,}558$ activity / $28{,}352$ $v \sin i$; stars seen in alignment training are excluded). Unimodal arms are fit on a single survey's frozen z_i; FAME arms are fine-tuned with dropout ($p = 0.5$) on the label-bearing survey (LAMOST for activity, APOGEE/GALAH for $v \sin i$), and both FAME rows of a column are the same model. The last row, with that survey withheld from every test star, is the cross-modal test. Both targets are modelled in log space and MAE is in dex: $\log R'_{\rm HK}$ for activity, $\log_{10}(v \sin i / {\rm km\,s^{-1}})$ for $v \sin i$.

	Activity ($\log R'_{\rm HK}$)		$\log_{10} v \sin i$	
	R^2	MAE	R^2	MAE
Unimodal				
LAMOST	0.658	0.098	0.423	0.147
APOGEE	0.151	0.159	0.713	0.105
GALAH	0.143	0.160	0.574	0.110
Gaia XP	0.352	0.140	0.564	0.129
TESS LC	0.260	0.151	0.389	0.155
Kepler LC	0.146	0.160	0.333	0.161
Astrometry	0.136	0.161	0.342	0.161
Multimodal				
XGBoost on z_i (pre-alignment)	0.671	0.096	0.884	0.066
FAME (fine-tuned)	**0.736**	**0.086**	**0.921**	**0.055**
Label survey withheld				
FAME, withheld	0.469	0.127	0.640	0.120

Table 8: Linear-probe AUC of known-peculiar populations against matched controls, on stars unseen by the alignment module. Same stars and folds for all columns; each feature block is reduced to 64 principal components fitted on the training folds.

Population	n	shared	residual	z_i
s-process (Ba/Sr)	322	0.56	**0.72**	0.69
GCAS (γ Cas/Be)	163	0.60	**0.70**	0.65
DSCT (δ Scuti)	105	0.55	**0.80**	0.76
ECL (eclipsing)	428	0.70	**0.91**	0.88
RR Lyrae	46	0.79	0.95	**0.98**
Li-rich giants	256	0.76	0.90	**0.95**

Table 9: All three AION-1 sizes (B/L/XL = base/large/xlarge) on the stars of Table 1; mean over 5 repeats. AION tokens are mean-pooled within each survey block and concatenated. *gaia*: *Gaia*-only restriction; AION's *Gaia* block also ingests parallax, position and fluxes; FAME-gaia reads the sampled XP spectrum and the astrometry channel.

		AION				AION-gaia		
target	FAME	B	L	XL	FAME-gaia	B	L	XL
activity	0.653	0.267	0.310	0.218	0.286	0.404	0.415	0.382
$\log P_{\rm rot}$	0.641	0.351	0.403	0.321	0.302	0.475	0.472	0.460
$v \sin i$	0.837	0.560	0.596	0.524	0.597	0.670	0.675	0.680
[Fe/H]	0.929	0.792	0.823	0.743	0.784	0.859	0.864	0.855
age	0.756	0.641	0.659	0.621	0.648	0.699	0.711	0.702
$T_{\rm eff}$	0.988	0.965	0.970	0.965	0.975	0.975	0.975	0.975
$\log g$	0.980	0.950	0.961	0.953	0.944	0.967	0.969	0.968

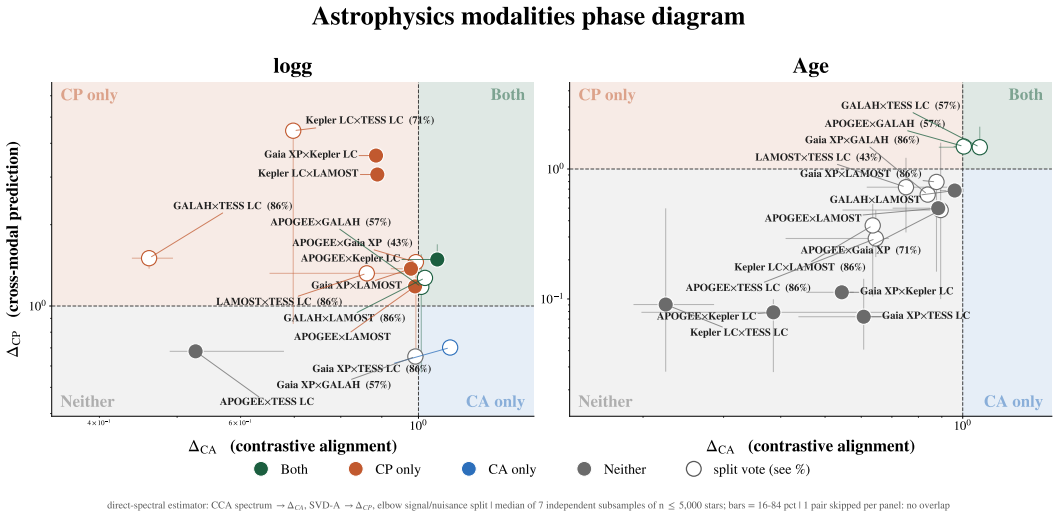

Figure 6: Phase diagram prediction for signal recovery based on Kamai et al. (2026b). The left panel shows the prediction of $\log g$ and the right panel the predictions of stellar age. Each prediction used a small number of labeled samples (5,000), and the pre-trained features to predict recovery regimes for cross-prediction and cross-alignment methods. It can be seen that many pairs, mainly in the stellar age case, sit in the *Neither* regime, where none of the methods are optimal.

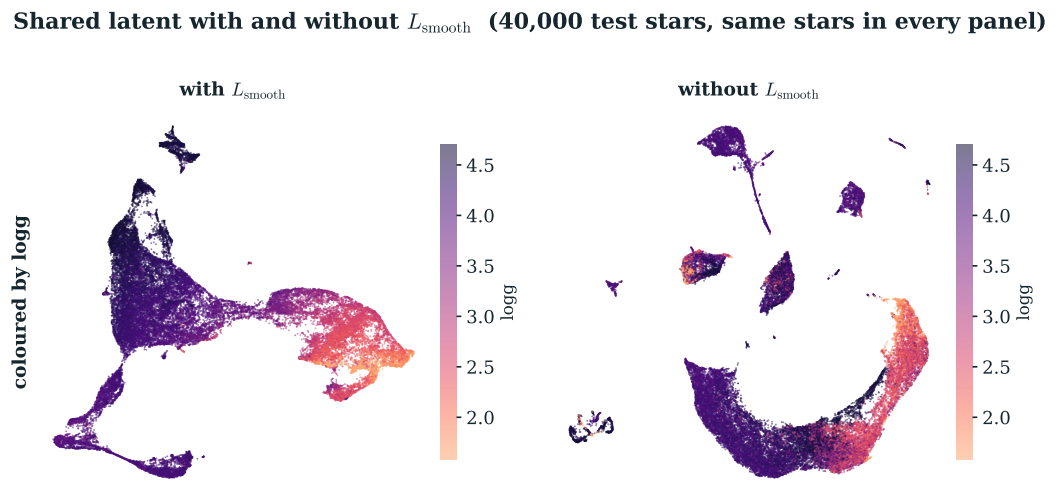

Figure 7: Effect of $\mathcal{L}_{\text{smooth}}$ on the latent space. **Left -** UMAP of the shared latent space of FAME with $\mathcal{L}_{\text{smooth}}$. **Right -** UMAP of the shared latent space of the same model without $\mathcal{L}_{\text{smooth}}$. It can be seen that the addition of $\mathcal{L}_{\text{smooth}}$ creates a more continuous latent space with a clean $\log g$ gradient.

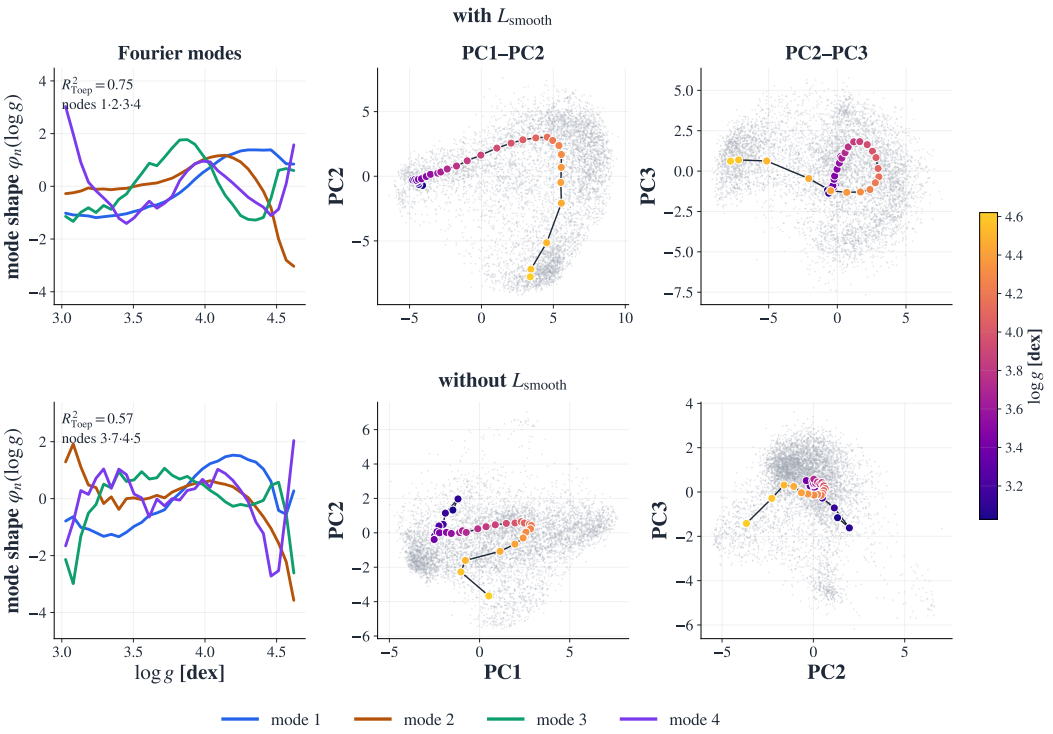

Figure 8: Same as Figure 3, but for $\log g$.

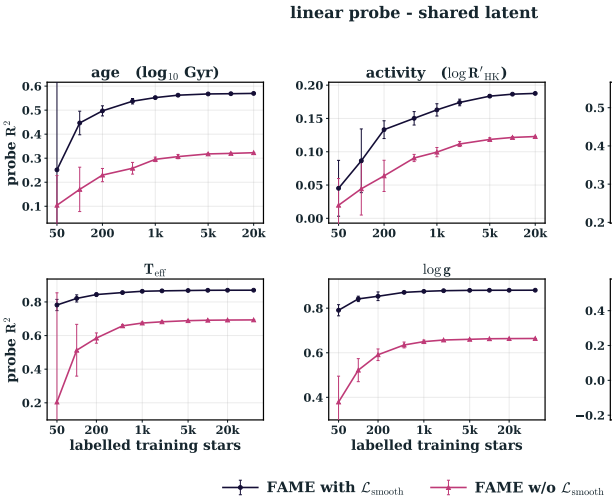

Figure 9: Effect of $\mathcal{L}_{\mathrm{smooth}}$ on downstream tasks. Linear probing of the shared latent space alone, with $\mathcal{L}_{\mathrm{smooth}}$ (black) and without (magenta), on test stars unseen by the alignment module; mean $\pm$ s.d. over 10 training subsamples.

Table 10: R^2 of FAME vs. Leung and Bovy (2023) (L&B) under a frozen ridge probe, on stars outside L&B's training set and FAME's alignment training; mean $\pm$ s.d. over 5 training subsamples (fixed test fold). L&B sees *Gaia* XP only; FAME-gaia adds the astrometry channel. Bold targets are the like-for-like labels; LAMOST labels are pretraining targets of FAME's encoders and APOGEE labels of both models.

target	L&B	FAME-gaia	FAME-full
age	0.763 ± 0.002	0.776 ± 0.001	$\mathbf{0.831} \pm 0.002$
activity	0.232 ± 0.007	0.246 ± 0.027	$\mathbf{0.587} \pm 0.009$
LAMOST T_{eff}	0.977 ± 0.001	0.986 ± 0.001	$\mathbf{0.995} \pm 0.001$
LAMOST $\log g$	0.934 ± 0.001	0.955 ± 0.001	$\mathbf{0.987} \pm 0.001$
LAMOST [Fe/H]	0.810 ± 0.005	0.812 ± 0.001	$\mathbf{0.958} \pm 0.001$
APOGEE T_{eff}	0.854 ± 0.011	$\mathbf{0.903} \pm 0.002$	0.897 ± 0.012
APOGEE $\log g$	0.903 ± 0.002	0.927 ± 0.001	$\mathbf{0.973} \pm 0.002$
APOGEE [M/H]	0.628 ± 0.014	0.604 ± 0.009	$\mathbf{0.930} \pm 0.007$

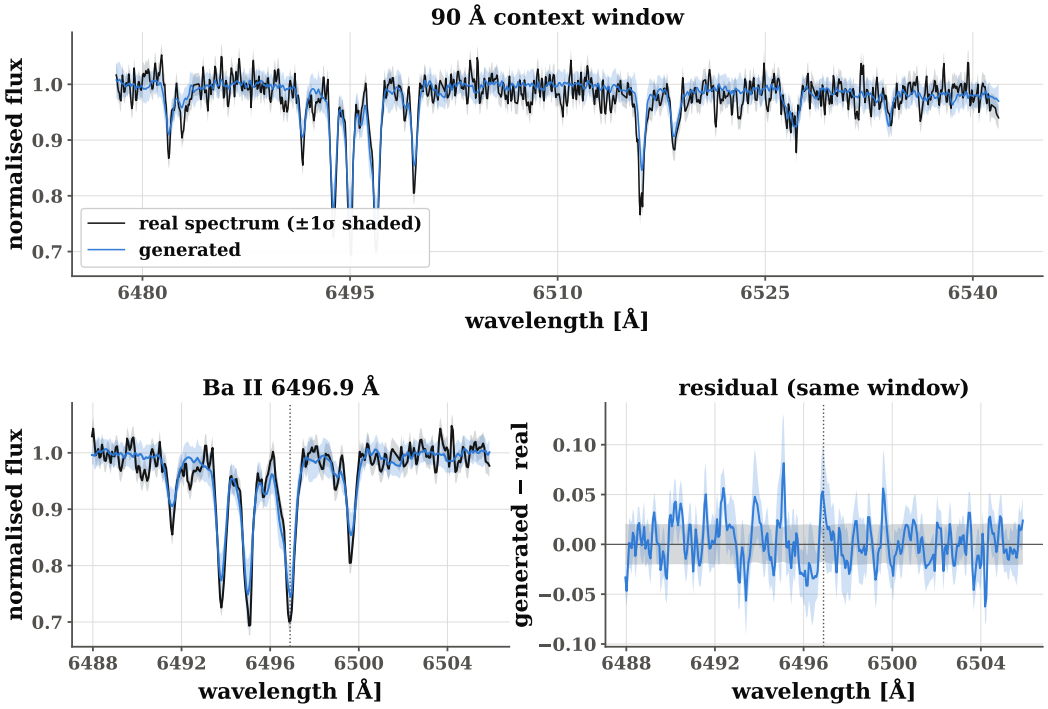

Figure 10: Cross-modal generation example of GALAH spectra. The blue curve is the sample mean with 32 samples. The shaded region is the standard deviation. The black curve is the true spectra. The lower panel shows a small window around Ba II line.

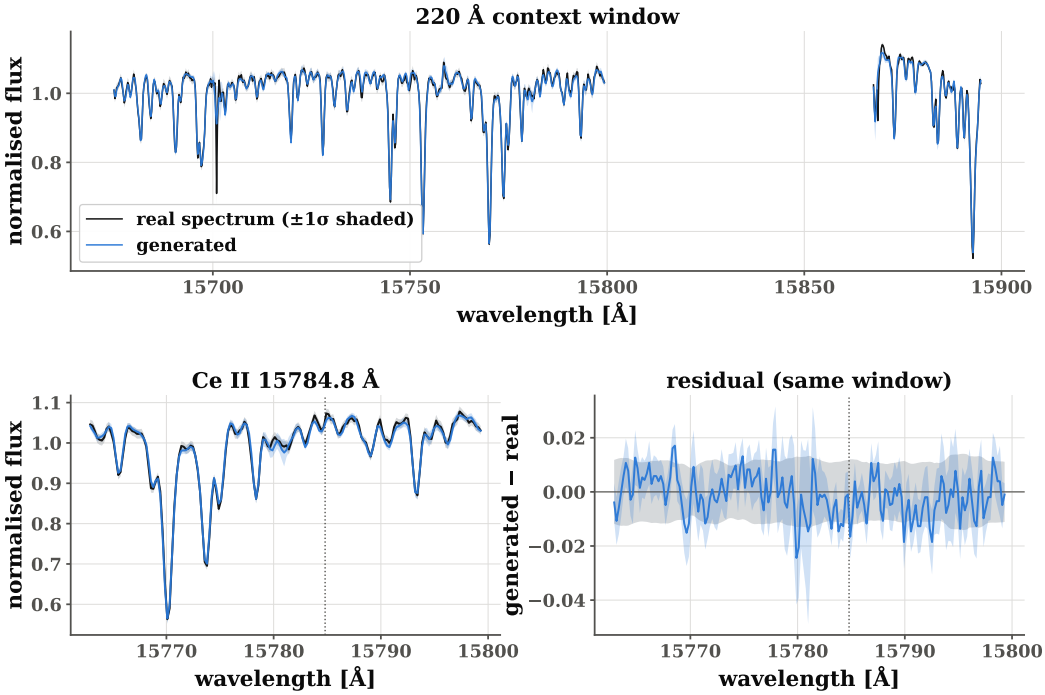

Figure 11: Cross-modal generation example of APOGEE spectra. The blue curve is the sample mean with 32 samples. The shaded region is the standard deviation. The black curve is the true spectra. The lower panel shows a small window around Ce II line.